



Ya-Ping (Amy) Hsiao,
Brenda den Oudsten &
Cor Sluijter (ed.)

A Practical Guide to Quality Assessment in Dutch Higher Education

**OPEN
PRESS
TiU**

TILBURG  UNIVERSITY

A Practical Guide to Quality Assessment in Dutch Higher Education

Ya-Ping (Amy) Hsiao, Brenda den Oudsten & Cor Sluiter
Editors



Open Press Tilburg University
Tilburg, The Netherlands

© 2026 Amy Hsiao, Brenda den Oudsten, Cor Sluijter

Design and typesetting: LINE UP boek en media bv
Cover image: Jean-Paul Wettstein (www.pexels.com/)

ISBN: 9789403935164
DOI: <https://doi.org/10.56675/9789403935164>



This is an Open Access book published under the terms of the Creative Commons 4.0 International license (CC BY NC-SA 4.0). The full licence terms are available at creativecommons.org/licenses/by-nc-sa/4.0/legalcode.

Contents

Preface	7
Contributors	8
A Note on How to Read This Book	9
Acknowledgements	12
Introduction	13

SECTION 1 INTRODUCTORY CHAPTERS	17
---------------------------------	----

1.1 Course-Level Assessment Design: Guidance for Course Coordinators	19
1.2 The Assessment Cycle	31
1.3 Role as Examiner and Essential Rules from Various Policy Documents	33
1.4 Estimating Written Exam Time	34
1.5 Setting a Passing Score and Transforming Raw Scores to Grades	37
1.6 Formative Assessment	44

SECTION 2 THEORY	53
------------------	----

2.1 Exams With Selected-Response Question Types	55
2.2 Exams With Constructed-Response Question Types	62
2.3 Exams With Both Selected- and Constructed-Response Question Types	68
2.4 Oral Exams	72
2.5 Open-Book and Stepwise Exams	78
2.6 A Follow-Up Oral Process Interview: What It Is and What It Should Not Be	82

INTERMEZZO – Boundary Crossing in Assessment: Integrating Examinations and Coursework	90
---	----

2.7 Assessment of Writing Assignments	92
2.8 Group Assignments	103
2.9 Simulation-Based Assessment	112
2.10 Portfolio Assessment	117
2.11 Reflection and Logbook Assessment	124
2.12 Internship Assessment	131
2.13 Thesis Assessment	139
2.14 Quick Guide to Thesis Oral Defenses	149
Appendix 2.14.1 Guidelines and Procedures: Examples and Instructions	154
Appendix 2.14.2 Checklist for Examiners – Preparing the Oral Defense	160
Appendix 2.14.3 Student Checklist – Thesis Submission and Oral Defense Preparation	163
2.15 Quality Assurance of Course-Level Assessment	166

SECTION 3 DESIGN CASES	173
A Note on Section 3	175
3.1 Formative Testing in an Interdisciplinary Minor Course	180
3.2 From Formative Quizzes to iRAT-tRAT: Strengthening Analytical Reasoning in Philosophy of Law	186
3.3 Exams with Mixed Question Types	194
3.4 Introducing Oral Examinations in a Program	200
3.5 A More Neuro-Inclusive Assessment of Presentation Skills	206
3.6 Becoming Critical Writers Through Dialogic Feedback and Co-Teaching	214
3.7 WIKI-Assignment in Psychopharmacology Medical Psychology	221
3.8 Groupwork in Personnel Psychology	228
3.9 Team Assignments in Strategic Consultancy	235
3.10 Simulation Design Case: Humanitarian Response Simulation Exercise	245
3.11 Simulation-Based Assessment in Collaboration in the Supply Chain	250
3.12 Simulation-Based Assessment in Data-Driven Design Approaches	260
3.13 Reflection and Logbook Assessment	269
3.14 Internship Assessment	276
3.15 Bachelor Thesis Assessment – Creating a collaborative research experience	284
3.16 Doing Capstone Projects Differently at Entrepreneurship and Business Innovation	291
3.17 Master Thesis Organization and Management Studies	302
Appendix 1. Technology Statements and Declaration based on AI index Tilburg	312
Glossary	315
References	321

Preface

This practical guide explores the principles and practices of **assessment** in higher education. It adopts a conversational style to engage with our readers. Therefore, we occasionally address you, the reader, directly in specific sections to make the content more relatable and accessible. To help you become familiar with the terminology of educational measurement, the online version of this document hyperlinks key technical terms the first time they appear in each chapter in the main text, so you can consult their definitions in the glossary immediately.

Although this guide is useful for all stakeholders involved in assessment organizations, its primary audience is lecturers (i.e., examiners) who are responsible for the quality of their course assessments, academic directors who oversee and monitor course assessments to ensure an appropriate and balanced program design, and members of examination boards who have to assure the **transparency, reliability** and **validity** of all assessments. Although this guide draws from the experience of personnel of Tilburg University, it is our hope that it will also be useful to staff at other Dutch universities and universities of applied science.

The guide is designed to help you understand and implement effective assessment strategies. Assessment refers to a structured process of collecting information on a student's knowledge, abilities, skills and competencies, enabling informed and valid educational decisions. Irrespective of its concrete form, an assessment can be seen as a stimulus – or a set of stimuli – prompting students' responses. The information thus collected is used to evaluate **learning outcomes** and guide academic progression.

Throughout their academic journey, students encounter various assessments that collectively inform the graduation decisions leading to the conferral of a bachelor's and master's degree. The purpose of this guide is to help you contribute to the accuracy of these graduation decisions. It provides key theory and examples of current assessment methods used at Tilburg University (Section 3) to help you increase your **assessment literacy**.

The editors wish to express their gratitude to the former and current chairs and members of the different Examination Boards at Tilburg University, colleagues from the Education Support Teams, assessment specialists, and lecturers for their contributions to the assessment handbooks of Tilburg School of Humanities and Digital Sciences (TSHD), Tilburg School of Economics and Management (TiSEM) and Tilburg School of Social and Behavioral Sciences (TSB) that were the foundation for the current document and for their help in preparing this publication.

Ya-Ping (Amy) Hsiao
Brenda Den Oudsten
Cor Sluijter
Editorial Group

Contributors

Tilburg School of Catholic Theology (TST)

Anton ten Klooster

Tilburg School of Economics and Management (TiSEM)

Astrid Kramer, Aswin van Oijen, Chong Zhang, Jolijn van Cruchten, Jorna Leenheer, Koen van den Oever, Louise van Hoek, Miranda Stienstra, Peter de Goeij, Roland van de Kerkhof, Werner Liebrechts, Wynand Bodewes

Tilburg School of Humanities and Digital Sciences (TSHD)

Constanze Hopfner, Frea Dooremaal, Jos Swanenberg, Hanne Jacobs, Marijn van Wingerden, Minne Bakker, Pytrik Schafrad, Sinsemilla Blokvoort, Yvette Drissen

Tilburg School of Social and Behavioral Sciences (TSB)

Antonio Massella, Brenda Den Ouden, Djurre Holtrop, Kerry Forrester, Leanne Prev-Heckman, Lonneke Schellekens, Paula Mommersteeg, Roel Rutten

Tilburg Law School (TLS)

Bart van de Moosdijk

Teaching & Learning Center

Ya-Ping (Amy) Hsiao¹, Vincent Kalis

The Association for Educational Assessment – Europe (AEA-Europe)

Cor Sluiter

Educational Development & Training, Utrecht University

Laura Koenders

Vrije Universiteit Amsterdam (VU Amsterdam)

Sam Passeport

¹ Also affiliated at Tilburg School of Economics and Management (TiSEM) and Tilburg Center of Learning Sciences (TiCeLS)

A Note on How to Read This Book

Three Sections

This book is organized into three sections. Section 1 provides introductory chapters that guide readers in reflecting on the starting points of their own educational contexts, with particular attention to Dutch higher education. It also introduces course-level assessment design principles that account for multiple forms of **assessment**, outlines the core quality criteria of good assessment, **transparency**, **reliability** and **validity**, and explains the evaluative function of assessment, including key considerations for quality assurance at the course level.

Section 2 presents theoretical perspectives on what constitutes good assessment across different assessment types and how each quality criterion can be met. These assessment types can be applied to course- and graduation-level contexts. Each chapter concludes with a practical checklist designed to support lecturers in applying the so-called four-eyes principle for their own quality assurance.

Section 3 features course- and graduation-level assessment design cases situated at Tilburg University. These cases illustrate ongoing efforts toward realizing the vision of assessment as learning. As such, they represent work in progress and intentionally leave room for further development in response to the continuous and emerging challenges of higher education, including technological advancements and broader societal uncertainties. Furthermore, these design cases also demonstrate the application of the recently developed TiU AI index (see Table o).

Targeted Audience

Although this document is useful to all stakeholders involved in assessment organizations, it is primarily intended for lecturers (i.e., examiners), who are responsible for the quality of their course assessments. Academic directors, who oversee and monitor course assessments to ensure a more appropriate and balanced program design may find this document of value as well. This document is also of help to Examination Boards. In the context of Dutch higher education, since 2010, when legislation increased their responsibilities in ensuring the quality of assessments, Examination Boards have become ultimately responsible for quality assurance and quality policy relating to exams and tests. Ultimately, Examination Boards guarantee the value of diplomas.

Quality Ownership of Assessment

We emphasize that academic directors and examiners must take ownership of ensuring the quality of their course assessments, next to an authorized entity such as an Examination Board (i.e., guarantee assessment quality, “borging” in Dutch) and faculty educational management teams (i.e., safeguarding assessment quality, “zorg” in Dutch). To that

end, this document is designed to guide and assist both lecturers and academic directors in fulfilling this responsibility. We integrate the Plan-Do-Check-Act (PDCA) cycle with the existing course design and assessment cycle (established in the University Teaching Qualification/UTQ trajectory), ensuring that quality enhancement becomes an inherent part of assessment practices rather than a separate, added task to the work of programs and lecturers. Additionally, policymakers and support staff, such as assessment specialists and educational experts, can also use this guide as a source of both practical and theoretical guidance for ensuring assessment quality.

Assessment Quality Criteria

Based on recent developments in assessment practices and research, we have revised the definitions of the three key criteria, transparency, reliability, and validity from earlier faculty assessment handbooks of Tilburg University.

The Approach of This Guide

We designed this book to be both practical and comprehensive for its readers. It is intended as a reference guide that readers can consult as needed, without the need to read it from beginning to end. For this reason, we use clear headings and a structured layout to support quick and easy navigation. Throughout the book, we also address readers directly using a conversational “you” style to create an engaging and accessible reading experience in Section 1 and 2.

Table o.

Index Tilburg University, Adapted by Hsiao, den Oudsten, Venhuizen, and Engelen (2025), Based on the Work of Perkins et al. (2024)

	Level in AI use for students	Required AI literacy level	Example instructions for students in the syllabus or assignment descriptions	Recommended lecturer monitoring throughout the process	Requirement(s) next to final product
1	No AI	None	The assessment is completed entirely without AI assistance. You must not use AI at any point during the assessment. You must demonstrate your core knowledge and skills.	Initial brainstorming and idea generation during interactive lectures to ensure that students rely on their existing knowledge, understanding and skills	No
2	AI-assisted editing	Basic TiU e-module	You are allowed to use GenAI tools to review completed drafts for editing, including spelling, punctuation, grammar, syntax, and referencing.	Submission of drafts, with lecturer random checks	<ul style="list-style-type: none">• Technology Statements & Declaration (Appendix 1)• Optional: submission of intermediate drafts
3	AI-assisted idea generation and structuring	Basic TiU e-module	You are allowed to use GenAI tools to develop or refine initial ideas, materials (e.g., interview guides, stimuli for experiments), paraphrasing, structures, or outlines.	Submission of initial ideas, structures, or draft outlines, with lecturer random checks	
4	AI task completion, human evaluation	Beyond Basic TiU e-module	You are allowed to use the GenAI tool to complete certain elements of the task (e.g., literature search, data analysis, audio transcription), and you are responsible for critically evaluating its output to meet the assessment requirements.	Submission of drafts, discussion or commentary on the AI-generated content, with lecturer random checks	
5	Full AI	Beyond Basic TiU e-module (or if the course is AI literacy itself)	You are allowed to use GenAI tools as a “co-pilot” to complete your work, and you are responsible for critically evaluating its output to meet the assessment requirements.		

Acknowledgements

We gratefully acknowledge the external reviewers of parts of this practical guide, Henk van Berkel² and Bas Hemker³, for their insightful comments and constructive feedback.

We also would like to thank all chairs of the Examination Boards and policy staff for their support on this practical guide and their advice at the early stage.

- Jac Kragt, Chair, Examination Board, Tilburg School of Economics and Management (TiSEM)
- Roel Rutten, Former Chair of Examination Board Social Sciences, Tilburg School of Social and Behavioral Sciences (TSB)
- Brenda Den Oudsten, Chair of Examination Board Psychology, Tilburg School of Social and Behavioral Sciences (TSB)
- Colette Cuijpers, Chair of Examination Board, Tilburg Law School (TLS)
- Jan Engelen, Chair of Examination Board, Tilburg School of Humanities and Digital Sciences (TSHD).
- Marcel Sarot, Chair of Examination Board, Tilburg School of Catholic Theology (TST)
- Jolanda Bachrach, Senior Policy Staff

We also express our gratitude to Marcel van Assen, former Vice Dean of the Tilburg School of Social and Behavioral Sciences, for his involvement in early discussions on the content of this guide and his encouragement to further develop it.

We would like to thank the reviewers of earlier versions of several chapters from the Assessment Handbook of the Tilburg School of Humanities and Digital Sciences (TSHD), which have been adapted for this guide.

- Amanda Cawston, Former Academic Director, Department of Philosophy
- Dries Deweer, Former Academic Director, Department of Philosophy
- Drew Hendrickson, Former Academic Director, Department of Cognitive Science and Artificial Intelligence
- Maria Mos, Former Academic Director, Department of Communication and Information Sciences
- Martin Hoondert, Former Academic Director, Department of Culture Studies
- Pascale Veenings, Former Policy Staff
- Rein Cozijn, Former Chair of the Examination Board
- Tessa Leesen, Former Academic Director, Liberal Arts and Sciences

This book is part of the innovation agenda of TUNED IN. We would like to thank TUNED IN for its support in the development of this book.

We would also like to thank Een Slimmer Collegejaar for their support.

This book was developed with the assistance of artificial intelligence tools for editing (TiU AI-index level 2) and initial brainstorming (TiU AI-index level 3). All final content has been reviewed and approved by the editorial group.

² Editor of Toetsen in het hoger onderwijs

³ Senior Educational Assessment and Psychometrics Researcher at Cito

Introduction

An **assessment** is a systematic procedure for gathering information about students' knowledge, skills, attitudes, and abilities to judge their achievement and learning. Its purpose is to form '...a judgment about the quality and extent of student achievement or performance, and therefore by inference a judgment about the learning that has taken place' (Sadler, 2005, p.177). The term derives originally from the Latin verb 'assidere', meaning 'to sit beside'. However, its current meaning in education came around in another way. The noun 'assessment' appeared in English in the sixteenth century, derived from the medieval Latin verb 'assessare', which, although derived from 'assidere', originally had a pure financial meaning: the valuation of property for taxation. Its general sense as 'estimation' dates from the next age, while its use as a technical term in educational measurement only gradually emerged in the twentieth century. Probably because of the need for a broader term in educational measurement literature than 'testing' or 'evaluation'.

However, in this document, assessment reflects its original roots. We interpret it as figuratively sitting alongside learners, to not only appraise or judge their learning, but also to support and guide it. Thus, it refers to a broad array of methods, including traditional ones like multiple-choice tests and theses, but also to more authentic tools like portfolios and essays. Independent of the methods chosen, examiners can use it to grade student achievement (**summative assessment**) and to give feedback to support and improve future learning (**formative assessment**). Whatever specific term is used, be it assessment, test, exam, or examination, the instruments or procedures these terms refer to all have the same basic tenet. Without exception, they are based on the assumption that a relatively small sample of a student's behavior in some competence domain, collected under controlled conditions, can be generalized to the student's behavior under much broader circumstances.

Assessment, often viewed as a modern practice, actually dates back to ancient China. Early assessments began around 2200 BC when emperors periodically evaluated officials' competence. By 1115 BC, formal procedures required candidates to demonstrate proficiency in writing, arithmetic, archery, horsemanship, music, and ceremonial skills. During the Han dynasty (202 BC–200 AD), written exams covered civil law, military affairs, agriculture, revenue, and geography. The system evolved over centuries, reaching its final form in 1370, where citizens needed to pass three competitive exams to attain public office. The first exam tested literary skills; only one to seven percent passed annually. The second, held triennially in provincial capitals, assessed erudition, with a pass rate of one to ten percent. Success led to the third exam in Beijing, granting eligibility for mandarin positions, but only about three percent succeeded. This assessment system persisted until the twentieth century, when it was replaced by university education. The Chinese model influenced examination systems in England, France, the United States, and beyond, and by the mid-1800s, written exams had become standard tools for student placement and evaluation in Europe and America. An elaborate description of the assessment system in ancient China can be found in Teng (1943). More concise representations can be found in DuBois (1970) and Wainer (Wainer 2000).

The use of quantitative measures in assessing student performance began in all probability with William Farish, an eighteenth-century Cambridge professor. He is said to be the first to break with the tradition of ranking written products of students in 1792. Instead, he expressed the quality of their work in a number. Farish was reportedly inspired by the shoe industry's practice of measuring product quality (Madaus & O'Dwyer, 1999). However, the quantification of learning performance only really took off at the beginning of the twentieth century, partly due to the rise of the intelligence test. World War I led to the development of mass testing methods, including multiple-choice **items**, which contributed to the widespread adoption of standardized tests for education and other uses. Now, more than a hundred years later, educators have become so accustomed to using numbers to indicate the quality of learning performance that they can hardly imagine how an education system could function without them.

The primary concern in the design of an assessment is that it can be used for its intended purpose. The three demands an assessment has to meet to be useful are those of **transparency**, **reliability**, and **validity**.

Transparency is the foundation of student-centered education and a prerequisite for any credible quality assessment. Transparency refers to the degree to which all necessary information about an assessment has been provided to both students and examiners (Hattie, 2015). This includes clear information on the nature, design and purpose of the assessment, scoring and grading policies, preparation materials, and assessment instructions. Transparency should not be perceived by students as mere compliance with criteria or as spoon-feeding instructions on how to achieve a grade (Balloo et al., 2018; Sadler, 2009), nor should it be interpreted by lecturers as the need to provide overly detailed or prescriptive instructions. Instead, a transformative approach emphasizes mutual communication, clarity of roles, and shared understanding of quality standards, enabling students and lecturers to take ownership and better manage learning and teaching (Balloo et al., 2018).

Reliability in educational measurement refers to the reproducibility and stability of assessment outcomes, whether produced by human or automated processing agents. These agents apply specific rules to interpret student responses to the set of stimuli in an assessment and generate scores that reflect performance. A reliable assessment yields similar outcomes across different administrations, different raters and different sets of stimuli, ensuring that the measurement is replicable. A reliable assessment therefore provides a sound basis for an accurate evaluation of student achievement.

Student behavior, however, is never totally consistent over different administrations and different sets of stimuli. And rater judgments of identical responses are unlikely to show complete agreement. Minor discrepancies may arise from differences in interpretation, attention, or application of scoring criteria. Edgeworth (1888) was the first to note that these inconsistencies or 'errors of measurement' could be quantified and estimated and therefore could be used to examine the quality of an assessment. Spearman (1904) developed the concept further and formally introduced the reliability coefficient to establish the measuring accuracy of an assessment. Reliability coefficients can be calculated to indicate to what extent an assessment produces stable results over the dimensions mentioned: instances, sets of stimuli, and raters.

Reliability is a necessary but insufficient condition for an assessment to be useful for its intended purpose. A reliable assessment outcome still needs to be validated, which means that it is necessary to explain why the intended evaluation and decision follow from the outcome. On close inspection, it becomes clear that all assessments are meant to make decisions. Either to grade or judge individuals, groups or systems, or to give feedback and adjust instruction. Validity is an overall evaluation of the degree to which evidence and theory support the intended use of specific assessments.

Classroom assessment differs fundamentally from traditional measurement contexts, requiring a different approach to reliability and measurement (Brookhart, 2003; Förster & Rojas-Barahona, 2023; Parkes, 2007; Parkes & Zimmaro, 2017; Smith, 2003). The assumptions underlying measurement theories used in these contexts, **classical test theory**, **generalizability theory**, and **item response theory**, including sampling, dimensionality, and local independence of items, limit their applicability in authentic classroom assessment contexts. For example, course learning goals span multiple domains, including knowledge, skills, and character, which inherently conflict with the

assumptions required for these traditional reliability measures. Therefore, psychometric indices should be regarded as interpretive tools rather than absolute indicators (Smits & Klinkenberg, 2026), and several factors, such as the exam type (e.g., only selected-response questions or mixed question types), course context, and student characteristics, should be considered when interpreting them (e.g., Cronbach's alpha) (Frisbie, 1988; Hoekstra et al., 2019).

Therefore, in classroom assessment, we adopt the definition of reliability as the sufficiency of information (Brookhart, 2003; Förster & Rojas-Barahona, 2023; Smith, 2003) within a reasonable time frame. That is, within the allocated time for exams or coursework, do lecturer(s) have enough evidence to make informed decisions about where students are in the instructional process? Will the assessment provide answers to instructional concerns? Have students been given meaningful opportunities to demonstrate what they can do?

Additionally, we emphasize objectivity by cultivating the integrity of examiners in grading (Brookhart et al., 2016; Guskey et al., 2024), recognizing that grades should reflect performance, not the student as a person. Institutions must define what constitutes an ethical grading process to minimize unconscious bias or inappropriate expectations. As part of this process, schools should require examiners to apply the four-eyes principle (i.e., peer review) when designing and reviewing exam items or assessment tasks, as well as when uncertainties arise regarding pass-or-fail decisions or exceptionally high grades (9–10). Grading should be aligned with course learning goals and assessment purposes, whether evaluating the final product or the learning process. Grades must be assigned through thoughtful and informed professional judgment, supported by well-articulated purposes, the collection of valid, representative, and sufficient evidence, as well as clear and understandable assessment criteria.

Validity is a unitary concept, although there are different sources of evidence (Messick, 1989). The validation process involves accumulating evidence to provide a sound scientific basis for the proposed use of the assessment (AERA, 2014). The development of a sound 'validity argument' (e.g., Kane, 2013) begins with an explicit statement of the proposed interpretation of assessment scores, along with a rationale for the relevance of the interpretation to the proposed use. The proposed interpretation refers to the **construct** the assessment should measure. The description of the proposed interpretation provides a conceptual framework that specifies the meaning of the construct in light of the purposes of the assessment, distinguishes it from other constructs, specifies the context in which the assessment will take place and indicates how measures of the construct should relate to other variables.

In terms of recent developments, it is important to note that the use of technology (e.g., GenAI and immersive technologies) alone does not guarantee valid learning or assessment; rather, its impact depends on how lecturers adapt and integrate it in ways that align with program intended learning outcomes and goals, thereby supporting **content validity**. When used thoughtfully, technology can also enhance **construct validity** by shaping the ways in which students demonstrate knowledge, skills, and character strengths through meaningful interactions. Because educational change is typically gradual, improvements in teaching, learning, and assessment should be *evolutionary* rather than *revolutionary*. For this reason, attention to **consequential validity** is essential, as the long-term effects of these innovations on students' learning and development only emerge over time (Corbin et al., 2026; Curtis, 2025; Dawson et al., 2024; Maney, 1999).

Consequential validity refers to a meta-level of validity that goes beyond the internal alignment and technical quality of an assessment. It focuses on whether the intended purposes of the assessment are actually achieved through the way the assessment functions within the broader educational or social system. In other words, it concerns the social impact and consequences of using the test, including whether it leads to appropriate, fair, and beneficial outcomes or produces unintended harmful and negative effects (Corbin et al., 2026; Curtis, 2025; Dawson et al., 2024; Maney, 1999).

The purpose of this book is to clearly demonstrate how, for a range of popular types of assessments, a designer can meet the demands of transparency, reliability, and validity. The higher the stakes of a decision, the greater the need to prove that the corresponding assessment is fit for purpose and will lead to decisions that will be considered defensible by all stakeholders.

Section 1

Introductory Chapters

Ya-Ping (Amy) Hsiao, Brenda den Oudsten, and Cor Sluiter

1.1

Course-Level Assessment Design: Guidance for Course Coordinators

Course Assessment Design as Part of Program Assessment Design

Course assessment design does not stand alone; it is embedded within a coherent program assessment design, which itself is an integral component of the overall curriculum. At the program level, an assessment program, defined as the integrated system that ensures all **assessments** across a program work together to validly, reliably, and meaningfully evaluate student learning (Charlton & Newsham-West, 2023; Hatfield, 2009), should be aligned with program intended **learning outcomes** and with the program's didactic approach (e.g., conceptual, experiential, or problem-based learning). This alignment ensures that assessment practices consistently reflect what the program values and aims to develop in graduates.

Within this broader framework, course assessment design operationalizes program intentions at the course level. In a constructively aligned curriculum, program intended learning outcomes explicitly describe what students are expected to know, understand, and be able to do by the end of a course or program. Course assessments are then directly aligned with these outcomes as well as with the program's exit qualifications. At the same time, teaching and learning activities are purposefully designed to prepare students to meet these expectations. This alignment ensures that assessment meaningfully measures the extent to which students have achieved the program intended learning outcomes, while also making expectations transparent and learning processes supportive.

The principle underpinning this coherence is **constructive alignment** (Biggs & Tang, 2011). It emphasizes the deliberate alignment of three key elements: learning outcomes, teaching and learning activities, and assessment. The “constructive” dimension highlights that students actively construct meaning through both cognitive and social engagement, drawing on principles of social constructivism. This means that effective assessment design must not only test knowledge acquisition but also support meaningful learning processes and interactions.

In recent years, the social dimension of learning has received increased attention due to the influence of social media and the growth of large, marketized classes in higher education (Biesta, 2015; Connell, 2019). Both cognitive and social engagement highlight that effective assessment design should ensure assessments are purposeful and meaningful, so that students are motivated to demonstrate their achievement of the program intended learning outcomes or goals.

A Good Assessment as Part of Course Assessment Design

A key feature of good course assessment design is the balanced integration of formative and **summative assessment**. These functions should not be treated as separate, but as complementary components of a coherent system, guided by an evaluative function that ensures quality and alignment. **Formative assessment** provides ongoing feedback that informs both teaching and learning, while summative assessment supports judgements about achievement.

Crucially, evidence from formative assessment must be used systematically, both to inform teaching and learning and to support collective judgements, such as evaluating pass rates, enabling adjustments that demonstrably improve student learning beyond what would occur otherwise (Black & Wiliam, 2009; Wyatt-Smith et al., 2024). In this way, good assessment is not an isolated activity but an integral part of course design, fully aligned with learning outcomes and teaching activities.

Balance between Formative and Summative Functions

Summative assessment and feedback alone, even when aligned with course learning goals, are not always sufficient to fully capture what students can achieve. When students are not monitored or do not receive feedback during the learning process, important opportunities for learning may be lost. This concern is particularly relevant in the age of generative artificial intelligence (GenAI), especially when students are permitted to use GenAI in coursework but lack sufficient AI literacy to do so effectively.

In the absence of structured formative assessment and feedback, students may rely on last minute preparation and adopt surface approaches to learning. As a result, they engage only minimally with course content and fail to develop deeper cognitive and social understanding. Moreover, summative feedback that accompanies a grade often arrives too late to address learning deficiencies or to inform subsequent learning. Research also shows that students frequently focus on grades without engaging with qualitative feedback comments (Mensink & King, 2020). This limitation is further compounded when courses are not clearly connected within a coherent program level learning trajectory, which is often the case.

At the same time, formative assessment and feedback alone are also insufficient. While they support learning by providing ongoing insights into student progress, they do not offer a structured judgement of whether learning outcomes have been achieved at the required standard. Without a summative component, there is no formal basis for certification, progression decisions, or accountability.

For this reason, formative and summative assessment should *not* be treated as separate or competing functions, but as complementary components of a coherent assessment system. Formative processes provide continuous feedback that supports learning and development, while summative assessment validates achievement against program intended learning outcomes. When intentionally integrated, these functions reinforce one another (Price et al., 2011). Formative assessment prepares students for success, and summative assessment confirms the extent to which that success has been achieved. See Chapter 1.6 for further details on how formative assessment can be integrated into course assessment design.

Although this book focuses on course level assessment design, such integration must be understood within the broader program context. Follow your institutional and program policy on the following topics:

- **Formulating course learning goal(s) (CLGs).** For example, some programs at Tilburg University use a three pillar approach, formulating CLGs in terms of knowledge, skills, and character development. In contrast, one school adopts an integrated approach in which these three pillars are combined within the CLGs to reflect their interdependent nature, meaning they are not assessed in isolation. It is important to align CLGs with program-level learning outcomes and to comply with the standard number of CLGs (e.g., 3-5).
- **Stating the CLGs in the specified platform and document.** Clearly state the CLGs, assessment types, and their corresponding weightings (percentages) in the official course catalog and syllabus, and ensure these remain consistent throughout the course. This information enables students to anticipate overall assessment workload and make informed decisions based on their personal context (e.g., selecting electives, balancing part-time work). Any changes should follow institutional procedures.
- **Use acceptable assessment types.** Confirm that your chosen assessment formats are permitted. For instance, take-home exams or oral exams may be prohibited in certain programs.

- **Communicating deadlines.** Review your institutional policy regarding the deadline for communicating detailed assessment information to students, whether it must be shared before the course begins or by a specific session (e.g., the first, second, or third session).
- **Ensuring complete information.** Ensure that all required assessment information is included in the syllabus (either on the Learning Management System or as a downloadable document), in accordance with your institutional guidelines. This should include study load based on the number of European Credit Transfer and Accumulation System (ECTS, 1 ECTS $\approx$ 25–30 hours of work), assessment workload, grading policies, types of exam questions, policies on **correction for guessing** and grade calculations, descriptions of assignments, assessment criteria or **rubrics**, and permitted uses of GenAI tools. For assignments, consider the overall workload students face, especially when they are working on multiple individual or group assignments across different courses. In the case of groupwork, consider the additional time and effort required for coordination and collaboration among group members.

Course coordinators should therefore ensure that course assessments are aligned with the overall program and guided by key assessment quality criteria such as **transparency**, **reliability**, and **validity**, as outlined below.

Transparency

Transparency refers to the degree to which all necessary information about an assessment has been provided to both students and examiners (Hattie, 2015). In line with the vision of student-centered education, transparency is positioned as a prerequisite for ensuring reliability and validity (Bamber, 2015; Bloxham, 2012; Tierney, 2014). Without a clear understanding of what is expected of students, it is not possible to meaningfully address the other two criteria. Recent discussions on improving transparency emphasize engaging students with lecturer(s) expectations regarding workload, criteria, and standards through a more student-centered approach that foregrounds dialogic interaction, enabling students to understand and internalize these expectations (Balloo et al., 2018; Bearman & Ajjawi, 2018).

Reliability

Reliability refers to the sufficiency and consistency of information (Brookhart, 2003; Förster & Rojas-Barahona, 2023; Smith, 2003) collected within a reasonable time frame and derived from valid and appropriate assessment tasks. It implies that enough evidence is gathered to support stable and dependable judgments about student performance. At the same time, reliability concerns the extent to which measurement error and bias are minimized, including variability arising from examiner(s), tasks, or contextual conditions. This requires clear criteria, consistent application of standards, and careful design of assessment conditions to ensure that outcomes are not unduly influenced by subjective or situational factors.

- Selected assessment types must provide trustworthy and sufficient evidence of student achievement (summative) and/or development (formative). Balance both functions of assessment.
- Ensure that students are given sufficient time to complete exams and assignments. This is perhaps what matters most to students in terms of reliability. For example, assess whether 40 complex calculation questions can be completed within a 120-minute exam, but keep in mind that in order to offer a reliable exam, a certain number of **items** must be present.
- Provide concise grading instructions and assessment criteria and standards. Describe as concretely as possible when a specific major grade level should be awarded (e.g., excellent, good, sufficient/pass and insufficient/fail).

Validity

In course-level assessment, validity refers to how well assessments measure the CLGs. This includes whether the chosen methods or types fit the course's role in the program and whether tasks are relevant, meaningful, and representative for demonstrating achievement of the CLGs.

- CLGs serve as the foundation for assessment design. However, the formulation of these goals or which taxonomy is used (e.g., Bloom's taxonomy, Miller's pyramid) may vary across universities, institutions, and programs.
- Assessments should appropriately measure to what extent students have achieved the intended CLGs and the corresponding relevant program learning outcomes. For example, a learning goal assessing lower knowledge and comprehension cannot contribute to the program learning outcomes of problem-solving or communication skills.
- The sequencing of CLGs plays a critical role in shaping both instructional activities and assessment design (see Table 1.1.1 for example). A logical sequence ensures that assessments align with the progression of learning and support coherent skill development. In addition, there should be a justified rationale for establishing the weight of each assessment in deciding on successful completion of the program.
- How well assessments measure the CLGs depends on the internal quality assurance policy. Coverage means all CLGs are assessed, though not always graded (e.g., affective outcomes like attitudes). When using a taxonomy, assessments should target the highest levels of the CLGs. Coverage does not mean assessing all content topics and all levels within CLGs; practical constraints such as limited exam time, depth requirements, and allowing student choice make this neither feasible nor necessary.

Table 1.1.1

Sequencing Methods for CLGs (Florida State University, 2002) and Its Impact on Assessment Design

Sequencing Method	Definition	Impact on Assessment Design
Job performance order	CLGs are taught in the same order as real-world job tasks.	Use task-based or simulation assessments that mirror work contexts. Weight assessments based on task complexity or job significance.
Chronological order	CLGs follow the natural order of events over time.	Design assessments that reflect the sequence of events. Weight may increase over time to reflect accumulating understanding.
Critical sequence	CLGs are prioritized and taught by their relative importance.	Major assessments should focus on high-priority CLGs. Assign greater weight to assessments linked to critical content.
Simple to complex	CLGs increase in difficulty over the course.	Begin with low-weight assessments. Increase weighting as CLGs become more complex to reflect deeper learning.
Comparative sequence	Familiar concepts are introduced before unfamiliar ones.	Early assessments may be lower-weighted and diagnostic. Later assessments involving unfamiliar material should carry more weight.
Relationship-based	CLGs are ordered based on dependency or support relationships.	Use scaffolding in assessments; ensure mastery of earlier CLGs before evaluating integrative tasks Weight earlier assessments moderately to ensure foundational understanding; increase weighting as integration and synthesis are required.
Part to whole	Components (parts) are mastered before addressing comprehensive tasks (whole).	Assess sub-skills or discrete CLGs early; culminate in integrative assessments or capstone tasks. Assign lower weight to each assessment; place higher weight on final integrative or holistic tasks.

Course Assessment Patterns and Grading Policies

As observed at Tilburg University, most courses use multiple assessment components to determine final grades. However, in the early stages of bachelor's programs (i.e., the first semester of the first year), some block courses rely entirely on a final exam due to their shorter duration. In such cases, it is important to include formative assessment and feedback to support students' progress towards the final exam. At this stage, relatively high dropout rates can also make graded groupwork less feasible, as groups may become unbalanced or, in some cases, reduced to very small numbers or individual students.

When formulating course grading policies, consider the following questions:

- Is there a balanced weighting between exams and coursework, and between individual and groupwork?
- In principle, each graded assessment should offer a resit opportunity. Is the resit comparable to the original assessment (i.e., assessing the same course learning goals)?
- Does the weighting of unsupervised or group-based coursework remain within program limits?
- Are first-year students protected from unsuitable assessment types (e.g., graded groupwork in semester 1) or overly complex assessment formats (e.g., challenge-based projects)?
- Are assessment criteria, rubrics, and grading scales applied consistently across assessments?
- Are bonus points aligned with the course learning goals?
- Are attendance or participation requirements educationally justified and not treated as assessment types?
- Do all grading and rounding procedures comply with institutional regulations (e.g., the Education and Examination Regulations)?
- Are all partial grades and bonus components properly recorded in the grade administration system?
- When correction for guessing is applied, is it explained in an understandable way (e.g., with examples) so that students clearly understand how their grade is calculated? Consult Chapter 1.5 for more information.
- Are the assessment and grading policies fair, transparent, and feasible for both students and staff?

At the program level, establishing a set of course grading policies is essential to ensure consistency and promote transparency in students' understanding of assessment and grading expectations.

Box 1.1.1

Example Program-level Grading Policies for a bachelor's Degree

- In the first semester of the first year of the program, no graded groupwork is permitted. However, formative groupwork is encouraged to support collaborative learning.
- Any invigilated (proctored) exam must include a comparable resit opportunity. The resit may take the form of an equivalent exam or a combined assessment covering overlapping learning goals. In all cases, the most recent attempt determines the final grade.
- Unsupervised or group-based coursework should not exceed 40% of the total course grade, unless the course learning outcomes can only be appropriately assessed through this method.
- Any graded assessment contributing more than 10% of the final course grade must include a resit opportunity, unless this is not feasible due to the nature of the task (e.g., process-based assessments).
- The resit for unsupervised coursework is capped at a maximum grade of 6.0.
- To ensure consistency, all rubric-based coursework assessments use a four-level grading format, converted into a numerical scale (1–10). Use the assessment criteria with relevant program learning outcome categories as stated in your program document.

- Do not state bonus points as an assessment type. State whether it is part of a graded assessment type or as a separate assessment type. They must be aligned with course learning goals and may contribute a maximum of 5% of the final grade. Bonus points cannot be used to convert a failing grade into a passing grade.
- Attendance or participation is not an assessment type. Mandatory attendance is only permitted when specific course learning goal(s) can only be assessed in-class or in person.
- Apply the formula for correction for guessing to exams with multiple-choice questions.
- Rounding rules must comply with the Education and Examination Regulations (EER).
- All partial grades for summative assessments, including any bonus components, must be recorded in the grade administration system.

Three Rules

When combining multiple assessments to determine a course grade, consider the following three rules (Chester, 2003) to determine a course grade:

- A conjunctive rule sets a passing grade on each of the assessments, and it requires students to pass each assessment to pass the course. Be aware of your institutional rounding rules.
- A compensatory rule does not set a passing grade on each of the multiple assessments. It aggregates grades from the assessments and thus it allows a higher performance on one assessment to counterbalance a lower grade on another one.

It is important to note that individual and groupwork cannot compensate each other. Individual students should be able to pass a course based on their own performance. When students with poor individual grades compensated by the group grades pass the course, we cannot be sure whether they have attained the course learning goals at a sufficient level.

- A complementary rule requires students to meet minimum requirements on one or more of the assessments. Students may pass one or several assessments, provided that all CLGs are assessed and each CLG is demonstrated at a satisfactory level.

In general, there are three types of course assessment designs: different assessments of different course learning goals, different assessments of the same course learning goals at the same cognitive levels, and different assessments of the same course learning goals at different cognitive levels. The following examples show how these rules are applied in the design of multiple assessments. For a more complicated course assessment design, please consult the assessment specialist.

Distinct Course Learning Goals Across Assessments

The most appropriate way to combine different assessments of *different* course learning goals is to use a *conjunctive* rule because we need to ensure that students have a sufficient level on all course learning goals. Table 1.1.2 shows that, to pass the course, students will need to pass both the team project (40%) and the final written exam (60%). For an insufficient grade, students resit each assessment separately.

Table 1.1.2*Course Assessment Overview*

Course Learning Goals <i>On successful completion of this course, students will be able to:</i>	Program learning outcomes ⁴	Team project (40%) ☒ Pass required	Exam (60%) ☒ Pass required	Formative assessment (no grade, Feedback only)
1. Identify key psychological concepts relevant to marketing (e.g., cognition, motivation, bias).	1		x	
2. Identify entrepreneurial marketing strategies (e.g., segmentation, value proposition, branding).	1		x	
3. Apply psychology and marketing concepts to ventures, companies, or current events.	2		x	
4. Apply adaptive experimentation (e.g., A/B testing, iteration) in marketing contexts.	3	x		Draft proposal & peer review
5. Collaborate to synthesize insights from multiple sources to address marketing challenges.	3	x		Draft report & peer review

A team project may consist of smaller assignments, such as a project proposal (10%), a project report (20%), and a presentation (10%). Since these small assignments assess the same course learning goals, it is not necessary to determine pass-or-fail on each assignment (see the compensatory rule).

Overlapping Course Learning Goals Across Assessments

When multiple assessments test the same course learning goals, a *compensatory* rule is appropriate (see Table 1.1.3). This pattern often takes place in skills-focused courses. Under this approach, results are combined into a weighted total or average to determine the final course grade. If this aggregated score meets or exceeds the **cut-off score**, the student passes the course. If not, the student may take a resit that covers the full scope of both the group report and presentation. This approach is suitable when there is substantial overlap between assessments in terms of both learning goals and cognitive demands. ⁴

Table 1.1.3*Course Assessment Overview*

Course Learning Goals <i>On successful completion of this course, students will be able to:</i>	Program learning outcomes ⁴	Group report (40%) ☐ Pass required	Group presentation (60%) ☐ Pass required	Formative assessment (no grade, Feedback only)
1. Execute key stages of product development to advance a concept toward implementation.	3	X	X	
2. Apply structured methods to generate, test, and refine solutions using user and market feedback.	3	X	X	
3. Evaluate business models and go-to-market options for new products or services.	2	X	X	
4. Communicate outcomes professionally using appropriate academic and business conventions.	6	X	X	Draft Interview protocol & peer review
5. Reflect on teamwork, incorporating self- and peer feedback for personal development.	4, 7	X	X	

Lecturers can use continuous low-stakes assessments to progressively develop specific skills, as illustrated in Table 1.1.4. In this design, a *complementary* rule is used to determine the course grade. The in-class individual R assignments

⁴ 1: Knowledge; 2: Applying knowledge; 3: Problem solving

are graded and they count for 10%, while there is no pass-or-fail requirement. For example, a student can be awarded a 4 (so the assignments are graded), but 4 counts for 10%. The reason why these assignments do not have a pass-or-fail requirement is because course learning goals (CLG 1, 2, and 4) are also assessed in the final exam. The final exam, however, does have a pass-or-fail requirement, and students need to resit it if they receive an insufficient grade. Once a passing grade on the final exam is achieved, the overall course grade is determined by combining the results of both the assignments and the final exam using a weighted total or average.

Table 1.1.4*Course Assessment Overview*

Course Learning Goals <i>On successful completion of this course, students will be able to:</i>	Intended learning outcomes	In-class individual R- assignments (10%) ☐ Pass required	Final exam (90%) ☒ Pass required	Formative assessment (no grade, Feedback only)
1. Apply core Microsoft Excel techniques for analyzing tabular data (e.g., pivot tables, lookup functions).	2	X	X	Peer review
2. Perform calculations using different R data structures.	2	X	X	Peer review
3. Clean and transform data in R (e.g., merging, reshaping, aggregating).	2, 5		X	
4. Write R code to solve basic optimization problems.	2	X	X	Peer review
5. Create data visualizations using R (e.g., ggplot2).	5		X	

Other Issues

Academic Integrity

Lecturers should treat academic integrity as an integral part of teaching and learning, rather than solely as rule enforcement, and actively support students in developing sound ethical academic practices. It should be embedded across teaching, supervision, and assessment by explicitly teaching skills such as citation, paraphrasing, and responsible use of data, and by designing learning activities that systematically develop these competencies. Integrity issues should be addressed early through formative dialogue focused on prevention and student development, while ensuring that institutional procedures are followed when misconduct occurs.

Attendance and participation

Attendance and participation are often used interchangeably by lecturers. However, attending lectures does not necessarily imply active student participation. As discussed in the Introduction chapter, grades should be based on valid and reliable evidence obtained through summative assessment. Since attendance is just a proxy for active participation and achievement, attendance alone should not be used as a basis for grading.

Participation, in contrast, generally refers to students' active engagement in learning activities, such as contributing to discussions, asking and answering questions, or collaborating on case-based tasks in small groups. This form of engagement may be influenced by individual differences in student personality, such as extraversion and introversion, which can affect how comfortable students feel speaking in group settings. Participation should only be graded when all students have equitable opportunities to demonstrate their thinking and engagement with the course content through these activities. For example, students should not be disadvantaged if they are unable to physically attend every lecture (Haladyna, 2019).

Bonus Points

Bonus points are assessment results, not a separate assessment type. Therefore, the awarding of bonus points must meet the same quality criteria as other assessments: they should be valid (aligned with course learning goals), reliable (based on clear evidence, deliverables, and grading instructions), and transparent (clearly described in the course syllabus).

Bonus points should only be used to reward excellent performance. This means that the associated assessment tasks should target the highest cognitive levels of the intended course learning goal(s). The contribution of bonus points may not exceed 5% of the final course grade and cannot be used to convert an insufficient final grade into a passing grade.

We also advise against awarding bonus points for class attendance alone.

Although assessment literature strongly discourages the use of bonus points, they can be considered when a new coursework assignment is introduced for the first time in a course design. For example, to mitigate the impact of GenAI, some lecturer(s) are experimenting with integrating GenAI into coursework assignments (Leenheer et al., 2026). Because these innovative assignments require new assessment criteria, it may not yet be possible to fully validate the criteria against the intended course and assignment learning outcomes. In such cases, lecturer(s) may use a pass-or-fail approach, where a passing performance earns a small bonus (e.g., 0.5 points), while a failing performance results in no bonus points.

Deduction Points

With multiple-choice exams, a scoring rule could be used according to which a correct answer earns 1 score point, an unanswered item 0 score points and for each item answered incorrectly 1 score point is deducted. The psychometric literature makes clear that using this rule significantly lowers the reliability of the assessment. We therefore advise avoiding working with deduction points.

Inspection

Students have a right to inspect their work. The institution has the right to set conditions under which this is done. The inspection is for students to review their responses in written examinations or coursework as well as in audio/video recording of oral assessments. The inspection is not a negotiation moment regarding grades for students. The discussion should be aimed at clarifying errors or omissions in the students' answers. Students can only ask for corrections on administrative errors (e.g., when a correct answer is not rewarded with one or more score points) based on a comparison between their own answer and the answer provided by the examiner during the inspection.

Re-Using Exam Questions

Writing quality selected-response questions is hard work. When lecturers have quality selected-response (SR) questions with satisfactory psychometric characteristics, they might want to reuse them. Consider the following caveats before reusing questions.

Reusing exam questions may pose a threat to validity (Joncas et al., 2018; Wagner-Menghin et al., 2013). Exam questions that require students to apply knowledge (and beyond) on their first use might require only recalling the correct answers when they are reused (i.e., exam questions do not assess what they are intended to assess). This might not be a problem for exam questions assessed at the knowledge level.

In addition, reusing exam questions within a short time span may potentially reduce the **item difficulty** level, especially for the questions intended to assess higher cognitive levels (Wagner-Menghin et al., 2013). Consequently, students may have a higher grade on an exam with a substantial proportion of reused questions due to a declined item difficulty level. The exam results could overestimate the actual performance level of those who have studied the reused questions (and some students regard this as unfair) (Gehring, 2004; Joncas et al., 2018).

When reusing exam questions in general, keep the following suggestions in mind (Gehring, 2004; Joncas et al., 2018; Wagner-Menghin et al., 2013):

- Avoid using a question within a short time span (e.g., 5 years).
- Consider the risk that re-used exam questions may no longer assess the intended cognitive levels.
 - Knowledge – Low risk: Reuse generally does not affect assessment, as these questions test simple recall of facts.
 - Comprehension – Moderate risk: Reuse may reduce depth, since students could rely on memorized explanations rather than demonstrating understanding.
 - Application and higher-order – High risk: Reuse can turn problem-solving or analytical questions into simple recall, undermining the assessment of intended skills.
- Consult the **item analysis** report to determine whether an item can be re-used or altered. For example, a difficult item can be re-used and adapted or an item in which the **distractors** were not functional can be removed.
- *Partially* re-use a question, for example, by changing some of the incorrect answer options, or by reversing the formulation, for example “which of the following is an advantage of focus groups, in comparison with single-person interviews?” can be flipped into “which of the following is a disadvantage of single-person interviews, in comparison with focus groups?”.
- Re-use or partially re-use only a low ratio (<5%) of randomly selected questions from a small item bank.

Use of GenAI

Follow your institutional guidelines (e.g., the AI Index or two-/three-lane approaches) to guide students in the responsible use of GenAI for coursework and exam preparation. This involves not only clearly stating in the course syllabus whether and to what extent AI tools may be used, but, more importantly, engaging students in ongoing discussions about AI use through course activities and formative dialogue. In addition, clarify the level of detail expected in AI disclosure statements. Both overly generic and excessively detailed disclosure requirements may unintentionally foster mistrust between lecturers and students. Finding the right balance is only possible when lecturers and students engage in accountable dialogue about responsible and transparent AI use.

Accountable talk refers to structured and respectful dialogue in which participants justify their ideas, respond to the reasoning of others, use evidence to support their claims, and collaboratively build understanding (Resnick et al., 2018). In the context of AI use in education, accountable talk encourages both lecturers and students to openly discuss expectations, ethical considerations, appropriate levels of AI assistance, and the rationale behind disclosure practices.

Checklist Course-Level Assessment Design

The following checklist summarizes key recommendations discussed in this chapter. It is intended as a practical tool to support the design and review of your course-level assessment design.

Follow your institutional guidelines

Formulating Course-Level Learning Goals (CLGs)

- CLGs are formulated at course level rather than lecture level.
- CLGs align with program learning outcomes.
- CLGs follow institutional model (e.g., knowledge–skills–character OR integrated approach). For example, if using pillars, they are either clearly separated or meaningfully integrated.
- CLGs are not overloaded (typically 3–5 per course).
- CLGs reflect appropriate cognitive level for the course status in the program.

Stating CLGs and Assessment Information

- CLGs, assessment types, and weighting (percentages) of each assessment type are clearly stated in the official course catalog and syllabus.
- Information is consistent across all platforms (Osiris, syllabus, learning management system).
- Any changes follow formal institutional procedures.

Use of Approved Assessment Types

- Assessment types comply with institutional regulations.
- Any un-listed innovative types (e.g., group debates) are verified for permission.

Completeness of Assessment Information in Syllabus

- Information is shared before course start or within required sessions (as policy states).
- Study load is clearly specified (ECTS workload estimate included).
- Students are informed early enough to plan workload and commitments.
- Assessment workload is realistic and distributed appropriately.
- Overall student workload across multiple coursework assignments is considered.
- Groupwork workload from coordination and collaboration effort is considered.
- Grading policies are clearly described.
- Permitted use of GenAI tools is clearly stated.

Transparency

Ensure students and staff clearly understand expectations, procedures, and grading.

Assessment Information

- Course documents clearly describe each assessment type.
- Final grade calculation is clearly explained.
- Pass/fail rules and resit possibilities are clearly stated.
- If applicable, compensation rules are clearly communicated.
- Grade conversion rules (from raw scores to a number grade 1-10) are clearly stated.
- Resit procedures are clearly described (what, when, how).
- Inspection (NL: inzage) information is provided.
- Groupwork rules and free-riding procedures are clearly explained.
- All necessary instructions are included (format, length, submission, deadlines).

Reliability

- Ensuring assessment results are stable, consistent, and fair.
- Four-eyes principle is applied (peer review of assessments).
- Assessment tasks provide sufficient and appropriate evidence of student achievement.
- Scoring system ensures consistent grading across examiner(s).
- Use clear and shared criteria and consistent application of standards.
- Time and workload are appropriate for the complexity and number of tasks.
- Test and item analysis is conducted (if multiple-choice items are used).

Validity

Ensuring assessments measure what they are intended to measure.

Course Learning Goals (CLGs)

- CLGs are student-centered and clearly formulated.
- CLGs are concrete, specific, and use measurable verbs.
- CLGs are appropriate for BA/MA level and course year.

Alignment with CLGs

- Assessment types appropriately match CLGs.
- All CLGs are assessed in at least one assessment type.
- Highest cognitive level of each CLG is assessed.
- Assessment tasks are representative of content and skills in the CLGs.
- **Construct validity** is ensured (thinking skills in the CLGs are assessed).
- **Content validity** is ensured (content topics in the CLGs are representatively covered).
- Criteria are aligned with CLGs and tasks.

Course Assessment Overview Quality

- Mapping between CLGs and program learning outcomes is appropriate.
- Mapping between CLGs and assessment types is explicit.
- Weighting across assessment types is logical and defensible.
- Whether a pass is required is indicated on each assessment type.
- Formative assessment is relevant to CLGs and meaningful for students to use feedback improve their performance on subsequent assessments.

1.2

The Assessment Cycle

An assessment cycle, see Figure 1.2.1, provides a structured framework for designing, implementing, and continuously improving assessment practices within academic programs at the level of individual **assessments**. The core of the cycle is formed by six interconnected steps (Van Berkel et al., 2023) that each correspond to a phase in a Plan–Do–Check–Act (PDCA) cycle, emphasizing continuous improvement.

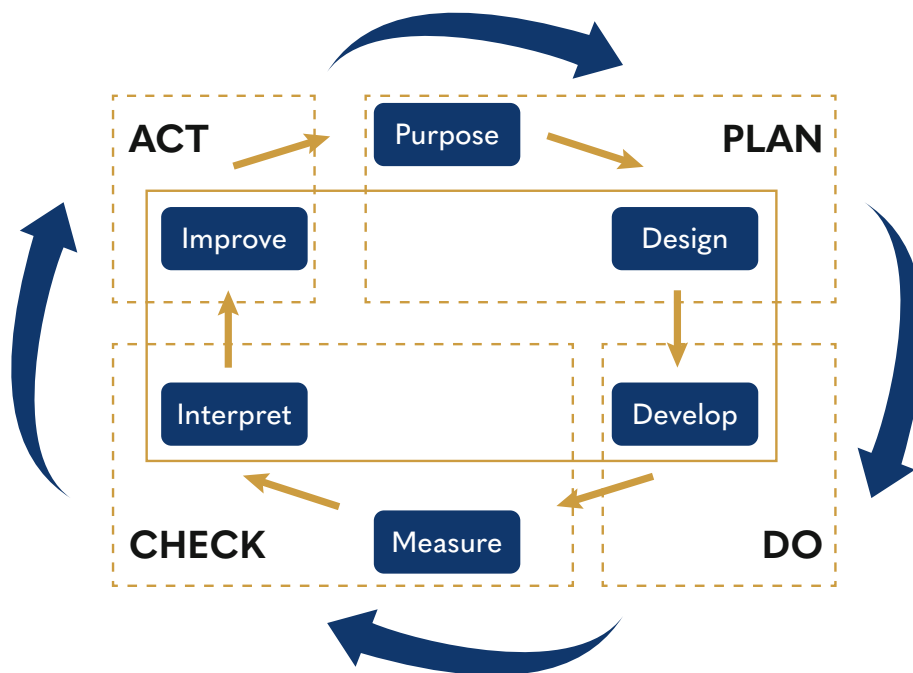


Figure 1.2.1
The Assessment Cycle Based on the PDCA

In the PLAN phase, the concrete *purpose* of the assessment must be clearly specified. This is decisive for all subsequent phases of the cycle. Three distinct purposes can be distinguished, each linked to specific inferences and decisions. **Diagnostic assessment** identifies students' prior knowledge, misconceptions and needs to guide instruction. It is prospective: it looks forward by identifying what should be taught next. **Formative assessment** uses evidence of learning to adjust teaching and to provide feedback to students. It is interactive: it looks inward to the ongoing learning process. **Summative assessment** involves making evaluative judgments about achievement at a particular moment, for grading, certification, placement, or accountability. It is retrospective: it looks back to determine the level of mastery attained.

Following a learning cycle, diagnostic assessment is typically used at the start; formative assessment during learning and summative assessment at the end to evaluate the effects of instruction.

Before course assessment *design*, consult the institutional assessment policy and the program assessment plan for acceptable assessment types determined for assessing each category of program learning outcomes. You are expected to select an assessment type that aligns with your course context and learning goals. Furthermore, develop an accountability table (e.g., a course assessment overview, table of specifications, test matrix, or another rationale for the assessment content) to clarify the assessment content and align it with the course learning goals.

Design also includes setting a preliminary standard (i.e., the **passing score**) and determining how assessment scores will be converted into course grades. Finally, decide on the implementation conditions, including whether the assessment will be paper-based or digital, what resources may be used during the assessment, and what support is permitted for coursework.

At the DO phase, *develop* assessment tasks or exam **items** in accordance with the established accountability table. Provide clear scoring instructions (e.g., model answers, answer keys, and point allocations) and appropriate instruments (e.g., criteria, **rubrics**, or checklists) to ensure consistent and transparent grading. Equally important are explicit instructions for students, outlining expectations, requirements, deadlines, and any relevant guidelines or resources to support their performance.

During the CHECK phase, conduct a pre-check of the developed assessment materials based on the four-eyes principle. This phase also includes the necessary steps for *measurement*, such as administering the assessments, scoring student work, and analyzing student responses using test and **item analyses** or by examining score consistency between different examiners to determine score **reliability**. Finally, based on these analyses, interpret student responses to identify poorly functioning assessment tasks or exam items and, where permitted by institutional policy, adjust the preliminary passing score. This post-check provides you with feedback, enabling you to examine your own assumptions about student ability and assessment difficulty for the next iteration.

At the ACT phase, the focus is on *improvement*. The analysis results can be used to *improve* future assessment quality. Interpret student responses to evaluate whether the collected evidence aligns with the course learning goals and the meaning of grades, and to determine the extent to which the course learning goals have been achieved. Based on these insights, adjust the course learning goals, instructional activities, and assessments in the next iteration of the cycle.

1.3

Role as Examiner and Essential Rules from Various Policy Documents

Examiners are officially appointed by the Examination Board. Their role and responsibilities are often outlined in the Rules and Guidelines.

While several documents are available to guide you in becoming a good examiner, we recognize the challenge of reviewing multiple documents alongside your busy work schedule (i.e., Education and Examination Regulations/**EER**, Rules and Guidelines/**RG**, Assessment Policy/**AP**, Practical Guide (this book)/**PG**). To make this process easier, we have compiled a list of the top ten questions you should understand before conducting any **assessment**, along with the relevant documents where you can find the answers, based on the context of Tilburg University.

1. Is it compulsory to offer a resit for each graded assessment? **EER**
2. If certain assessment components must be completed during tutorials, do I still need to specify a compulsory attendance requirement (e.g., 80% attendance)? **EER**
3. Do all assessments require a minimum grade to pass the course? Is the minimum grade always 5.5, or do I have the flexibility to determine it myself? **EER/AP/PG**
4. Is **correction for guessing** mandatory⁵ for multiple-choice exams? Which option should I choose? Also, what grading method should I apply for grade calculations—should I always use a split-linear transformation, and why? **RG/PG**
5. What grading policies and assessment information must we communicate to students in a student administration system (SAS) (e.g., Osiris), the course syllabus, and the learning management system (LMS) (e.g., Canvas)? **AP/PG**
6. Do we need to inform students about consulting the EER (e.g., the rule that the last attempt counts), or can we assume they will read it when needed? **AP/PG**
7. Is the four-eyes principle mandatory? How do I implement it to ensure compliance with this requirement? **EER/RG/AP**
8. How should we communicate grades and the possibility for grade inspection to students after each attempt? Do we need to offer inspection opportunities after a resit? Should we send grades to the Examination Organization first, or communicate them via the LMS? If via the LMS, what is the best method? What is the proper procedure for conducting an inspection (e.g., what can students bring and take from the inspection)? **EER/RG**
9. Do we need to submit partial grades to the Examination Organization for registration in the SAS, or only the final grades? **EER/RG**
10. Am I responsible for detecting fraudulent behavior in coursework? How should I communicate any findings to students, and what is the process for reporting them to the Examination Board? **EER/RG**
11. What if a student contacts you with special circumstances and asks for an exemption related to attendance, assignments, exam? **EER/RG**

5 RG: Can only state that correction for guessing should be applied.

1.4

Estimating Written Exam Time

Introduction and Learning Goals

This chapter focuses on estimating exam time. However, exam design should always be situated within the broader framework of assessment quality. A well-designed exam meets key criteria of **transparency** (e.g., clear information about available time and guidance on time management), **reliability** (i.e., a sufficient number of **items** to ensure adequate evidence of student performance while also allowing sufficient time for completion), and **validity** (i.e., exam duration should allow adequate coverage of the course learning goals). For further guidance on these aspects, consult your institution's assessment policy or handbook.

After applying the instructions of this chapter to their own assessment practices, lecturers are able to:

- Identify factors that influence exam duration, such as question type, complexity, and cognitive demand.
- Apply methods to estimate the time students need to complete exam questions.
- Evaluate and adjust exam design to ensure that the allocated time supports fair and effective **assessment**.

Ensure Sufficient Exam Time

One pre-requisite to ensure exam reliability is whether the exam time is sufficiently long for all or almost all students to complete the exam so that they can demonstrate their thinking skills of the course content (Traub & Rowley, 1991). This means it is essential to determine the number of questions to fit the time frame (e.g., two-hour exams) and to estimate exam time carefully at the same time. This guide only focuses on how to estimate exam time.

When estimating response time, consider:

- Task type – whether the question is a selected-response (see Table 1.4.1) or a constructed-response (see Table 1.4.2).
- Reading time – influenced by the length and complexity of question stems and answer alternatives.
- Thinking and drafting time – depends on the cognitive complexity, varying between simple recall and higher-order thinking skills (application, analysis, evaluation). Drafting time depends on the expected length of the response and the mode of production (handwriting or typing).
- Review time – include additional minutes for students' final checking of answers (referred to as “last answer time” for example, in TestVision).

When estimating response time for complex or calculation-based items, consider the following additional aspects:

- **Consider time to execute mechanical (arithmetic, algebraic, or statistical multi-steps)** on top of basic response time.
- **Ensure that higher-order thinking skills or multi-step questions do not disproportionately consume exam time.** This means checking that one complex item does not take so long to solve that it reduces students' ability to attempt the rest of the exam⁶.

6 If the course learning goals also contain higher-order thinking skills, that it is not sufficient to just produce items aimed at recalling knowledge, just for the sake of having many items.

Selected-Response questions

Table 1.4.1 provides suggested evidence-informed response times for selected-response questions, taking into account their cognitive complexity, length, and format.

Table 1.4.1

Estimated Response Times by Selected-Response Question Type
(Brothen, 2012; Kennette & McGuckin, 2025; Wendt & Kenny, 2009)

Selected-Response Question Type	Response time
Knowledge & Understanding, questions with 25 to 45 words	30–45 seconds
Knowledge & Understanding, questions with 35 to 55 words	40–60 seconds
Complex (Application, Analysis, Evaluation)	70–90 seconds
With calculations	2–5 minutes
Ordered response	110–120 seconds
Multiple response	70–85 seconds
Hot spot	40–50 seconds
Chart/Exhibit	70 seconds

Constructed-Response Questions (Open-ended or essay questions)

When estimating exam time for constructed-response questions, consider both reading speed, handwriting, and typing speed.

Reading Speed

- When planning reading activities (e.g., longer cases), keep in mind that the optimal reading rate for skilled adult readers, including undergraduate students, is about 300 words per minute, if the language is the reader's native one.

This estimate applies in a normal reading environment where (Rayner et al., 2016):

- The text does not introduce unfamiliar words or complex new concepts.
- The reading goal is straightforward comprehension: understanding the meaning of each sentence as it is read. Avoid irrelevant difficulty, unnecessary linguistic complexity in wording of stem and options (Bosher, 2003).

Estimate additional time if the exam text introduces new vocabulary, abstract concepts, or requires interpretation/analysis beyond simple comprehension.

Handwriting and Typing Speed

- Handwriting:** The handwriting fluency test shows that undergraduates can write 15–17 words per minute (WPM) (Connelly et al., 2005; Summers & Catarro, 2003). Use this rate as a baseline for estimating the time students need to complete extended written responses.
- Typing:** In copy-typing tasks, a university student may produce about 31 WPM (Sperl et al., 2024). Expect speed to be closer to 20–25 WPM in real exam situations.

Table 1.4.2 outlines how to calculate exam time by considering cognitive levels, expected word counts in student responses, and the time required for reading, thinking, drafting, reviewing, and editing.

With the growing impact of GenAI, many lecturers are moving toward extended constructed-response questions. To ensure reliability, it is essential that exams are designed to provide students with sufficient time to complete these tasks.

Table 1.4.2

*Example of Estimated Response Times for Constructed-Response Questions
(in Minutes: Reading, Thinking & Drafting, Typing, Reviewing & Editing, and Total)*

Question	Levels tested in the question	Word counts expected in the answers	Reading	Thinking & Drafting	Typing	Reviewing & Editing	Total
1a	Comprehension	50	1	1	2	2	6
1b	Application	50	1	3	2	2	8
2a	Comprehension	50	1	1	2	2	6
2b	Application	50	1	3	2	2	8
3a	Comprehension	50	1	1	2	2	6
3b	Application	50	1	3	2	2	8
4a	Application	80	1	3	3	3	10
4b	Analysis	80	1	5	3	3	10
5a	Analysis	80	1	5	3	3	10
5b	Evaluation	80	1	5	3	3	10
6a	Analysis	100	1	5	4	3	11
6b	Evaluation	100	1	5	4	3	11
Total		820	12	40	32	30	114

1.5

Setting a Passing Score and Transforming Raw Scores to Grades⁷

Introduction and Learning Goals

Setting performance standards, also called norming, involves defining the level of achievement required for different performance levels, such as passing or excellent performance (Cizek & Bunch, 2007). A performance standard, or criterion level, specifies how well a student must perform to meet the required expectations. This standard can be expressed as a minimum number of correct answers (for example, 4 out of 5 **items**) or as a percentage of correct answers (for example, 80%). The score that represents this standard is called the **cut-off score**. The cut-off score separates students into categories, such as those who demonstrate mastery or competence and those who do not yet meet the required level (Berk, 1986).

This chapter does not aim to provide a comprehensive overview of the technical **standard-setting** methods used in educational measurement. Instead, it focuses on the concepts and methods that are most relevant in classroom practices within the Dutch higher education context.

After applying the instructions of this chapter to their own assessment practices, lecturers are able to:

- Explain the purpose of setting performance standards in educational **assessment**.
- Determine an appropriate **passing score** for an exam based on the institutional guideline.
- Apply methods to combine **correction for guessing** with the determination of a passing score.
- Apply methods for transforming raw exam scores into number grades.

Methods for Setting Performance Standards

Very often, the raw scores on an assessment are not on the same scale as the final grade (i.e., 1–10). It is therefore important to transform raw scores into course grades in an appropriate and consistent way. Before this transformation can take place, a pass-or-fail or cut-off score must be established; this is the lowest sufficient score corresponding to a grade of 5.5.

There are two fundamentally different approaches to determining this cut-off score. The first approach is to compare a student's test score with those of other students who have taken the same test. This is known as relative norming, as the grades students receive depend on how their performance compares to that of their peers. The second approach is absolute norming, in which test results are evaluated against a predetermined standard set by the lecturer.

⁷ This chapter is adapted from two assessment handbooks developed by the Tilburg School of Social and Behavioral Sciences (TSB) and the Tilburg School of Humanities and Digital Sciences (TSHD).

Both approaches have their advantages and disadvantages. A drawback of relative norming is that it is less suitable for small student groups and resit examinations, and it may allow students to influence outcomes through strategic behavior. In contrast, a limitation of absolute norming is that if an assessment turns out to be too difficult, a large proportion of students may fail (Sluijter & Van Berkel, 2023).

To address these limitations, a third approach known as compromise methods combines elements of both relative and absolute norming. One such method is Cohen's method (Cohen-Schotanus & Van der Vleuten, 2010), which provides a practical solution for setting pass-or-fail standards, particularly when more complex standard-setting procedures are not feasible. Rather than relying solely on a fixed cut-off score or exclusively on the distribution of student performance, Cohen's method integrates both perspectives to produce a more balanced and fairer standard. Specifically, the method uses the performance of high-achieving students as a reference point (i.e., 60% of the highest score or 60% of the 95th percentile) to determine the pass score. By anchoring the cut-off score to top performers, it adjusts for variations in test difficulty while maintaining a consistent benchmark. In this way, Cohen's method reduces the risk of setting standards that are either too lenient or too strict, offering a practical and defensible approach to grading.

A note on Setting Passing Scores

Absolute norming is commonly used in Dutch higher education as the procedure for setting a passing score at the institutional level. Except for multiple-choice tests, the pass-or-fail score can be obtained by multiplying the maximum score with the mastery percentage resulting from the absolute norming procedure.

Although frequently used in practice, it is not recommended to apply a pre-fixed percentage correct as the passing standard. Treating such a fixed percentage as a proxy for the required achievement level is a reasoning error, because it ignores the fact that the difficulty of assessments varies in practice. As Sluijter and Straetmans (2024) point out, applying a pre-fixed percentage does not take the actual difficulty of the specific assessment into account. They show that commonly used fixed percentages, such as 55%, often correspond to relatively low levels of achievement. They therefore argue that institutions should use one of the many accepted procedures for setting a defensible passing score as described in, for instance, Cizek and Bunch (2007).

However, if there are institutional policies that require working with pre-fixed mastery percentages, to mitigate the effect of an assessment that turns out to be too difficult, the Cohen method (Cohen-Schotanus & Van der Vleuten, 2010) can be used. This method is particularly useful for novice lecturers who have little experience with the design of assessment tasks or exam questions and have few observations of actual student performance. The Cohen method does not take the maximal score on an assessment as a starting point, but the highest scores that are achieved in practice. The pass-or-fail score is derived from the scores of the best performing students. The underlying idea is that these students have prepared themselves optimally, by studying well, attending lectures, etcetera. And thus, this group deserves the maximum grade of 10. To neutralize the influence of random outliers, the maximum achievable score can be set as the average score achieved by the five percent best students, the so-called 95% percentile score.

Setting Passing Scores

Except for multiple-choice tests, the pass-or-fail score can be found by multiplying the mean score of the best performing students or the highest score achieved with the pre-fixed mastery standard.

Example 1

For a 40-item test where each item is assigned one point with a pre-fixed 70% standard, the passing score S_p is 28:

$$S_p = 0,70 \times 40 = 28$$

Example 2

For a 40-item test where each item is assigned one point with a pre-fixed 70% standard and a mean score of the best performing students of 32, the passing score S_p is 23. Because the highest insufficient score is 22:

$$S_p = 0,70 \times 32 = 22,4$$

Correction for Guessing When Setting Passing Scores

Although it is quite popular in practice, a questionable procedure to correct for guessing with multiple-choice tests is based on the unrealistic idea that students will choose the right alternative by chance for items they do not know the answer to. From this invalid assumption, it follows that their raw scores should be corrected by subtracting $1/(N_a - 1)$ for each item answered incorrectly from the total score, where N_a is the number of alternatives per item. So, for instance, a score of 24 on a 40-item test with multiple-choice items with 4 alternatives, is corrected to $40 - (4 \times 16)/3 = 40 - 21 = 19$. And a score of 10 is corrected to $40 - (4 \times 30)/3 = 40 - 40 = 0$. Next, the passing score is calculated for the corrected scores. This procedure, however, negatively affects the **reliability** and **validity** of the assessment (Sanders & Verstralen, 2017) and must be avoided. The proper way to correct for guessing is to take it into account when establishing the pass-or-fail score:

Passing score (S_{pc}) with proper correction for guessing:

$$S_{pc} = cT + P(T - cT)$$

- c: chance score, the proportion of guessing;
- T: total number of items;
- P: the pre-fixed percentage of mastery standard.

With multiple-choice tests, the Cohen method can of course also be used to mitigate the effect of an assessment that turns out to be too difficult.

Passing score (S_{pc}) with proper correction for guessing, following Cohen's method:

$$S_{pc} = cT + P(B - cT)$$

- c: chance score, the proportion of guessing;
- T: total number of questions in the test;
- P: the pre-fixed percentage of mastery standard;
- B: score of the best performing student(s).

Example 3

For a 40-item, 4-alternatives multiple-choice test (and each item is assigned one point) with a pre-fixed 55% standard, the passing score with correction for guessing, S_{pc} , is 27. Because the highest insufficient score is 26:

$$S_{pc} = 0,25 \times 40 + 0,55 \times (40 - 0,25 \times 40) = 10 + 16,5 = 26,5$$

Example 4

For a 40-item, 4-alternatives multiple-choice test (and each item is assigned one point) with a pre-fixed 55% standard and a mean score of the best performing student(s) of 36, the passing score, S_{pc} , is 25.

$$S_{pc} = 0,25 \times 40 + 0,55 \times (36 - 0,25 \times 40) = 10 + 14,3 = 24,3$$

Example 5

For an exam consisting of 40-item, 4-alternatives multiple-choice questions (and each question is assigned one point) with a pre-fixed 60% standard and 8 short answer questions (all together count for 60 points), the passing score with correction for guessing is 64.

Multiple-choice part:

$$X_{pc} = 10 + (40 - 10) \times 0.60 = 28$$

Short-answer part (no correction for guessing):

$$60 \times 0.60 = 36$$

Passing score for the entire test:

$$28 + 36 = 64$$

When applying correction for guessing, provide students with the following instructions:

Multiple-choice questions ask you to select an answer from several options. It is possible that you answer a question that you do not know the answer to correctly by a lucky guess. We take this into account in the pass-or-fail score. It is therefore important that you answer *all* questions, even if you have no or only partial knowledge to answer them.

This exam consists of [] multiple-choice questions. Each question has [] options, and each question is assigned [] point (s).

The pass score corrected for guessing is [] (= 5.5 pass grade), which means, you will pass the exam if your score is [] or higher.

Transforming Scores to Grades

Despite the emergence of **formative assessment** and programmatic testing, grading, valuing a score on an assessment based on an absolute or relative frame of reference like the ones presented above, is common practice for examiners. In the Netherlands it is a legal requirement that performances on an assessment should be graded on a scale ranging from 1 to 10. This requirement was formally introduced in 1870 and then only for the final exams on the 'Hogere Burgerscholen'⁸ (Meuffels, 2004a). But it has now turned into an almost sacred framework for grading

8 'Higher Civic Schools'; this was a type of secondary school that existed in the Netherlands and the Dutch Empire between 1863 and 1974. It was replaced in 1974 by 'Voorbereidend Wetenschappelijk Onderwijs (VWO): pre-university education

in all types of education. It is, however, quite arbitrary and differs substantially from systems in other countries. Only Belgium and Spain have a similar arrangement (Meuffels, 2004a). In Germany, for instance, the grading system ranges from '6' to '1'. Grade '1' is the highest grade and grade '6' is the lowest. Austria has the same scale, but the Austrian grade '1' is lowest and grade '6' highest. Note furthermore, that from a formal point of view it is not allowed to calculate the mean of a set of grades, or to grade using one, let alone two decimals. Grades '...are based on conventions and the grading system has nothing to do with our decimal system.' (Meuffels, 2004b, p.16). Grades form an ordinal scale; they are mere categories and not real numbers: a grade '8' is not twice as good as a grade 4. And the difference between a grade '5' and '6' is not the same as the difference between a grade '6' and a grade '7'. In 1929 a structural change of the grading scale in the Netherlands took place: the grade of '5', representing a just sufficient achievement until then, was declared to mean 'just insufficient'. This was done because there was much discussion in educational circles at that time about the precise meaning of that grade (De Groot, 1970). Since then, nothing fundamentally has changed, and the different grades have the following interpretation %:

- **10:** Excellent
- **9:** Very good
- **8:** Good
- **7:** Very Satisfactory
- **6:** Satisfactory (Sufficient/Pass)
- **5:** Almost satisfactory
- **4:** Unsatisfactory
- **3:** Very unsatisfactory
- **2:** Poor
- **1:** Very poor

When the pass-or-fail score has been established through relative or absolute norming or the Cohen method as described above, the scores can be transformed to grades. The proper way to do this is by using a split linear method as described by Sanders and Verstralen (2017, p. 151-152). Using this method ensures that the lowest score will correspond with grade 1 and the maximal score – or, in case the Cohen method is used, the 95th percentile score achieved in practice – will correspond with grade 10. The split linear method warrants that both below and above the pass-or-fail score unrounded grades are equidistant and the rounded grades are distributed as evenly as possible.

A score S below or equal to the passing score S_p can be transformed to a grade G by applying the following formula and by appropriately rounding of:

$$G = S \times \frac{4,5}{S_p} + 1)$$

A scores S equal to or above the passing score S_p , with a maximum score S_{max} , can be transformed to a grade G by applying the following formula and by appropriately rounding of:

$$G = S \times \frac{4,5}{S_{max} - S_p} + 10 - S_{max} \times \left(\frac{4,5}{S_{max} - S_p} \right)$$

9 Source: <https://www.nuffic.nl/en/education-systems/the-netherlands/grading-systems>

Example 6

For a 40 open-item assessment with a pre-fixed 75% standard (and a correct answer on an item is assigned one point), the pass-or-fail score is calculated as $0,75 \times 40 = 30$. So, the grade 5,5 corresponds with a score of 30. A grade of 1 corresponds with a score of 0 and a grade of 10 with the maximum score of 40.

For scores below the pass-or-fail score of 30, grades can be calculated using formula (1) above. This results in the following grades rounded to one decimal in the columns with the label **'Fail'** below. For scores above the pass-or-fail score of 30, grades can be calculated using formula (2) above. This results in the following grades rounded to one decimal in the columns labeled **'Pass'** below.

Table 1.5.1
Score-Grade Transformation

Fail						Pass	
Score	Grade	Score	Grade	Score	Grade	Score	Grade
0	1	10	2,5	20	4	30	5,5
1	1,2	11	2,7	21	4,2	31	6
2	1,3	12	2,8	22	4,3	32	6,4
3	1,5	13	3	23	4,5	33	6,9
4	1,6	14	3,1	24	4,6	34	7,3
5	1,8	15	3,3	25	4,8	35	7,8
6	1,9	16	3,4	26	4,9	36	8,2
7	2,1	17	3,6	27	5,1	37	8,7
8	2,2	18	3,7	28	5,2	38	9,1
9	2,4	19	3,9	29	5,4	39	9,6
						40	10

Figure 1.5.1 shows how this score-grade transformation results in a 'kink' (hence split-linear) at the pass-or-fail score of 30. Note that the grades underlying the figure are not rounded as in Table 1.5.1, to graphically reflect that the score-grade transformation is a linear one. Using rounded values, would result in grades not being on a perfectly straight line in the figure below.

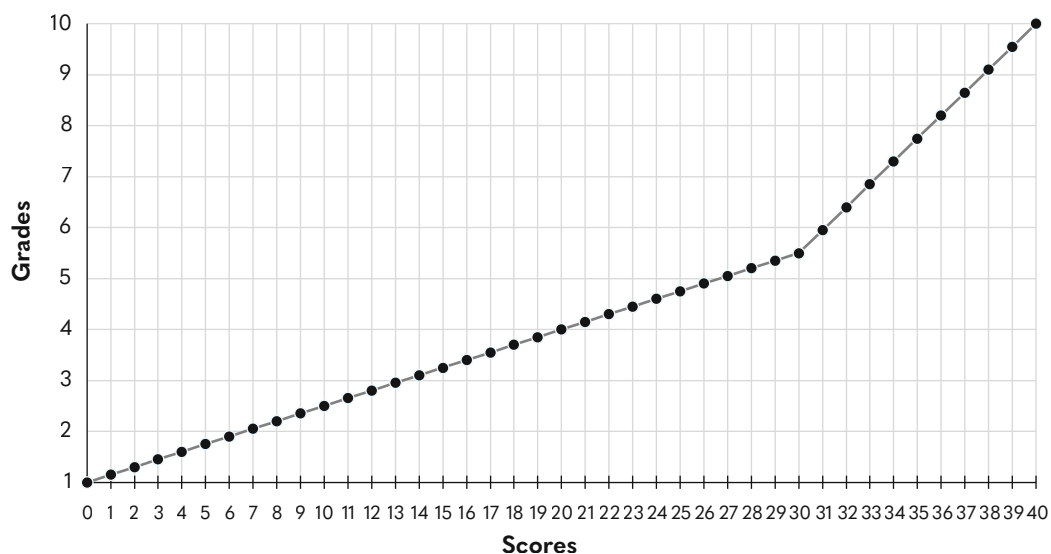


Figure 1.5.1
Score-Grade Transformation

Concluding Remarks

With both absolute and relative norming, the passing score may be insufficiently aligned with the actual difficulty of the test items. In practice, panel-based standard-setting methods are often not feasible for course examinations due to constraints in time, resources, and expertise. As a result, compromise methods that combine elements of both approaches present a workable alternative. However, because these methods retain a partially norm-referenced component, they may still result in relatively high passing scores. Therefore, it is important to evaluate their use carefully and to build an adequate evidence base by applying them in parallel with a more established standard-setting method.

When combined with the application of guessing correction and score transformation methods to grades, it is essential to ensure that the overall approach remains transparent and comprehensible. A key priority is to be able to clearly justify why these methods are used and to provide concrete examples or illustrations that help students understand how their final grades are determined.

With the increasing use of e-assessment software, examiners no longer need to perform manual calculations. However, it remains essential to consider the following points in the checklist.

Checklist of Norming Practices

- Follow institutional policy for standard setting (absolute, relative, or a compromise approach such as Cohen's method), including rules for correction for guessing (whether it is required, for which item types, and which formula is applied) and score-to-grade transformations (e.g., linear or piecewise linear methods, and the rationale for the chosen approach).
- Review institutional formulas used in paper-based or e-assessment systems, and seek clarification whenever calculations, assumptions, or implementation details are unclear.
- Ensure **transparency** by documenting the method and providing students with clear examples or illustrations of how passing score is determined, how correction for guessing is applied, how raw scores are transformed into final grades.
- Evaluate exam difficulty (is the assessment appropriately calibrated, or systematically too easy or too difficult?).
- Consider cohort characteristics (including group size, presence of resits, and representativeness of the student population).
- Be aware of what happens when an institution sees an unexpectedly high or low pass rate. Most quality assurance frameworks treat it as a signal for review, not automatically as an error.
- Assess the impact on failure and pass rates (are outcomes reasonable, defensible, and in line with expectations for the course level?).
- Validate results against an alternative established method (to check robustness and support an evidence-based justification of the chosen standard).

1.6

Formative Assessment¹⁰

Introduction and Learning Goals

During the Teaching Meeting at Tilburg University on January 21, 2021, Dominique Sluijsmans outlined how **formative assessment** practices can be embedded in the curriculum and what important steps can be taken towards an assessment culture in which the formative and **summative assessments** are balanced.

Since **assessment** often drives students' learning, it is no surprise that students highly focus on the assessment and assessment format. The risk of too many summative assessments is cramming and short-term learning; for students it becomes most important to pass the test. To promote long-term learning in which students are encouraged to take responsibility for their own learning, it is worthwhile to emphasize the formative purpose of assessment. The purpose of formative assessment is to organize activities undertaken by lecturers and students which help them to modify teaching and learning activities. Formative assessments yield information that helps students understand where they are in their learning, where they need to be, and how they are going to get there.

This chapter provides lecturers with more detailed instructions on providing formative assessment. The core concepts are:

- assessment patterns¹¹ when using formative assessment (no or low stakes) (Russell, 2010);
- feedback loop: Feed-up, Feedback and Feedforward (Hattie & Timperley, 2007);
- formative assessment strategies (Fisher & Frey, 2014);
- effective and constructive feedback (Bloxham & Boyd, 2007).

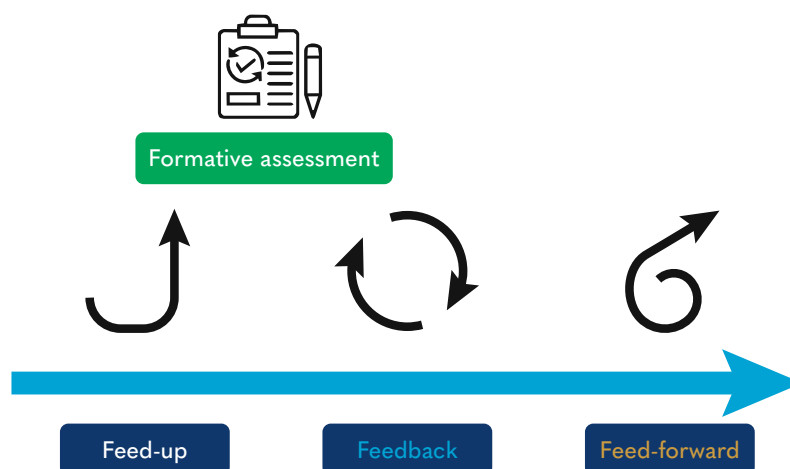


Figure 1.6.1

Formative Assessment and Feedback Loop

¹⁰ This chapter is adapted from the assessment handbooks developed by the Tilburg School of Humanities and Digital Sciences (TSHD).

¹¹ <http://pcwww.liv.ac.uk/~jwgaynor/TAPS/TAPS-MarkRussellslides.pdf>

After applying the instructions of this course to their own assessment practices, lecturers are able to:

- Recognize the conditions that make formative assessment both truly formative and effective;
- Use or adapt an assessment pattern that combines the use of formative assessment and feedback;
- Use formative assessment strategies to prompt student learning towards the final attainment of course learning goals;
- Identify student learning deficiencies and provide students with effective and constructive feedback;
- Evaluate the formative assessment results and adapt follow-up or future instructional activities.

Conditions That Make Formative Assessment

The main confusion around the term formative assessment lies in the word assessment. Many educators think assessment is the same as grading, but this is a misconception. Assessment is much broader than grading; in fact, grading is only a small part of assessment practices. Grading involves giving a final, official judgment at the end of an assignment, unit, or group of units. These results are used for formal purposes, such as certification, promotion to the next stage, or entry into a higher level of education. Another misconception is that grading results must be numerical, whereas they can also be expressed qualitatively, such as through pass-or-fail decisions or performance categories.

This misunderstanding also leads to debates about whether formative assessment should be graded. Most agree that it should involve no grading or only be graded when a crucial decision needs to be made, particularly at key moments such as a thesis proposal. However, because formative assessment is essential for effective learning, it is possible to design conditions that still meet course requirements. This is especially effective when formative assessment is embedded in interactive lectures, where students can demonstrate their thinking and progress, receive feedback from others, and engage in meaningful feedback discussions.

Formative Assessment is Purposeful

Students should have a clear understanding of the purpose and direction of their learning. By clarifying and sharing learning intentions and success criteria, lecturers ensure that students know what they are expected to learn and how their progress will be assessed, not only at the end of the course, but continuously throughout the learning process (Black & Wiliam, 2009).

Formative Assessment is Assessment FOR Learning

Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting students' learning. It thus *differs* from assessment designed primarily to serve the purposes of accountability, or of ranking, or of certifying competence. An assessment activity can help learning if it provides information to be used as feedback, by lecturers, and by their students, in assessing themselves and each other, to modify the teaching and learning activities in which they are engaged. Such assessment becomes 'formative assessment' when the evidence is actually used to adapt the teaching work to meet learning needs (Black et al., 2003, p.9) and for students to use to improve their learning in the process and performance on the sub-sequent assessment practice.

Formative Assessment is an Integral Part of Assessment AS Learning

According to Yan and Boud (Yan & Boud, 2022), assessment as learning is defined as: "[a]ssessment that necessarily generates learning opportunities for students through their active engagement in seeking, interrelating, and using evidence" (p.13). Beyond assessment for learning, it emphasizes integrating formative and summative purposes, aligning learning activities with assessment tasks, and promoting student agency, metacognition, and self-regulation.

In the age of GenAI, it has become increasingly important to foster student ownership, responsibility, and independence in learning (Liu et al., 2024). Assessment as learning is grounded in the assumption that students are willing to engage in the learning process (Dann, 2014). Formative tasks should therefore be designed to help students recognize and understand their learning gaps, while formative feedback should guide them in closing those gaps.

To Close the Gap

Formative assessment is to help the learner close the gap between the present state of understanding and the learning goal. The lecturer's role is to communicate appropriate goals and promote improvement as learners work towards the goals. Feedback in the classroom is interactive: it should operate from lecturer to students and from students to lecturer (Dobson & Fudiyartanto, 2023).

The purpose of formative assessment involves diagnosis at different points, namely:

- at the beginning of a program, in initial guidance or formal **diagnostic assessments**;
- during lessons and sessions, through dialogue and question and answer;
- during tutorials and reviews;
- when reading and assessing assignments;
- during groupwork, through listening to students talk to each other;
- from specific diagnostic activities: tests, quizzes, "fear in a hat".

It is the purpose of the assessment that makes it formative, not the method, timing, or activity: a timed examination/test can be used solely for formative purposes, for example.

Frequently Asked Questions About Policies and Implementations

If a formative assessment is only used for feedback (no grading, no consequences), then there is no additional rule on how it should be organized. Lecturers have the full autonomy to design it. But if the formative assessment is *not only* for feedback (i.e., no grading but has consequences), please consult the following FAQs.

Q: Can I make formative assessments compulsory that students need to complete them?

A: Consult your institutional Education and Examination Regulations, Rules and Guidelines, or Assessment Policy for more information.

Q: Can I assign a bonus point for those who complete a formative assessment?

A: It depends on your institutional policies (see above).

An example of policy states: "Awarding bonus points must be done sparingly. They should be used formatively for motivating students and rewarding excellence (students showing a performance that exceeds the normally expected **learning outcomes** of the course)" (Assessment Policy of Tilburg School of Economics and Management, 2024, p.14). Keep in mind that grades can mostly be assigned to assessments that show student attainment of course learning goals.

In some institutions, bonus points are not used as part of formal assessment. However, in general, a student's grade should reflect their achievement in a specific subject, as assessments are intended to measure learning outcomes. Bonus points are generally considered a form of extra appreciation, but they can introduce construct-irrelevant variance and therefore potentially risk distorting the **validity** of the measure of student performance.

Q: Do I need to indicate formative assessments in the accountability document such as a Course assessment overview or an accountability table?

A: It depends on your institutional policies (see above).

The information of formative assessments is important for the academic director to make the program assessment plan. You can add formative assessment information to your regular accountability table (e.g., course assessment overview).

If your formative assessment leads to any consequence (e.g., fail/pass, compulsory completion to finalize the course), it is possible that you will need to make a table, according to your institutional Assessment Policy.

Q: In my assessment design, students have to peer review each other's work and give each other feedback. I regard this activity as a formative assessment. But then I already noticed that some students did not do it because they did not get a grade for this investment. What should I do?

A: This is a common issue reported by many lecturers. Two perspectives should be considered.

From a didactic perspective, we can communicate the purpose of this activity. Do not give students the impression that we want to save ourselves time regarding giving feedback. Instead, we should emphasize why this activity can promote their learning performance: help them get a higher grade on the final assessment. Also, we will still give students collective feedback, based on our observations in some selected samples. The aim of lecturers' feedback is to find common issues and work on them.

As for assessment, if evaluation skills are part of the course learning goals, the peer feedback results can show their understanding of the content and quality criteria. It is possible to assign a grade on the quality of peer feedback. But you will need to specify the quality criteria on how peer feedback is graded. Consider whether this is feasible for a large class.

Q: What are the design principles of formative feedback? (Gibbs & Simpson, 2004; Nazaretsky et al., 2026)

A:

- Formative feedback should be precise, personalized, accurate, and useful.
- Sufficient feedback is provided, both often enough and in enough detail.
- The feedback focuses on students' performance, on their learning and on actions under the students' control, rather than on the students themselves and on their characteristics.
- The feedback is timely in that it is received by students while it still matters to them and in time for them to pay attention to further learning or receive further assistance.
- Feedback is appropriate to the assignment's purpose and its criteria for success.
- Feedback is appropriate in relation to students' understanding of what they are supposed to be doing.
- Feedback is received and attended to.
- Feedback is acted upon by the student.

Why Formative Assessment?

10 benefits of formative assessment (Roediger et al., 2011):

- Benefit 1 – The testing effect: retrieval aids later retention
- Benefit 2 – Testing identifies gaps in knowledge
- Benefit 3 – Testing causes students to learn more from the previous or next learning episode
- Benefit 4 – Testing produces better organization of knowledge
- Benefit 5 – Testing improves transfer of knowledge to new contexts
- Benefit 6 – Testing can facilitate retrieval of information that was not tested
- Benefit 7 – Testing improves metacognitive monitoring
- Benefit 8 – Testing prevents interference from prior material when learning new material
- Benefit 9 – Testing provides feedback to lecturer(s)
- Benefit 10 – Frequent testing encourages students to study

Whether or Not to Use Formative Assessment

Black and Wiliam (1998) found overwhelmingly positive results in their meta-analysis of effectiveness of formative assessment on student learning gains. In addition, implementing quality formative assessment can effectively reduce a wide spread of achievement levels because it helps low achievers more than other students (Black & Wiliam, 2010).

When formative assessment is well planned and properly implemented, all lecturers are recommended to use formative assessment to gather more accurate information than gut-feeling on student learning. When acted upon, this process will improve how lecturers adjust their instruction and, in turn, improve how well students learn.

One downside of using formative assessment for lecturers is the extra workload involved in integrating it into the overall course design, creating meaningful assessment tasks, and providing high-quality feedback. This includes explaining the purpose of formative assessments, motivating student participation, and designing frequent quizzes

with sufficient feedback. With advances in GenAI technologies, hybrid feedback approaches are becoming feasible. By combining AI-generated formative tasks and feedback with human oversight or validation, educators can enhance trust and ensure the quality of feedback while reducing workload (Nazaretsky et al., 2026).

How to Design and Implement Formative Assessment

Assessment Patterns

The following assessment pattern example (Russell, 2010) illustrates how to engage student learning and balance study load between formative and summative assessments.

- Making more of feedback
- Moving away from high-takes, end-of-process assessment

W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	Exam
			Problem Definition & Analysis plan			Data & Results			Group Project (40%)	Quiz	Quiz	Mock exam	Exam (60%)
						Feedback		Feedback		Feedback			

Benefits of this pattern.¹²

- Engages students with the course from the start.
- Distributes workload evenly across the semester.
- Uses two medium-stakes assessments to moderate the impact of assessment outcomes.
- Offers a balanced mix of no- and medium-stakes assessments.
- Provides opportunities for feedback after each formative assessment.
- Encourages students to use feedback from formative assessments on the subsequent assessment.
- Allows students to identify and correct misconceptions through low-stakes formative assessments.
- Allows the final exam to integrate and assess learning from group projects.

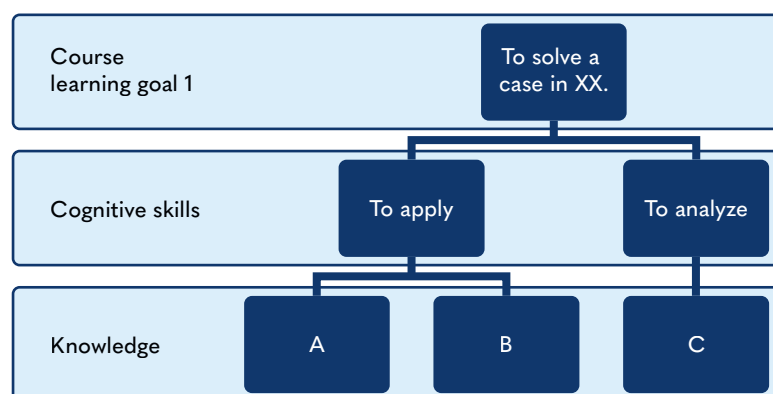
Decision Moments in a Learning Path (Popham, 2008)

Formative assessment aims to improve teaching and learning. “To improve” in the process refers to making adjustments when they are needed (i.e., it is difficult to continue with learning without these adjustments).

A learning path is a logical sequence of several building blocks that students need to master to successfully attain more distant and higher-level course learning goals. These *requisite* building blocks consist of constituent knowledge and sub-skills.

For example, for a course learning goal, students need to solve a case in a specific type of organization (see Figure 1.6.2). To reach this distant goal, students first need to apply theoretical concepts A and B into the organization’s context C. This requires students to analyze context C by using a set of case analysis skills and then to relate theories A and B to the context. Therefore, formative assessments (FA) are needed for both lecturers and students to determine whether they need to adjust teaching and learning before application and analysis.

¹² With permission to use

**Figure 1.6.2**

Example Learning Path to Achieve Course Learning Goal 1

Appropriate Formative Assessment Types.

Based on the phenomenon backwash effect¹³, the use of formative assessment should result in desirable adjustments in teaching or learning. As Ramsden (2003) pointed out, “from our students’ point of view, assessment always defines the actual curriculum” (p.182). Choose appropriate formative assessment types to assess student mastery level of requisite precursory cognitive levels (formative assessment).

Consult “Tools for formative assessment techniques to check for understanding”¹⁴.

Formative Assessment Cycle

With shared understanding of course learning goals (CLGs) and assessment criteria, lecturers can involve student peers and themselves to carry out formative assessment and formulate feedback (Black & Wiliam, 2009; Hattie & Timperley, 2007) (see Table 1.6.1).

Table 1.6.1

Activities of a Formative Assessment Cycle

	Feed-up	Formative assessment	Feedback	Feedforward
	<i>Where the learner is going</i>	<i>Where the learner is right now</i>		<i>How to get there</i>
Lecturer	Clarify and share learning goals and criteria for success.	Carry out effective classroom discussions and other learning tasks that elicit evidence of student understanding.	Provide feedback on the status of student learning: what meets the criteria and what does not.	Provide suggestions that move learners forward. Adjust teaching.
Peer	Engage in the meaning of learning goals and criteria for success.	Activate students as instructional resources for one another. Peer assessment & peer feedback		
Student	Engage in the meaning of learning goals and criteria for success.	Activate students as the owners of their own learning. Self-assessment & self-feedback		

¹³ A consequence of assessment design on what students pay attention to and how they approach learning.

¹⁴ <https://www.utwente.nl/en/examination/faq-testing-assessment/60formativeassessment.pdf>

Lecturers' Instructional Adjustments (Popham, 2008)

To determine when to adjust instructions, lecturers can set an adjustment trigger *before* implementing formative assessment, based on (1) a minimum per-student performance level (i.e., a required level of individual student performance) and (2) a minimum per-class performance level (i.e., a required level of total group performance).

For example, a lecturer sets an adjustment trigger (i.e., a rule) that each student should score 80% or higher on weekly quizzes. If fewer than 70% of the students reach this score on a quiz, the lecturer will hold a review session to go over the mistakes.

Students' Learning Tactic Adjustments (Popham, 2008)

Students often do not know what to do with formative assessment results and feedback. Lecturers should provide students guidance on how to deal with feedback (see Table 1.6.2) and arrive at defensible decisions about whether to adjust learning tactics. Lecturers can take the following four steps to develop student feedback literacy (Carless & Boud, 2018):

1. Appreciate feedback

Lecturers should emphasize students' active role in understanding feedback. Lecturer or peer feedback should not merely focus on *telling* students what they must do.

2. Make judgments

When students can make quality evaluative judgements (i.e., self-assess the quality of the work and generate internal feedback), it is more likely that they understand external feedback (from lecturers or peers) better and are more willing to use it for improvement. It is thus important that students practice using assessment criteria and making judgments frequently. In addition, lecturers should examine the accuracy of student evaluative judgements.

3. Manage affect

The effectiveness of feedback depends on whether students perceive it as useful and act on it. Unfortunately, critical feedback can sometimes result in negative student responses. How feedback is perceived not only depends on its constructiveness (i.e., how feedback is formulated) but also on a trusting culture created by the learning environment. To improve positive perception, lecturers can consider using a student-initiated feedback sheet so that students get more focused feedback and feedback that suits individual's learning needs (Bloxham & Campbell, 2010).

4. Take actions

As was observed in the course assessment packages, most of the lecturer feedback focuses on the task level, which is specific and makes it difficult for students to generalize and apply the meaning to other tasks and to take subsequent action. According to the Hattie and Timperley's (2007) model, process and self-regulation level feedback are generally most likely to lead to learner uptake and improvement.

Table 1.6.2*Four Types of Feedback (Chou & Zou, 2020; Hattie & Timperley, 2007)*

Feedback Type	Example comments
Personal-level feedback: Focuses on evaluating and expressing positive affect about the learner. This feedback type contains less task-related or meta-cognitive information and leads to fewer learning gains but affects students' motivation and beliefs	Good job! You worked very hard!
Task level feedback: Informs how well tasks are understood or performed. It directly informs students about correctness and task completion.	XX is not correct! The XX is missing.
Process level feedback: Guides the processes needed to complete tasks or correct errors.	You've correctly set up the first part of the calculation but check step 3 because you need to apply the formula for the second variable. Try recalculating this step and see if your final answer changes.
Self-regulation level feedback: Addresses how students monitor, direct, and regulate their own learning, including self-assessment, goal setting, and other regulation actions.	You rated your understanding as 'excellent,' but your quiz shows some gaps. Review Section XX and adjust your self-assessment accordingly."

SECTION 2

THEORY

2.1

Exams With Selected-Response Question Types¹⁵

Ya-Ping (Amy) Hsiao

Introduction and Learning Goals

Selected-response (SR) question types include multiple-choice questions (MCQs) (one correct answer), true-or-false (binary choice) questions, multiple-response questions (requiring several correct answers), statement-based questions, hotspot and hot-match questions, ranking questions, and matching questions (McMillan et al., 2013). While these formats are efficient for grading, they require greater expertise and experience in item writing to ensure high-quality **assessments**.

After applying the instructions of this chapter to their own assessment practices, lecturers are able to:

- Provide assessment information for the exam in the course documents and on the exam cover sheet.
- Construct an accountability table in accordance with institutional policy to ensure the **validity** and **reliability** of an exam with selected-response question types.
- Develop sufficient and effective selected-response questions aligned with the course learning goals.
- Apply **corrections for guessing** accurately.
- Review item quality using the four-eyes principle based on a checklist at the end of this chapter.
- Evaluate exam **items** based on test- and **item-analysis** to determine whether each item performed satisfactorily.

Transparency

Include the following assessment information in the course document (e.g., syllabus, course manual) when using an exam with SR types.

- Indicate whether it is an open- or closed-book exam and which materials (e.g., dictionary) or tools (e.g., calculator) can be used during the exam.
- State the weight of this exam towards the course grade, question type(s) (e.g., multiple-choice or short answer questions), and number of questions.
- State whether there is a fail/pass decision based on the results of this exam and resit possibilities. Consult Chapter 1.5 to set a **passing score** and to transform raw scores to course grades.

15 This chapter is adapted from the assessment handbook developed by the Tilburg School of Humanities and Digital Sciences (TSHD).

- State clearly when the correction for guessing is applied to determine the **cut-off**/passing score, both in the course documents and exam cover sheet. See Chapter 1.5 for more information on correction for guessing.
- State that it is important that students answer all questions, even if they have no or only partial knowledge.
- List the course materials (e.g., textbook chapters, slides, handouts) that will be covered by this exam.
- State the possibility of exam inspection and how it is arranged.
- Organize a mock exam study session effectively (see Info Box 2.1.1). Stimulate student agency by asking them to generate questions for in-class plenary discussions or online peer-review activities (see Info Box 2.1.2). Student-generated questions can also be adapted by the examiners to use in the real exams afterwards.
- Use the standard exam cover sheet from the institution, if available.

Box 2.1.1

Design Guidelines for Organizing a Mock Exam Study Session

Research suggests that the most effective exam preparation combines practice exam questions with structured study sessions that review course materials and provide targeted feedback on misconceptions or difficult concepts. To maximize student learning, educators should design study sessions that integrate realistic practice, clear performance expectations, and lecturer(s)-guided feedback, rather than relying solely on peer discussion or unstructured Q&A (Dotson, 2010).

When organizing mock exams or study sessions, educators should:

- Align mock exam questions directly with the course learning goals.
- Ensure questions cover key topics and reflect the level of difficulty expected in the formal exam.
- Deliver questions using paper, a learning management system, or an e-assessment platform, ensuring that students can attempt them in a way that *mimics the formal exam experience*.
- Allow multiple attempts to provide practice and reinforce learning.
- Compile anonymized student responses along with lecturer comments to provide constructive feedback.
- Encourage students to submit discussion questions before the study session.
- Facilitate discussions orally in class and actively engage students in analyzing questions and reflecting on their own actual performance to prevent occurring of the Dunning–Kruger effect (i.e., students overestimating their own potential to perform) (Kun et al., 2023).

Box 2.1.2

Multiple-Choice Item Development Assignment (MCIDA) (Doyle & Buckley, 2022; Fellenz, 2004)

Purpose

To promote active engagement with course content, encourage reflection on cognitive learning levels, and develop **assessment literacy** through the creation and analysis of MCQs.

Learning Benefits

- Supports both lower- and higher-order learning.
- Encourages meta-learning by requiring students to consider the knowledge needed to answer questions.
- Develops understanding of how assessment items measure **learning outcomes**.

Preparation Materials

Students should review guidance on multiple-choice item design and Bloom's Taxonomy.

Task Requirements

For each submitted multiple-choice item, students must:

1. Write a question stem.
2. Provide three answer options (one correct and two incorrect **distractors**).
3. Identify the correct answer.
4. Explain why the correct answer is accurate.
5. Explain why each incorrect option is incorrect.
6. Identify the highest level of Bloom's Taxonomy required to answer the item correctly.

Item Structure

- 1 question stem
- 1 correct answer option
- 2 incorrect answer options (distractors)

Cognitive Framework

Bloom's Taxonomy should be used to classify the highest cognitive level required for each item.

Submission Examples

Example of a completed student submission in the course within a master's program of educational assessment: A university lecturer(s) notices that students achieve high scores on weekly quizzes but perform poorly on the final examination. After reviewing both assessments, the lecturer(s) concludes that the quizzes may **NOT** be measuring the same learning outcomes as the final examination.

Which assessment concept best explains the lecturer(s)' concern?

- A) Reliability
- B) Construct validity
- C) Authenticity

Correct Answer: B) Construct validity.

Why option B) is Correct:

Construct validity refers to the extent to which an assessment accurately measures the knowledge, skills, or abilities it is intended to assess. If the quizzes and final examination are intended to assess the same learning outcomes but produce substantially different patterns of performance, the quizzes may lack construct validity.

Why the Other Options are Incorrect:

A) Reliability is incorrect because reliability concerns the consistency of assessment results. The scenario does not provide evidence that quiz scores are inconsistent or unstable.

C) Authenticity is incorrect because authenticity concerns how closely an assessment reflects real-world tasks or applications. The issue described is whether the quizzes measure the program intended learning outcomes, not whether they resemble real-world practice.

Bloom's Taxonomy Level: Level 5: Evaluation.

Students must evaluate a realistic assessment scenario, compare multiple assessment concepts, and determine which concept provides the most appropriate explanation. This requires judgment and application of assessment theory rather than simple recall of definitions.

Reliability

In the context of SR exams, reliability depends on whether the exam includes enough well-constructed items to produce stable and consistent scores across students. Although scoring in this context is typically objective, reliability can still be influenced by factors such as test length, time, item clarity (see Checklist of Item Quality), **item difficulty** and discrimination (see Post-Hoc Test and Item-Analysis), guessing (see Correction for Guessing When Setting Passing Scores), and testing conditions (e.g., noise or distractions in the room, technical issues in digital testing). Ensuring reliability therefore requires careful item design, representative sampling of the course content and thinking skills, and standardized administration procedures to minimize measurement error and support dependable interpretations of student performance.

Length

To determine the number of questions in an exam, consider the allowed exam time (practical constraints set by your institution) and number of CLG assessed. Within the allowed exam time, to reach an acceptable level of reliability, the exam should contain a minimal number of questions.

In addition, to reach a satisfactory degree of test reliability, consider the number of options (i.e., alternatives) for each question and determine the minimum number of questions (Milius, 2007). It is recommended, especially for novice lecturers, to use *three* options for MCQs because writing effective distractors (i.e., alternatives that are not the answer) is difficult and time-consuming. And adding more alternatives does not lead to better quality (Arifiyanti et al., 2023; Rodriguez, 2005). With more alternatives, some options are seldom chosen (Rodriguez, 2005), which indicates that they are distractors of low quality. Moreover, with three alternatives, more items fit within the allotted exam time, thus increasing reliability. Another reason for this recommendation is that novice lecturers do not have much input about student misconceptions or common errors to be used in writing effective distractors.

Finally, when MCQs are used, it is recommended to use the same number of alternatives for the entire exam. This reduces cognitive load (i.e., mental effort) for students and makes it easier to apply a correction for guessing, unless a digital exam is used so that the scores and grades are automated calculated by the system.

Time

One pre-requisite to ensure exam reliability is whether the exam time is sufficiently long for all or almost all students to complete the exam so that they can demonstrate their thinking skills of the course content (Traub & Rowley, 1991). It is therefore essential to determine the number of questions to fit the time frame (e.g., two-hours exams) and to estimate exam time carefully at the same time. When estimating response time (Brothen, 2012; Wendt & Kenny, 2009) (see Chapter 2.4), consider the reading time based on the length of question stems and alternatives, cognitive levels, task types, and additional time to review the whole exam quickly before submission (i.e., so-called last answer time). For complex questions and questions requiring calculation, allocate thinking time next to response time. Additionally, consult your institutional accommodation rules for students with special needs to have extended exam time.

Post-Hoc Test- and Item-Analysis

After test administration, exam quality can be evaluated based on student responses. The four most commonly used psychometric indices for this purpose are test reliability, item difficulty, **item discrimination**, and **Standard Error of Measurement** (SEM) (Haladyna, 2015).

The most commonly used reliability coefficient is Cronbach's alpha although it can be shown that it is an underestimate of the measuring precision of an assessment. It ranges from 0 to 1, with values closer to 1 indicating higher reliability. Note that Cronbach's alpha is also group-dependent; meaning the value can vary depending on the composition of the group of students taking the test. Generally speaking, the more homogeneous the group, the lower

the score variance and the lower alpha is. In addition to this, note also that it is often stated that alpha measures the internal consistency of a test, in the sense that a high alpha indicates that all items measure the same underlying **construct**. But this is not correct, as shown by Sijtsma (2009). Internal consistency is the extent to which the items in a test cohere as expressions of a shared psychological process or behavioral principle. Internal consistency is therefore an aspect of the test's construct validity.

Item difficulty indicates how easy or difficult a question is. It is defined as the proportion of students who answered the question correctly, with a dichotomous (binary-score; 0-1) item. And is then called the **p-value**:

- Difficulty = 0: no students answered correctly (very difficult).
- Difficulty = 1: all students answered correctly (very easy).

The p-value is calculated by dividing the number of correct responses by the total number of students. In case of a polytomous (partial-credit; multi-points) item, the item difficulty is defined as the mean item score divided by the maximum item score. Note that a better term would be item easiness, because the p-value statistic is counterintuitive: a high score means that the item is easy, while a low score means that the item is difficult.

Item discrimination measures how well a question distinguishes between higher-performing and lower-performing students. It is calculated by comparing the performance of the top 27% of students with the bottom 27%, based on their overall exam scores.

- Positive discrimination: higher-performing students answer correctly more often (desirable).
- 0: both groups perform similarly.
- Negative discrimination: lower-performing students answer correctly more often, which may indicate a problem with the question. Such items should be reviewed and checked with a colleague with content expertise.

The SEM provides an estimate of the *uncertainty* around each student's observed score. It should be used to interpret scores as ranges rather than precise point values, to evaluate whether differences between students are statistically meaningful, and to support fair and defensible decisions about student achievement. In exams with multiple-choice items, SEM is particularly important for interpreting scores as ranges of ability, identifying when small score differences do not reflect real differences in performance, and informing fair grading decisions, especially for students *near cut-off or boundary* points.

Validity

Accountability

When designing exams with SR question types, ensure that:

- The most important and representative content areas are assessed. Constructing an overview to show how course learning goals are assessed by SR questions helps, such as course assessment overview or a test matrix.
- The cognitive levels of exam questions align with the intended course learning goals, especially ensuring that the highest levels of the goals are adequately assessed.
- If a particular sub-domain is more important, include more questions on that topic rather than assigning greater weight to individual items.
- The weighting of lower- and higher-order thinking skills reflects the course's position within the overall program. For example, avoid assigning excessive weight to basic knowledge-level questions (e.g., memorization) in second- or third-year bachelor's courses or in master's courses—unless the content is new to most students. This does not require assigning different weights to individual question types.

Cognitive Levels

SR question types are mostly used for assessing lower cognitive skills. When written properly with contexts, such as cases and scenarios, SR question types can also assess higher cognitive skills (see Table 2.1.1).

Table 2.1.1

Cognitive Levels Assessed by Commonly Used SR Question Types (Haladyna et al., 2004; Van Blerkom, 2017)

Types	Cognitive levels based on Bloom's taxonomy
With 2 options True/False, Yes/No, Correct/Incorrect, Fact/Opinion	Knowledge (factual)
Matching	Knowledge & comprehension (understanding of concepts, principles, and procedures)
Conventional MCQs with 3 (recommended number) or 4 options	Knowledge & comprehension When using a context (with cases/scenarios): application, analysis, and evaluation
Hot spot Questions are presented with an image and must click or select a specific area on it	Knowledge & comprehension
Multiple true/false	Knowledge
Multiple-statement questions	Comprehension (conceptual understanding) & analysis

Checklist of Item Quality

To ensure item quality, use the checklist below to review your own closed questions. This checklist is based on the writing guidelines suggested by Haladyna et al. (2002) and Burton et al. (1990).

- Each question has been constructed to assess the targeted course learning goal.
- A specific problem is stated clearly in the stem, in the form of either a question or an incomplete statement.
- The question uses paraphrased textbook language or language used during instruction to avoid testing for simply recall.
- The content of each question is independent from content of other questions on the exam.
- The question is based on the course content instead of student opinion.
- The stem includes sufficient information to answer the question, without including irrelevant material.
- The stem is stated in a positive form.
- The alternatives are worded clearly and concisely.
- The alternatives are mutually exclusive.
- The alternatives are homogeneous in content.
- The alternatives are free from clues as to which response is correct. They avoid giving clues to the right answer, such as:
 - Specific determiners including *always*, *never*, *completely*, and *absolutely*.
 - Clang associations¹⁶, choices identical to or resembling words in the stem.
 - Grammatical inconsistencies that clue the test-taker to the correct choice.
 - Conspicuous correct choice.
 - Pairs or triplets of options that clue the test-taker to the correct choice.
 - Blatantly absurd, ridiculous options.

¹⁶ Clang associations are groupings of words, usually rhyming words that are based on similar-sounding sounds, even though the words themselves do not have any logical reason to be grouped together.

- Avoid using 'all of the above' and 'none of the above,' especially when options are *randomized* in an e-assessment system.
- Avoid using double statements* in a single question. If more than one topic is involved, split the question into separate questions.
- The question includes one and only one correct or clearly best answer.
- The questions are laid out in a clear and consistent manner.
- The grammar, punctuation, capitalization, and spelling are correct.
- The question uses simple vocabulary (avoid unnecessarily difficult vocabulary).
- All alternatives should be of similar length to avoid the tendency of item constructors to use shorter formulations for distractors.

***What is wrong with questions with double statements?**

Questions with double statements present two statements in the stem, and students must indicate whether none, one, or both statements are correct.

- They often combine different topics, making the question harder than necessary without adding learning value (i.e., extraneous cognitive load).
- They focus on the relationship between statements rather than the intended content, making it difficult to analyze results and understand what students truly know.
- They are vulnerable to guessing, because students with partial knowledge can use eliminating strategies to find the correct answer.

2.2

Exams With Constructed-Response Question Types¹⁷

Ya-Ping (Amy) Hsiao

Introduction and Learning Goals

A constructed-response (CR) question requires the examinee to generate an answer, typically in the form of one or more sentences (Livingston, 2009; Miller et al., 2009). Unlike selected-response formats (such as multiple-choice questions/MCQs), constructed-response questions do not provide answer options. Instead, students must compose their own responses. In many cases, there is no single predetermined correct answer. The accuracy, relevance, depth, and overall quality of the response are evaluated by a subject expert using professional judgment.

After applying the instructions of this chapter to their own assessment practices, lecturers are able to:

- State the assessment information of the exam with constructed-response questions in the course documents and on the exam cover sheet.
- Construct an accountability table in accordance with institutional policy to ensure the **validity** and **reliability** of an exam with constructed-response question types;
- Develop sufficient and effective constructed-response questions with grading instruments (e.g., model answers or **rubrics**) based on the course learning goals and cognitive levels specified in the accountability table;
- Review item quality and grading instruments using the four-eyes principle based on a checklist at the end of this chapter.
- Recognize possible bias and use grading strategies to restrict subjectivity.

CR question types

CR questions are generally classified into two main formats: short answer (restricted response) and long answer (extended response) (Miller et al., 2009).

Short answer or restricted response questions require brief and focused responses. The scope of the answer is clearly defined, and specific limits may be placed on content, length, or time. Student responses are typically graded using model answers or a simple scoring rubric (for example: more than sufficient = full points; sufficient = partial points; insufficient = no points).

¹⁷ This chapter is adapted from the assessment handbook developed by the Tilburg School of Humanities and Digital Sciences (TSHD).

Long answer or extended response questions require more detailed and comprehensive responses. These questions assess higher-order thinking skills such as analysis, synthesis, and evaluation. Within this format, questions may be either closed-ended or open-ended.

- In closed-ended essays, the response is structured and guided, with clear limitations on content coverage, length, and/or time. Student answers are often graded using model answers or structured rubrics similar to those used in restricted response questions (Reiner et al., 2002).
- In open-ended essays, students have considerable freedom in organizing and presenting their ideas, with minimal restrictions on length or structure. Responses are typically evaluated using broader assessment criteria with defined performance levels (for example, excellent, good, satisfactory, and unsatisfactory), focusing on argument quality, organization, depth of analysis, and clarity of expression (Reiner et al., 2002).

Table 2.2.1

Examples for Each CR Question Type

Assessment Type	Example Question	Typical Response Length
Short Answer / Restricted Response	Define construct validity (the extent to which an assessment measures what it intends to measure) and briefly explain why it is important in educational assessment in no more than three sentences.	1-3 sentences
Long Answer / Extended Response: Closed-Ended Essay	Compare and contrast reliability (consistency of measurement) and validity (accuracy of measurement). In your answer, explain their relationship and implications for designing educational assessments. Limit your response to 500 words.	50-100 words
Long Answer / Extended Response: Open-Ended Essay	Assessment shapes student learning experiences. Critically evaluate the role of assessment in higher education and discuss how current assessment practices support or hinder meaningful learning. Use relevant theories and examples to support your argument.	100-300 words

Transparency

- In the course document (e.g., syllabus), state the following information:
 - the course materials (e.g., slides, handouts, textbooks, and compulsory or recommended reading materials) assessed by the exam;
 - the exact type of exam: closed- or open-book, which tools/materials are allowed during the exam;
 - the type and number of CR questions.
- In the preparation lecture for the exam, include the following in the slides and publish the slides on the learning management system:
 - information of the above-mentioned points;
 - information of past or sample exam questions;
 - preparation strategies (optional);
 - answering strategies regarding the interpretation of questions, by using the definitions of action verbs.
- In the exam, add the number of points assigned to each main question and sub-question.
- In the model answers, describe what makes good answers (e.g., a list of key words, criteria, or simple rubrics) and add the instructions on how points are assigned to each aspect.
- Use the standard **exam cover** sheet from the intranet:

Box 2.2.1

*Sample Instructions on the Exam Cover Sheet on How to Answer CR Questions*¹⁸

Constructed-response questions: Answer the questions concisely and formulate your answers clearly. Please pay close attention to the phrasing of the questions, this will tell you what type of answer is expected (e.g., “name” or “mention” requires you only to list/mention things, whereas “describe” or “explain” require you to elaborate on something).

Knock-out criterion: Responses must be written in comprehensible English that allows the examiner(s) to understand and evaluate the answer. Submissions that cannot be understood due to insufficient language clarity will not be graded and will receive no points.

Reliability

Length

Providing students with sufficient time is an essential condition for ensuring reliability. The required length and complexity of student responses directly determine the amount of time needed to complete the task (see Table 2.2.2).

To determine the number of questions used in exams with CR questions, please consider the types of CR questions and response time required (see Chapter 1.4).

Table 2.2.2

Response Time for Different Lengths of CR Questions based on Chapter 1.4

Types of CR Questions	Response Time in Minutes
answer requires 1 word or 1 sentence	4-5
answer requires ¼ A4 page	10-12
answer requires ½ A4 page	20-30
answer requires 1 A4 page	40-60
answer requires 2 A4 pages	80-120

Item Quality

Make CR questions based on the accountability table and keep the following points in mind:

- Examine whether each CR question aligns with the content topics indicated in the course learning goals.
- Choose an appropriate type of CR question to measure the highest level (based on Bloom’s taxonomy) indicated in the learning goals (see Table 2.2.3).
- Check whether each question is measured at the correct level of cognitive skills.
- Examine whether the points/weights of the course learning goals and levels of cognitive skills reflect the emphasis and balance of the course.
- Check whether the number of questions is feasible within the limited exam time (see Table 2.2.2 and Chapter 1.4).
Revise the accountability table if necessary.

Grading Constructed-Response Questions

The following recommendations on grading CR questions are based on Hogan and Murphy’s summary of literature review (Hogan & Murphy, 2007).

- Score anonymously.

¹⁸ Written by Nynke van der Laan, with permission to use

- Score one **item** at a time (i.e., horizontal grading).
- Use a model answer or rubric.
- Separate mechanics (i.e., language and writing skills) from thinking skills of the course content knowledge.
- Use a second reader.
- Rescore some papers; recalibrate.
- Provide feedback, comments.
- Score all responses to an item in one sitting.
- Rearrange the sequence of exam papers.
- Read the sample before grading.

Box 2.2.2

*Example of A Good CR Question with Model Answers*¹⁹

Question:

Give a short (maximum of three sentences per concept), academic definition of the following concepts. You can use definitions from the papers you read (such as Sherman and Craig (2002), Craig (2013), or others), or construct your own.

(NB: a good academic definition is concise, specific, and does not include examples)

- Virtual reality [1 point]
- Immersion [1 point]
- Telepresence [1 point]
- Augmented reality [1 point]
- In C, you were asked to define telepresence. Different types of definitions exist. Which types of definitions exist? Explain and defend why you are using this type of definition. Give at least two arguments in favor and one against your definition. You can use arguments based on Steuer (1992). [3 points]

Grading instructions:

All definitions should be concise, use academic language, and not contain examples.

- Definition should mention or describe immersion and/or (tele)presence, as well as the real world being replaced (at least at the visual level) with a digital (virtual) world. Can be described both as an experience and as a collection of technologies/media.
- Definition should mention the literal immersion of several senses OR immersion as a broader concept of forgetting the world around you and feeling connected with the virtual world.

Validity**Cognitive Levels**

Table 2.2.3 shows the two common types of CR questions and the cognitive levels they are best suited to assess.

¹⁹ Written by Lisa E. Rombout, with permission to use

Table 2.2.3

Cognitive Levels, Types of Complex Learning outcomes and Examples of Short and Long Response Questions (Miller et al., 2009)

	Short Answer / Restricted Response Questions	Long Answer / Extended Response Questions
Cognitive levels based on Bloom's taxonomy	From knowledge to analysis	From comprehension to synthesis/creation
Types of complex learning outcomes	The question mostly assesses whether students can interpret and apply data in a specific area. Explain cause-effect relationships Describe application of principles Present relevant arguments Formulate tenable hypotheses Formulate valid conclusions State necessary assumptions Describe the limitations of data Explain methods and procedures	The question assesses higher order thinking skills and gives students freedom to demonstrate their ability to select, organize, analyze, evaluate, and integrate ideas. The choice of questions and the scoring procedure needs considerable attention. Produce, organize, and express ideas Integrate learning in different areas Create original forms (e.g., designing an experiment) Evaluate the worth of ideas
Examples	Name and explain (in one sentence) the three techniques used to transfer knowledge within an organization based on the theory of cognitive apprenticeship. (9 points, ¼ A4 page, 10 minutes)	Evaluate the significance of using three techniques of cognitive apprenticeship (scaffolding, mentoring, and coaching) on employees' cognitive and metacognitive development within an organization. (24 points, 1 A4-page, 45 minutes)

Accountability

The guidelines for making an accountability table are the same as indicated in Chapter 2.1 Exams with SR question types.

Structure of Constructed-Response Questions

When creating CR questions, it is essential to ensure that each component, namely context, question, and instructions, aligns with the program intended learning outcomes and cognitive levels. Table 2.2.4 provides practical guidelines for structuring CR questions effectively.

Table 2.2.4

3-part Structure to formulate CR questions (TU/Delft, n.d.)

Parts	Guidelines
Context	<ul style="list-style-type: none"> • optional for lower-order thinking skills: knowledge and comprehension • compulsory for higher-order thinking skills, from application to synthesis/creation
Question (a clear task)	<p>Context and question should be linked</p> <ul style="list-style-type: none"> • Make sure the context is relevant for the question. If not, delete it. • If the question can be answered without using the context, then change/remove the context OR change the question. Note: irrelevant information is only allowed in the context when the learning goal includes the thinking skills of judging irrelevant information.
Instructions for answering For example, "Justify your answer, showing which formulas you used. Write no more than 3 sentences."	<ul style="list-style-type: none"> • Use imperative sentences ("List three characteristics of X." rather than "What are the characteristics of X?"). • Specify what you expect in the answer (e.g., "List the <i>three</i> characteristics of X"). • Avoid asking personal opinions such as "What do you think...?"

Checklist Item Quality

To ensure item quality, use the checklist to review your own CR questions. This checklist is based on the writing guidelines suggested by Hogan and Murphy (2007).

Transparency

- Are examinees given appropriate guidance for answering?
- Are recommended words used in the prompt (e.g., analyze, explain, justify, evaluate)?
- Have you stated the time limit, page/word limit, and point value clearly?

Reliability

- Are the number of items and time per item balanced?
- More items = less time per item (short response)
- Fewer items = more time per item (extended response)
- Does the question avoid opinion-based answers unless assessing argumentation or writing skills?
- Have you developed a model answer or scoring rubric?
- Have the questions been reviewed or pretested?
- Is each question/task defined clearly and unambiguously?
- Is the total number of questions manageable within the exam's time limit?
- Is the question determinate, with a clearly better or more supported answer?
- Are an appropriate number of questions used to capture sufficient evidence for each learning goal?

Validity

Alignment with Learning Goals

- Does each question align with the content topics in the course learning goals?
- Does the question measure the cognitive level specified in the learning goal (e.g., comprehension, application, analysis, evaluation, synthesis)?
- Is the type of CR question appropriate?
- Are restricted response questions used for comprehension and application?
- Do extended response questions assess the levels of analysis, synthesis, or evaluation?
- Have you revised the accountability table if needed for better time balance?

Cognitive Level Appropriateness

- Is the question pitched at the correct level of Bloom's taxonomy?
- Does it require students to demonstrate the intended thinking skill (e.g., evaluate, analyze, apply)?
- Does the question clearly indicate the mental processes students must use (e.g., compare, critique, justify)?
- Does the question test complex processes rather than just recall?
- Does it require evidence-based responses?

Assessment Balance and Emphasis

- Are the questions weighed according to the importance of each learning goal?
- Does the distribution of questions reflect the overall emphasis of the course?

2.3

Exams With Both Selected- and Constructed-Response Question Types

Ya-Ping (Amy) Hsiao

Introduction and Learning Goals

Exams with mixed question types use different types of questions, like multiple-choice questions (MCQs), matching, and short- or long-answer questions in the same test (Garg & Lee, 2016). Each type of question checks different skills and putting them together helps give a better overall picture of what students really know and can do. With the introduction of e-assessment in higher education, it is easier to make exams with diverse and mixed question types.

After applying the instructions of this chapter to their own assessment practices, lecturers are able to:

- Explain the strengths and weaknesses of selected- and constructed-response questions and justify the use of a mixed exam format.
- Determine the optimal balance of question types in a mixed-format exam based on coverage and time.
- Design effective preparation strategies for mixed-format exams, considering cognitive and time demands.

Why Mixed Exam Format?

The examination method aims to assess multiple course learning goals across different content topics and cognitive levels. To achieve this, a mixed exam format combines different types of questions, such as MCQs and open-ended questions, to test a wider range of skills and understanding. By integrating selected-response (SR; for example, MCQs) and constructed-response (CR; that is, Very Short Answer Questions (VSAQs), short/long answer questions, essays) questions, this approach makes good use of the advantages of both formats while mitigating their individual limitations, resulting in a more balanced and comprehensive assessment.

SR question types such as MCQs are useful for testing students' understanding of facts and basic ideas. They are quick to answer and cover a lot of material. But in real life, we don't get to pick options. Students need to come up with their own answers.

To improve authenticity (Liu et al., 2014), CR questions ask students to write their own answers, like a short explanation, a written argument, or solving a problem step by step. These types of questions show how well students can think, explain their reasoning, and apply what they have learned to new situations. Research shows that being able to come up with an answer is very different from just recognizing the right one (Mullen & Schultz, 2012).

Kaipa (2021) has compared MCQs and CR questions in several studies. The following table summarizes the advantages and disadvantages of both question types.

Table 2.3.1

Advantages and Disadvantages of MCQs and CR Question Types

Advantages		
Aspect	MCQs	CR Questions
Efficient for large classes	Easy to grade and provide quick feedback	Grading takes time and is more complex
Objective grading	Reduces grading bias; answers are either right or wrong (Scouller, 1998)	More subjective; depends on examiner interpretation
Broad content coverage	Allows testing of many topics in one exam	Limited content due to time needed per question
Faster for students	Students complete them quickly; good for recall-based prep (Parmenter, 2009)	Longer to answer; more effort required
Guessing possible	Can still score points by guessing	Guessing is not helpful; must generate answers
Versatile	Widely used across disciplines	Also applicable to most subjects, especially those needing deeper thought
Disadvantages		
Aspect	MCQs	CR Questions
Encourages surface learning	Focuses on recognition, not deeper thinking (Struyven et al., 2005)	Promotes deeper content understanding
No partial credit	Binary grading doesn't show partial knowledge (Schultz, 2011)	Allows partial credit for partial understanding
Does not promote writing skills	No development of written expression	Encourages structured thinking and writing
Can be guessed	Students may guess instead of reasoning (Frery, 1988)	Guessing unlikely; must show actual understanding
Potential bias	Gender bias observed, men tend to perform better, especially when correction for guessing is applied (Beller & Gafni, 2000; Gellman & Berkowitz, 1993)	Less evidence of such bias
Requires good writing skills	Not required	May disadvantage students with weaker writing skills
Grading workload	Low for lecturer(s)	High—more time and effort to evaluate

By combining both SR and CR question types, a mixed-format exam...

- tests a wider range of learning goals,
- helps students show both what they know and how they think (e.g., testing reflection and reasoning),
- tests more like real-life problem solving,
- gives students more ways to demonstrate what they can.

Assessment Quality Considerations for Mixed-Format Exams

While the general principles of item quality still apply (see Chapter 2.1 and 2.2), designing mixed-format exams that combine SR and CR question types require additional attention.

Combine Scores Meaningfully

Assign appropriate weightings to each question type based on what you want to assess. The proportion of SR versus CR should align with the importance and complexity of the learning goals.

Use Very Short Answer Questions (VSAQs)

In a mixed exam format, it is essential to provide students with adequate time to respond to different types of questions. VSAQs, which typically require brief responses of three words or fewer, can be particularly effective when assessing foundational knowledge or applied knowledge that cannot be well tested by multiple-choice format.

VSAQs are especially useful in professional contexts where quick and accurate recall is necessary for efficient performance. In such cases, MCQs may fall short for several reasons:

- It can be difficult to create plausible **distractors** for certain content.
- The presence of answer choices can lead to guessing or recognition-based answers rather than true recall.
- Some **learning outcomes** require concise, precise responses, which are better assessed through CR formats like VSAQs.

Convert MCQs to VSAQs

If a MCQ can be answered correctly by reading only the question stem (i.e., passes the cover test), it can often be converted into a VSAQ. This encourages active recall over passive recognition.

Box 2.3.1

Example of Converting a MCQ to VSAQ

Psychology (MCQ): “Which part of the brain is responsible for regulating emotions?”

- A. Cerebellum.
- B. Amygdala.*
- C. Hippocampus.
- D. Thalamus.

This question can be rewritten as a VSAQ:

“What part of the brain regulates emotions?”

Answer: Amygdala.

Formative Assessment and Exam Preparation

Include both SR and CR versions of similar questions during **formative assessments** and exam preparation. This approach helps students recognize that CR questions are more cognitively demanding and require deeper understanding and active recall, compared to the recognition-based nature of SR questions.

Additionally, students’ responses to CR questions can be analyzed to identify common errors or misconceptions, which can then be used to create realistic and effective distractors for MCQs in the actual exam.

Estimate Exam Time Carefully

The ideal number of SR and CR questions in an invigilated exam depends on several factors, including:

- Complexity and cognitive level of questions (see Chapter 2.1 and 2.2)
- Reading and response time per **item** (see Chapter 1.4)
- Weighting of assessed learning goals
- Balance between breadth and depth

Assess Internal Consistency Across Formats

Consult your institutional guidelines on how to account for the **reliability** of a mixed-format exam, for example whether Cronbach's alpha should be calculated for the entire exam.

Checklist for Quality Mixed-Format Exam Design

In addition to the checklists in Chapters 2.1 and 2.2 addressing item quality, the following questions can help guide the design of a mixed-format exam.

Transparency

- When **correction for guessing** is applied, is it clearly explained to students whether the correction applies only to SR questions or to the entire exam?

Reliability**Time Management**

- Have you estimated time per item (e.g., 1–2 min for MCQs, 5–10 min for short answer)?
- Can the exam be reasonably completed within the allotted time, including revisiting time?
- Have you provided students with guidance on how to manage their time per section?

Weighting and Scoring

- Are question weights proportional to the importance of the learning goal(s)?
- Is the overall score distribution clear (e.g., 60% MCQs, 40% short answer)?
- Are scoring **rubrics** or model answers prepared for CR questions?

Validity

- Does the exam cover a representative sample of the course content?
- Are question types matched to the cognitive level of the learning goals?
- Do SR questions assess factual knowledge, understanding, and basic application?
- Do CR questions assess foundation knowledge or higher-order skills (e.g., analysis, synthesis, evaluation)?

2.4

Oral Exams²⁰

Ya-Ping (Amy) Hsiao

Introduction and Learning Goals

Oral exams, defined as **assessments** conducted through spoken interaction between the student and examiner(s), offer unique opportunities to evaluate higher-order cognitive skills (Buehler & Schneider, 2009) and communication skills (Moore & Kaplan, 2015; Rawls et al., 2015). These assessments are particularly well-suited for measuring the upper levels of Bloom's taxonomy, from applying, analysis, synthesis, to evaluation (Buehler & Schneider, 2009). The oral exam is one of the oldest assessment forms and is commonly used in thesis defense at both graduate and undergraduate level and for medical clinical certifications (Akimov & Malin, 2020).

All of the guidelines on constructed-response questions in Chapter 2.2 still apply to oral exams. In addition, after applying the instructions of this section to their own assessment practices, lecturers are able to:

- Provide students with clear assessment information, preparation materials, and strategies for taking an oral exam.
- Establish manageable and objective procedures to ensure **reliability** and **transparency**;
- Construct parallel questions based on the accountability table such as course assessment overview..

Transparency

- As for transparency, make sure that you provide students with sufficient information. See “Box 2.4.1 Sample oral examination format for communicating to your students” (Rawls et al., 2015) and “Box 2.4.2 Brief summary of oral exam strategies for students” (Kesselman-Turkel & Peterson, 2004).
- Provide clear time management information before the exam. Inform students about the total time allocated per student and the estimated time for each question. Let students know that you will signal them when their time for a question is nearly finished or has ended.
- Allow thinking time after asking questions. After presenting a question, give students a short period to think and write down their main ideas in a schematic form. Clearly communicate how much preparation time they have before they must start answering.
- Offer an opportunity to revisit unanswered questions. If time allows near the end of the exam, give students the chance to return to and answer any questions they were unable to address earlier.
- Communicate the final grade transparently, according to your institutional guidelines. When sharing the final grade, summarize key points from the completed assessment forms to justify the evaluation.

²⁰ This chapter is adapted from the assessment handbook developed by the Tilburg School of Humanities and Digital Sciences (TSHD).

- Use assessment forms during inspection sessions. Refer to the completed forms when students review their results, ensuring transparency and consistency in the grading process.

Box 2.4.1

Sample Oral Examination Format for Communicating to Your Students, Adapted from Rawls et al. (2015)

General Structure

You will be asked five questions during the oral exam. Your answers will need to have an introduction and conclusion and must be structured. You will have approximately 30 minutes to complete the entire oral exam.

Number of Examiners

There will only be two examiners assessing you. They will use a **rubric** to grade your answers as you are speaking.

Canvas Conference Room and Time

You will log in to the learning management system Conference room according to the scheduled time.

Level of Interaction

When the oral exam starts, we will present you a question. You receive three minutes thinking time, the opportunity to write down keywords. After that we will not answer any questions you may have. We may ask for clarification during the oral exam, but we will not ask probing questions other than a clarification of your answers.

Topics to Cover

The exam study guide serves as the range of questions you will be asked in the oral exam. Instead of writing the essays, you will provide succinct answers orally.

Box 2.4.2

How to Prepare for An Oral Exam, A Brief Summary of "Oral Exam Strategies" for Students (Kesselman-Turkel & Peterson, 2004)

Like written exams, oral exams also require preparation. In fact, preparation is often even more important because you have less time to think about your answers during the exam.

First, find out which topics will be covered. In many cases, the scope of the oral exam is clearly defined. If it is not, ask your lecturer for clarification. You can ask which topics will be included, but avoid asking directly about the exact questions on the exam.

Once you know the topics, prepare well. For each topic, try to learn:

- Three key points, and
- One or two examples that help explain the topic.

Sometimes oral exams also evaluate your public speaking skills. To prepare for this, practice explaining the topics with classmates or friends in small groups. This can help you become more confident and allows others to give feedback on how you can improve. Practicing is especially helpful if you feel nervous speaking in front of people. On the day of the exam, try to make a good first impression. Dress appropriately and behave professionally. This is not about influencing the examiner to give a higher grade, but about showing respect and professionalism.

When answering questions, you do not need to use complicated vocabulary (unless certain terms are part of the exam) to impress the examiners. Instead, use clear and simple language that anyone can understand.

When the examiner asks a question:

- Write down the key points of the question.
- Identify how complex the question is.
- Take a moment to organize your thoughts before answering.

If a question seems very complex, do not panic. If it is allowed, ask the examiner for clarification or try to rephrase the question to confirm that you understand it correctly.

Finally, stay calm until the end of the exam. When the exam is finished, thank the examiners for the conversation.

Reliability

- Follow institutional regulations. Arrange and conduct oral exams according to your institutional education and examination guidelines (see Box 2.4.3 Example prerequisites for oral exams or presentations).
- Minimize assessment bias. Be aware that factors such as a student's articulateness, shyness, speed of thought, gender, language skills, or familiarity with the examiner can influence judgment. Ensure that these factors do not unfairly affect grades by involving two examiners for the orals.
- Recognize that students may be anxious or nervous during the exam. Adjust your behavior to support a fair assessment, decide whether to appear encouraging, neutral, or stern, and be consistent with all students.
- Plan interventions carefully. Determine in advance if and when you will intervene:
 - Will you repeat a question once if a student does not understand it?
 - Will you switch to another topic if a student cannot answer?
 - Will you ask clarifying questions if a student's answer is unclear?
- Prevent question leakage. Consider whether students sharing questions with peers could affect **fairness**. Use parallel or equivalent questions for different students to avoid advantages or disadvantages.
- Use simple grading rubrics. Apply a clear, manageable rubric (see Box 2.4.4) that evaluates factors such as completeness and correctness of answers. Ensure it can be used to grade efficiently during the exam.
- Consider alternative assessment methods. Recognize that fully reliable oral exams are difficult to achieve. Whenever possible, consider alternative assessment types that may better measure the targeted course learning goals.

Box 2.4.3

Example Prerequisites for Oral Exams or Presentations

It is important to check your institutional Education and Examination Regulations²¹. For example:

- In principle, an oral examination or an oral interim test is conducted by two examiners or by one examiner assisted by a member of the academic staff.
- The examiner may conduct an oral examination or an oral interim test without the presence of a second examiner or a member of the academic staff under the condition that an audio recording is made.

21 Source: Tilburg School of Economics and Management Education and Examination Regulations

More detailed guidelines²²

Consent | Request permission for recording.

- When TWO EXAMINERS are present at the assessment, they must ask the student for permission to record the session at the start of the assessment. This recording will be used for inspection. If the student does not grant permission, the assessment can take place, but the student agrees to have only limited inspection possibilities (i.e., based on examiners' notes if available).
- When only ONE EXAMINER is present at the assessment, the examiner informs the student of the need to record the session for objective assessment as well as inspection. If the student does not grant permission, the assessment cannot take place, and the student needs to wait for the next assessment opportunity. The opt-out policy applies.

Box 2.4.4

A simple rubric for grading an oral exam

Completeness

- 2 pts: The answer completely covers what is asked in the question.
- 1 pt: The answer partially covers what is asked in the question.
- 0 pt: No answer is given.

Correctness

According to the course materials...

- 2 pts: The answer is correct.
- 1 pt: The answer is partially correct.
- 0 pt: The answer is incorrect.

Validity

- Plan exam timing carefully. Decide how much total time will be spent per student, including both preparation and grading.
- Determine the number of questions per student. Ensure that the number of questions is consistent with the intended depth and scope of the oral exam.
- Align questions with course learning goals. Prepare questions that are parallel in difficulty and content across all students, covering the targeted course learning goals. Use standardized question stems and include a list of relevant content topics.
- Use a structured assessment form. Create an assessment form (see Table 2.4.1) that includes all planned questions and model answers. Model answers may reference course materials or key terms. Include spaces to record:
 - The exact question asked (by circling)
 - Notes on the student's response
- Plan for additional questions if needed. Decide in advance whether students may receive extra questions. Clearly define the purpose of these questions and how their answers will be evaluated.

22 Source: <https://www.tilburguniversity.edu/intranet/education-support-portal/teaching/online-hybrid-blended/exam-types/oral-presentation>

- Archive the assessment forms for quality assurance. The completed assessment forms are essential for maintaining assessment quality. Store them according to your institutional procedures and retain them for the number of years required by institutional guidelines.

Table 2.4.1*Template of an Assessment Form for an Oral Exam*

						Rubric	
Course learning goal	Cognitive level(s)	Content topic(s)	Question stems	Model answers (e.g., course materials)	Notes of student answers	Completeness	Correctness
LG1	Knowledge	Theory A, B	<ul style="list-style-type: none"> • What is __? • List ____. 	Textbook Chapter X		2 1 0	2 1 0
LG2	Comprehension	Theory C, D	<ul style="list-style-type: none"> • What do you conclude from ____? • Explain why ____. 	Lecture 1-3 slides		2 1 0	2 1 0
.....			•				
			•				

Note. During the oral exam, examiners can simply circle the options (such as the red circles shown in this table).

Checklist for Organizing An Oral Exam

This checklist provides practical guidance for lecturers on planning and conducting oral exams, drawing on established best practices (Joughin, 1998, 2010; Theobold, 2021). It covers key aspects such as time management, question design, assessment forms, and strategies to ensure **validity** and reliability. The checklist is designed to help lecturers create a transparent, fair, and structured oral examination process while supporting students' ability to demonstrate their knowledge and skills under exam conditions.

To ensure item quality, use the checklist below to review your own closed questions. This checklist is based on the writing guidelines suggested by Haladyna, Downing, and Rodriguez (2002) and Burton, Sudweeks, Merrill, & Wood (1990).

Before the exam

- Provide students with written information in advance (e.g., in a test folder), including:
 - Duration of the exam
 - Type of responses expected
 - Exam procedure
 - Assessment criteria
- Standardize the exam setting to minimize grading bias or external influence.
- Prepare clear and well-distributed questions across the study material.
- Write down all exam questions/topics in advance or prepare a card-draw system.
- Allocate more time to important or high-weight topics.

At the start of the exam

- Begin with simple questions to help students relax.

During the exam

- Skip to a new question if the student cannot answer the current one.
- Ask clarifying questions if answers are vague or incomplete.
- Ask questions clearly and rephrase briefly if needed.
- Ensure students answer the actual question. Do not allow them to sidetrack.
- Assess each answer as it is given and record the score before moving on.
- Base grades only on the responses given during the exam.
- Evaluate solely in relation to the course learning goal(s).
- Start with a clean slate: Remain neutral and unbiased for each student, especially if lecturers have small groups of students. Don't let student course behavior influence your professional judgement.

At the end of the exam

- Summarize key answers and ask the student to confirm accuracy; or:
- Take brief notes during the exam before assigning a grade, if there is no recording.

2.5

Open-Book²³ and Stepwise Exams

Ya-Ping (Amy) Hsiao

Introduction and Learning Goals

Chapters 2.1, 2.2, and 2.3 refer to conventional invigilated close-book exams. This chapter introduces two alternatives: open-book and stepwise exams.

Open-book exams usually refer to conventional on-campus, timed, and invigilated exams, with varying degrees of access to resources and references. Open-book exams are primarily used to assess higher-order and critical thinking skills, where success depends on applying, analyzing, and evaluating knowledge rather than simply locating information in a book²⁴. They allow students to bring permitted resources to an exam, such as notes, textbooks, handouts, slides, and dictionaries. They assess students' ability to find and apply information and knowledge, so they are often used for courses requiring direct reference to written materials, like law statutes, statistics, or acts of parliament.

Stepwise exams break **assessment** into a progressive sequence of stages that together form a single evaluative process. For example, they may involve at-home pre-exam preparation followed by a conventional on-campus close- or open-book examination (Parkes & Zimmaro, 2017). In this structure, each step contributes differently to the final demonstration of learning. The preparation phase allows students to engage with materials in advance, organize their knowledge, and build a foundation of understanding. The examination phase then requires students to draw on this prepared knowledge and use it to apply concepts, analyze information, and construct well-reasoned arguments under timed and supervised conditions.

Both open-book and stepwise exams expect students to apply deeper learning strategies to prepare for the exams. The exam questions can be selected- or constructed-response. All the guidelines on selected-response- and constructed-response questions in Chapter 2.1, 2.2 and 2.3 still apply to open-book and stepwise exams.

In addition, after applying the instructions of this chapter to their own assessment practices, lecturers can:

- Determine whether open-book and stepwise exams are suitable for their own course context and course status in a program curriculum.
- Provide students with clear assessment information, preparation materials, and strategies for taking open-book and stepwise exams.
- Develop an exam process that fosters student learning in both the preparatory and examination phases.

²³ This part is adapted from the assessment handbook developed by the Tilburg School of Humanities and Digital Sciences (TSHD).

²⁴ In Theology, since locating information is a supporting skill for interpretation and argumentation, open-book exams can be appropriate when they are designed to assess those higher-order skills rather than mere retrieval of information.

Transparency

- Clearly explain what is expected from students in an open-book exam. Many students may not understand how this format differs from traditional exams. As a result, they might simply copy information from their materials instead of using the materials to support their answers. Additionally, some students may underprepare, assuming they can rely entirely on their books during the exam. Emphasize that success still requires a solid understanding of the material and the ability to apply it effectively.
- For open-book exams, clearly specify which materials are allowed (e.g., textbooks, handouts, cheat sheets, slides) and in what format (paper-based or digital). For instance, if a cheat sheet is permitted, you might require it to be handwritten and limited to a specific size. This encourages students to actively engage with the material by selecting, organizing, and prioritizing key information in preparation for the exam (Song & Thunete, 2015). For reasons of equality and **fairness**, the lecturer can provide a standard set of materials to all students so that each candidate has access to the same information during the examination. This helps ensure that performance is determined by how students use the provided material rather than by differences in access to external resources or personal notes.
- For case-based, stepwise exams, inform students that you expect specific answers supported by evidence from the case. Make it clear that the goal is not to debate all possible alternatives but to identify the most appropriate response based on the case details (King, 2016).
- Provide rehearsal opportunities. If students have no experience with open-book exams, they need at least some practice to familiarize themselves with the approach they need to adopt. It can be useful to provide a dry run, asking students to time themselves when doing a set task with resource materials at home, and bringing their results to a class session where they are guided through marking their own (or each other's) work, helping them to see what may constitute a good response to the question.

Reliability

- Formulate exam questions, including context and answer instructions, in a clear and straightforward manner.
- Keep in mind that open-book exams usually take more time than closed-book exams (Agarwal & Roediger, 2011). Provide students with sufficient time to prevent a decrease in **reliability** of their scores (Traub & Rowley, 1991). To maximize exam time for answering, share the exam cover page and detailed instructions in advance.

Validity

- Open-book exams aim to assess higher-order thinking skills rather than simple recall or reproduction of knowledge. They are not suitable for testing the lowest cognitive level of remembering facts. Instead, questions should require students to actively use the information available to them, moving beyond summarizing toward application, analysis, synthesis, and even creation.
- From a design perspective, this has direct implications for both assessment and teaching. According to **constructive alignment**, implementing open-book exams requires more than changing the exam format; it requires a deliberate shift in **learning outcomes** toward higher-order thinking skills such as conceptual understanding, critical reasoning, and problem-solving, supported by teaching activities that explicitly develop these abilities (Feller, 1994).
- Similarly, stepwise exams often incorporate case-based or problem-based approaches that function not only as assessment preparation but also as a central instructional strategy throughout the course (King, 2016; Teodorczuk et al., 2018). In such designs, students are encouraged to engage with cases in advance and anticipate potential

questions, for example through classroom activities or by using GenAI tools to explore multiple perspectives and deepen understanding.

- Stepwise exams break assessment into a progressive sequence of stages that together form a single evaluative process. For example, they may involve at-home pre-exam preparation followed by a conventional on-campus close- or open-book examination (Parkes & Zimmaro, 2017). In this structure, each stage contributes to the final demonstration of learning: preparation supports knowledge building, while the examination phase focuses on application, analysis, and synthesis of that knowledge.
- Both open-book and stepwise exams are considered more authentic to real-world practice than traditional closed-book examinations (Packard & Austin, 2009; Teodorczuk et al., 2018). These formats reflect professional contexts where access to information is typically available during task performance, such as general practitioners consulting resources during patient care or job applicants preparing for staged or case-based tasks in advance.
- However, implementation needs careful consideration across different course levels. First-year bachelor's courses often focus on foundational disciplinary knowledge, and in such contexts open-book exams may not always be appropriate. A stepwise closed-book format, where students are given a list of potential questions in advance without knowing which will be selected, can support mastery while still encouraging preparation and understanding.
- Finally, assessment criteria and timing should align with these formats. Performance should be judged based on what students do with information rather than their ability to locate it. In addition, because students spend time searching for and interpreting resources during open-book exams, responses are typically shorter and more focused than in traditional closed-book, memory-based examinations.

Checklist: Setting Up Open-Book Exams

In summary, this checklist highlights key considerations for setting up open-book exams, including exam design, communication with students, and how to prepare students for open-book exams.

Validity

- Ensure learning goals and instructional activities support deep understanding.
- Align questions with higher-order thinking skills (e.g., application, analysis, evaluation, synthesis).
- Avoid questions that rely on simple recall or copying from materials.

Reliability

- Allow extra time to account for searching and integrating information.
- Share exam instructions and cover page beforehand to reduce in-exam confusion.

Transparency

Permitted materials

- Clearly list allowed resources (e.g., textbooks, slides, handouts). Or the lecturer could provide the materials, so that every student has access to the same materials.
- Specify format: paper-based or digital.
- If using cheat sheets, require them to be handwritten and size-limited.

Student preparation

- Explain how this format differs from traditional exams.
- Clarify that success depends on understanding, not just resource availability.
- Provide a practice task with feedback.

Checklist: Setting up Stepwise Exams

In summary, this checklist highlights key considerations for setting up stepwise exams.

Validity

- Ensure that learning goals emphasize the application of prepared knowledge to the analysis of information and the construction of well-reasoned arguments under timed conditions.
- Use realistic, staged case scenarios to reflect professional tasks.
- Base questions on the application of case details, not general theory.
- Design questions that assess decision-making and reasoning.

Reliability – Assessment criteria

- Require specific answers supported by evidence from the case.
- Make clear that students are not expected to explore all possible alternatives.

Transparency

- Encourage students to explore the case before the exam.
- Consider the use of class activities or GenAI tools to analyze the case from multiple angles.
- Clarify expectations for answering case-based exam questions, which can be different from case study exercises at the tutorials.
- Provide a list of potential question types to guide preparation.
- Offer rehearsal opportunities or sample cases with feedback.

2.6

A Follow-Up Oral Process Interview: What It Is and What It Should Not Be²⁵

Ya-Ping (Amy) Hsiao and Laura Koenders

Introduction and Learning Goals

Detecting potential breaches of academic integrity is part of an examiner's responsibility. For this, you need guidance from the university on how to respond, the procedures you must follow, how to report your suspicions, the type of evidence you need to provide, and how to submit it. You also need to be informed about the next steps in the process, such as a decision made by the Examination Board (Glendinning, 2024).

This is not new. However, the presence of Generative Artificial Intelligence (GenAI) means that it is becoming more difficult to detect these breaches of academic integrity, since GenAI has so many different types of use. Note: if the use of GenAI in question is part of the assignment, then of course this should not be seen as a breach of academic integrity. What we mean here is that you've designed an assignment and have decided that the use of GenAI or some form of GenAI use is not permitted (see Table o).

This chapter aims to offer this guidance for both you as an examiner, and for Examination Boards (or Exam Committees, Board of Examiners), with a special focus on possible steps to take in the coursework assessment method (e.g., individual writing assignments, groupwork).

After applying the instructions of this chapter to their own assessment practices, lecturers are able to:

- Recognize the examiner's role and responsibilities as a gatekeeper of academic integrity.
- Apply university procedures to organize a fair and transparent follow-up oral process interview.
- Formulate appropriate questions and demonstrate professional conduct during the oral process interview.

Before We Start

To illustrate this, we would like to start with a case:

When grading essays, you suspect GenAI use in the essays of three of the students due to the general phrases and 'empty' use of language. There were no clear guidelines for GenAI use, although copying a text in full is considered fraud. You decide to ask the students to come by, to hear their side of the story. Two students admit to having used GenAI, one student does not. You decide that

²⁵ This chapter is an adapted version from <https://www.uu.nl/en/education/educational-development-training/knowledge-dossiers/knowledge-dossier-generative-ai-in-education/a-follow-up-oral-process-interview-what-it-is-and-what-it-should-not-be>

the students who have admitted to GenAI use should improve their work and hand this in again for a grade. You grade the work of the student who said not to have used GenAI without penalization.

You see that this approach is problematic for several reasons.

Firstly, you can create inequalities between students. Students who stated they had used GenAI are graded similarly to students without any suspicion of using GenAI. In addition, the fact that the third student did not admit to using GenAI does not clear the student. However, according to the case description, the student is not penalized, which may mean that denial is a beneficial strategy, also for future students. Another inequality issue can develop when each examiner handles the situation on their own. One examiner may decide that a student cannot finish the course, while another examiner will only warn a student.

If the Examination Board examines a case of suspicion of fraud, it will result in a written decision stating whether fraud has been committed and its consequence (i.e., sanction). The Examination Board will decide in line with the rules and regulations as well as jurisprudence (i.e., how previous cases were dealt with). Students are also informed that they can appeal against the decision. Examiners handling the case themselves will probably not act like the Examination Board. Another downside when examiners handle the case themselves is that it remains unknown to the Examination Board or to other examiners whether students committed fraud more than once over time. When the Examination Board is involved, they take recidivism into account when they sanction.

Secondly, there is the risk that students feel falsely accused, which might damage your relationship with them. On this subject, the most asked question from examiners is: when to report a suspicion to the Examination Board? Please follow the policy guidelines issued by the Examination Board or your institution.

What the case above illustrates is that you must consider quite carefully whether you should or how to approach conversations like this, in terms of ungraded formative or graded **summative assessment**. In this chapter, we hope to offer some suggestions and will use the term 'oral process interview' for a conversation used to figure out the learning process, or authenticity of the work of the student. For this oral process interview, the approach you follow is closely tied to the goal of the **assessment**. Therefore, this is the first thing you need to consider.

Our core message is that cultivating academic integrity in our students is a fundamental responsibility of educators. Balancing the dual roles of educator and examiner is no easy task.

If the Examination Board (or your organization) has provided clear guidelines on how or when to report a suspicion (of fraud or irregularity), please follow them. If such guidance is not available, the following sections are designed to help fulfill your responsibility as an examiner.

The Goal of the Assessment Defines the Goal of the Oral Process Interview

For this chapter, we define an assessment as *everything you do to get information about the learning of a student* (Nitko & Brookhart, 2013). This means an assessment can take many forms, such as written assignments, group projects, or multiple-choice exams. What you do with this information is an important distinction, and this is called the *function of the assessment* (William & Black, 1996). For example, you can use the information from the assessment to decide about student learning (summative assessment, also called 'assessment of learning'). Next to that, you can use the information to guide student learning and to know your students better (**formative assessment**, also called 'assessment as learning' or 'assessment for learning'). For a good overview of the different terms 'assessment of/as/for learning' and their relationship to educational outcomes, see the review of Schellekens et al. (2021).

To illustrate this:

- A student can use the information from an assessment to gauge how well they understand the content of the course. In this case, the student uses the assessment as a learning opportunity to further guide their study efforts (assessment as/for learning). For example, at the first attempt of a written exam, a student gets a very low grade of 3.0 and (s)he concludes to study harder or differently on a resit. Or a student who is doing an internship and

the progress is evaluated halfway: is the student on track? what does the student need to learn? what is already mastered?

- A lecturer can use the information from the assessment to decide if the student has enough knowledge to complete a course (assessment of learning). Using the same example as above, if the student still receives a low grade of 4.0 on the resit, the lecturer concludes that the student has not sufficiently met the course learning goal(s) and must re-enroll in the course next time.

In higher education, we have many assessments throughout our courses with combined functions. For example: a student will receive feedback on a writing assignment (for learning), use the feedback to reflect and improve the subsequent work (as learning), and finally the assignment also counts toward the final grade (assessment of learning). With GenAI in the picture, it is becoming increasingly important to be very clear about these different functions in the design of your assessments. Using the same internship example, while the halfway evaluation has no direct consequences, if the student fails to address the feedback received, the consequences should be clearly reflected in the final graded summative assessment. If you would like to know why, we would advise you to read the Swiss Cheese Model suggested by Curtis (2025).

In assessments designed to help students learn (assessments as/for learning), the impact of a breach of academic integrity is different than in assessments designed to make a formal decision (assessments of).

Since in the former there is no consequence (no or low stakes) involved, it does not impact on the decisions about the level of the students, neither at course nor at curriculum level. There is of course a negative impact on the learning process of a fraudulent student. For example, you have designed a writing assessment for learning, to help students write a concise and clear paragraph. Here the learning outcome is a writing skill, so you do not want students to outsource this to GenAI. If a student decides to do so, your feedback will be less valuable and therefore the student won't learn as much.

However, if you want to use the same assessment for a formal decision about how well a student can write a paragraph, it is of course more problematic if a student uses GenAI. This will impact the **validity** of your judgement, because you are not sure whether the results show student attainment of the learning outcomes; either good writing skills or good GenAI use (if permitted). Also, your decision may not be reliable, because you cannot be certain when the student works on the same writing assignment again in a different context (e.g., under supervision), can (s)he demonstrate the same results?

In both cases above, a breach in academic integrity will harm the goal that you have with the assessment. Thus, in both cases it could be valuable to combine the assessment with some sort of oral process interview. This approach is often more easily integrated into instructional formats that involve individual supervision (such as theses or internships) than into others. However, implementing an oral component will take additional resources and time from you as a lecturer, so it is important to carefully consider its feasibility and impact. Oral process interviews can take various forms, during (group) presentations, interactive lecture activities (e.g., tutorials), or as individual conversations with all students, or just with a few selected students. Knowing the goal of the oral process interview will also help examiners make choices regarding the procedures.

How to Organize A Follow-Up Oral Process Interview?

Assessment integrity scholars suggest balancing the process monitoring (focus on better understanding of your student and persuasion) and product detection (determining when to report) is the most appropriate and educational way to ensure academic conduct (Ellis & Murdoch, 2024).

Therefore, there are three scenarios for this kind of follow-up orals:

After Submitting Formative Assessment Components (e.g., Drafts, Sections, Logbooks): process monitoring

In addition to written feedback, oral process interviews are great opportunities to give students oral feedback. These moments can also be used to address unauthorized uses of GenAI tools by persuading students to adopt the expected behavior, especially when they still have the chance to revise their own work. The oral feedback can be integrated as in-class discussion group activities, making the process more interactive and reflective. For example, in a large business education course with more than 300 students, Olsen and Hunnes (2024) implemented dialogic feedback through small-group tutorials (3–4 students), where lecturers provided oral feedback on students' draft essays. These 45-minute sessions encouraged peer learning and revision, guiding students to reflect and refine arguments from the reader's perspective before final submission.

Random Interview Before Grading: persuasion

Some examiners will use the complementary post-assignment interview as an anti-fraud measure. It is important to mention that this is not a grade determining part of the examination, it is only applied to check if the submitted work is done by the student herself/himself. Only a sample of the students (e.g., 5–20%) will be selected to do the complementary oral check. The goal of this random check is to discourage those who consider dishonest practices.

In case you do this based on a sample of your student population, please consider the following:

- We recommend doing the interview shortly after the assignment has been submitted and before you start to grade.
- Preferably pick the students totally random (so not the first 20% in alphabetical order, but for example based on a randomly picked last number of their student number).
- If you decide to use an algorithm to select students, make sure to make the selection criteria explicit to prevent bias.

Targeted Cases During Grading: product detection

- During grading, we recommend documenting the suspected issues and list the questions you plan to ask during this follow-up, as you would have done before GenAI.
 - When inviting students to an oral process interview, keep the message neutral and focused on clarification. See Box 2.6.1.
- In case you are not convinced that the student did not commit fraud after the oral process interview, please follow the regular processes and report this to the Examination Board. In case you are convinced that the student committed fraud, inform the Examination Board as well. Note: The assignment cannot be graded as long as the Examination Board is involved.
- Students have the right to be informed about these processes, so if they are not (yet) described in the Education and Examination Regulations, be conscious of informing them about it.
- The oral interview should be recorded with only one examiner but should only be stored if there is a suspicion of fraud. The recording and storing should be done in a similar way as described in guidelines for Oral Exams of and according to or in line with the Education and Examination Regulations (EER).

Box 2.6.1

Dear Student,

After reviewing your work, I have a few questions about authorship and the authenticity of your work that I would like to discuss with you. To clarify these points, I would like to invite you to a brief oral process interview. Please let me know which of my availability listed below suits your schedule so we can arrange a suitable time. My available time slots are: XX. The duration of the session is XX minutes.

In practice, various assessment methods and interview formats can be integrated into a course design to strengthen educators' dual role as both lecturers and examiners. However, when implementing follow-up oral process interviews—especially those conducted during grading—it is important to consider their potential to induce anxiety and stress in students, which may negatively affect cognitive performance (Ringeisen et al., 2019). Factors such as fear of punishment for poor performance and social pressure can contribute to this anxiety (Ryan & Deci, 2017; Schürmann et al., 2022). To mitigate these effects, educators should keep the message neutral and focused on clarification during the interviews while fostering a supportive, non-threatening environment. This approach promotes **fairness** and reduces the risk of false positives.

Questions You Can Ask During the Oral Interview

With the Goal to Help the Student Learn (Assessment As/For Learning)

As stated above, in assessments designed to steer student learning, there is no formal decision involved. Therefore, the questions you ask are designed to give yourself and the student insight in their learning process, and the impact of GenAI use on this process. It is very important to create an open atmosphere, so be open about the goal of this meeting (e.g., discussing how GenAI is used based on the logbook and AI index) and what you will do with the answers the students give (e.g., students are not sanctioned for mistakes, instead student learn how to use a different approach).

Use eliciting, clarifying, justifying, defending, and explaining questions. These types of questions should resemble the formative feedback that lecturers give during the process of student work (Wellington, 2010a).

Try to align the questions with the level of GenAI use that is permitted for the assignment, and with what students have learned from their earlier AI literacy courses, namely the ethical and responses of AI (such as <https://student.tilburg.ai/#/>). See for examples of questions the column 'example questions to increase student learning, and to give feedback on the use of GenAI' of Table 2.6.1.

With the Goal to Determine Student Learning Outcomes (Assessment Of Learning)

As stated above, the goal of this type of interview is to establish authorship and authenticity of the work. For this, you will need to ask enough questions to increase the **reliability** of your judgement. A conversation like this is also known as 'oral inspection'²⁶. However, be aware that you can never be 100% sure that the student indeed did not use GenAI in a greater amount than allowed. In case you cannot establish the authorship or have serious doubts about the authenticity of the work, please follow the regular processes and report this to the Examination Board and the student.

In order to increase the reliability of your judgement, it is important (as is with the other type of oral verification reviews) that you create an open atmosphere and there is room for a dialogue. **Transparency** about your intentions and the goal of the meeting will help. See Box 2.6.1.

Do not use leading and vague questions (Pearce & Chiavaroli, 2020). If you use leading or vague questions, you may unintentionally help the student provide answers that align with what you expect, rather than allowing the student to demonstrate their own independent understanding of their writing. For example, if you say: "Your paper argues that technology has more benefits than drawbacks, correct?", this is leading the student toward agreeing with this statement, even if they are unsure or did not originally frame their argument that way. Inauthentic students (who may have used GenAI or external help) might simply agree without being able to explain their reasoning, while even genuine students might unconsciously follow the examiner's wording rather than recalling their original thought process.

See for examples of questions the column 'example questions to verify authorship and authenticity' of Table 2.6.1.

26 https://canvas.utwente.nl/courses/15815/pages/what-can-i-do-if-i-suspect-the-unauthorised-use-of-ai-in-assessment?module_item_id=549659

How to Effectively Communicate to Students?

It is crucial to inform students about the permitted GenAI uses and the organization of the follow-up oral process interview: why it is conducted, when it will take place, how long each session will be, and what is expected from students (e.g., answering questions or explaining how the work was done). Additionally, students should be informed that the oral follow-up will be recorded and what happens after the review.

While it is the lecturer's responsibility to ensure the authenticity of student work, the growing integration of human and machine intelligence means that GenAI's role will continue to evolve (Luo, 2024). Therefore, the emphasis should be on ensuring student accountability for their work, rather than attempting to 'determine where the human ends and where the artificial intelligence begins' (Eaton, 2024, p.10). This requires a balance between the lecturer's responsibility and the student's accountability, which should be established from the very beginning of the assignment process. Finally, higher education institutions also have a responsibility to ensure the credibility of their degree diplomas. Disruptive AI technologies make it increasingly challenging to maintain this certifying function. Although structural and systematic changes are needed in the long term, any short-term measure that has the potential to uphold certification standards and diploma quality is worth trying — such as the proposed oral interview process.

Acknowledgements

We would like to thank Rein Cozijn, Dries Deweer, and Brenda den Oudsten for their review and feedback.

In writing this piece, the authors used ChatGPT (Version March 7th, 2025) to improve English language. Full responsibility for the contents lies with the authors.

Table 2.6.1

AI Index Tilburg University, Adapted by Hsiao, den Oudsten, Venhuizen, and Engelen (Hsiao et al., 2025).

Level	Instructions for Students	Recommended lecturer monitoring throughout the process	Assessment as/for learning: Example questions to increase student learning, and to give feedback on the use of GenAI	Assessment of learning: Example questions to verify authorship and authenticity (Hacker et al., 2009)
1. No AI	The assessment is completed entirely without AI assistance. You must not use GenAI at any point during the assessment. You must demonstrate your core knowledge and skills.	Initial brainstorming and idea generation during interactive lectures to ensure that students rely on their existing knowledge, understanding and skills	<ul style="list-style-type: none"> • How did you decide on this structure or approach for your ideas? • What sources did you use, and how did you assess their credibility? • Can you explain how you developed your main idea or argument? 	<p>Based on your observations of the student's earlier work, if the content and writing appear significantly more advanced, you might ask:</p> <ul style="list-style-type: none"> • What challenges did you face when structuring your thoughts? • Looking at this specific section, can you now explain it in your own words? • What are the major differences between the draft and final version?
2. AI-assisted editing	You are allowed to use GenAI tools to review completed drafts for editing, including spelling, punctuation, grammar, syntax, and referencing.	Submission of drafts, with lecturer random checks	<ul style="list-style-type: none"> • What are the major differences between the draft and final version? • Did you use GenAI for editing? In what way? • What type of edits did GenAI suggest? How did you incorporate this? • What was the biggest grammatical issue you had to fix? 	<ul style="list-style-type: none"> • Can you point out a part of your product that you revised manually? • Did AI suggest any sentence rewrites? If so, how did you decide which ones to accept? • What strategies do you use to improve your writing before AI-assisted editing? • How did you ensure that AI did not change the meaning of your ideas or the style of your voice?
3. AI-assisted idea generation and structuring	You are allowed to use GenAI tools to develop or refine initial ideas, materials (e.g., interview guides, stimuli for experiments), paraphrasing, structures, or outlines.	Submission of initial ideas, structures, or draft outlines, with lecturer random checks	<ul style="list-style-type: none"> • Why did you think that the AI-generated ideas suit the assignment goals? • Can you explain why/how you integrate AI-generated ideas with your own ideas this way? • What additional databases did you use beyond AI-generated content? • What are the major differences between the draft and final version? 	<ul style="list-style-type: none"> • How did AI influence the direction of your work? • How did you modify AI-generated ideas to make them your own? • How did you ensure that your final work reflects your own understanding and critical thinking?

4. AI Task completion, human evaluation	You are allowed to use the GenAI tool to complete certain elements of the task (e.g., literature search, data analyses, audio transcription), and you are responsible for critically evaluating its output to meet the assessment requirements.	Submission of drafts, discussion or commentary on the AI-generated content, with lecturer random checks	<ul style="list-style-type: none"> • What specific sections of your product were generated by AI, and how did you ensure they aligned with the assignment requirements? • How did you ensure that AI did not introduce biases or errors into your work? • How did you evaluate the output of GenAI? 	<ul style="list-style-type: none"> • Which AI tools did you use, and why did you choose them for this task? • Did you disagree with any AI-generated content? How did you adjust it? • How did you document and acknowledge AI's contributions in your work?
5. Full AI	You are allowed to use GenAI tools as a “co-pilot” to complete your work, and you are responsible for critically evaluating its output to meet the assessment requirements.		<ul style="list-style-type: none"> • How did you ensure the AI-generated content meets the assessment criteria? • What human input was needed to finalize your work, and how did you ensure accuracy? • How did you check the accuracy of AI-generated references or data? • How did you ensure that AI did not introduce biases or errors into your work? 	<ul style="list-style-type: none"> • Which AI tools did you use to assist you in different stages of the work (idea generation, structuring, writing, refining, etc.)? • How did you document and acknowledge AI's contributions in your work? • Do you think using AI improved your understanding of the subject? Why or why not?

Note. Based on the Work of Perkins et al. (2024) and Examples of Questions to Increase Student Learning or to Verify Authorship and Accountability

Intermezzo – Boundary Crossing in Assessment: Integrating Examinations and Coursework

Ya-Ping (Amy) Hsiao

In most assessment literature and textbooks, assessment methods are typically framed as a binary: either invigilated examinations or unsupervised coursework (as discussed in the following chapters) (see Table I). In the age of generative artificial intelligence (GenAI), however, this distinction may not function appropriately to address emerging challenges, including the integration of GenAI into course design and concerns about academic integrity, particularly student authorship. There is therefore a growing need to move beyond this dichotomy and adopt more integrated approaches. Recent practices and emerging publications already highlight efforts to combine these two modes of **assessment** (Bridges et al., 2002; Richardson, 2015).

Importantly, course-level assessments should be integrated, with assessment components combined meaningfully rather than treated as isolated components. The different assessment components must be carefully balanced so that, collectively, they meet key quality criteria and ensure that final grades meaningfully reflect students' learning and achievement.

Table I.
Comparisons between Exams and Coursework (modified from Richardson, 2015)

Aspect	End-of-Module Examination	Coursework Assessment
Primary purpose	Primarily tests student knowledge and near transfer of knowledge	Both teaches and assesses learning
Quality of work	Time pressure limits depth and academic excellence	Allows students to produce more thoughtful, high-quality work
Mode of expression	Relies mainly on timed written or oral responses	Enables diverse forms of expression (e.g., projects, presentations, multimedia)
Student well-being	Can induce anxiety that affects performance and health	Generally associated with lower anxiety levels
Generalizability over time	Reflects performance in a single high-stakes setting; learning may not be retained or transferable to other contexts	Assesses performance over a longer period, accommodating variations in health and circumstances

What does such integration look like in practice? When GenAI is embedded in course design and instructional activities, assessment strategies must reflect this integration. A coherent course assessment design can therefore follow several key steps:

First, determine an appropriate level of GenAI use (Hsiao et al., 2025) aligned with the course learning goals.

Second, select suitable assessment methods appropriate for course learning goals. In most courses we have observed at Tilburg University, this will involve a combination of an examination and one or more coursework components.

Third, determine the relative weighting of each component based on their scope and demands. Assigning only marginal weight to coursework that requires substantial effort and higher-order thinking is often perceived as unfair by students. For example, group assignments involve significant coordination and collaboration; allocating less than 20% of the total grade may discourage meaningful engagement. The impact of GenAI has driven many lecturers to heavily weight turn to exams with most weights. Keep in mind that heavy reliance on high-stakes final examinations in many university subjects is poorly justified by the balance of empirical evidence (French et al., 2024).

Fourth, given that current technologies cannot reliably detect misuse of GenAI, an alternative approach is to assess students' understanding of their coursework under controlled conditions. This can be done by requiring students to apply what they have learned in more constrained exam-like settings. For instance, through modified essay questions (see Chapter 2.8), or short reflective tasks that assess their ability to interpret, differentiate, explain relationships, and justify assumptions (MacKenzie, 2024) (see Chapter 2.8). A modest weight (e.g., 10–15%) can be assigned to these components. This integrated approach can be applied to all of the coursework assessments in the following chapters.

Finally, examine the relationship between individual students' performance on examinations and coursework. Such analysis can inform adjustments to the relative weighting of each component in future iterations of the course, helping to ensure a more valid and balanced assessment design (Bridges et al., 2002).

This design supports **criterion validity** by examining the relationship between students' coursework performance and their performance on controlled assessment tasks that serve as an external criterion of the same underlying **constructs**. The extent to which these two measures correlate provides evidence of concurrent validity, while the extent to which coursework performance predicts controlled performance or final examination outcomes informs predictive validity. In addition, analyzing the relationship between coursework and examination scores can guide adjustments to assessment weighting, strengthening the overall **validity** of the assessment system.

The following chapters on coursework assessments illustrate how different assessment methods are appropriate for assessing different constructs. The choice of assessment should be guided by the construct of interest and aligned with the course learning goals, including cognitive constructs (such as critical thinking and conceptual understanding) and skills-based constructs (such as communication, writing, and problem-solving skills).

2.7

Assessment of Writing Assignments²⁷

Ya-Ping (Amy) Hsiao

Introduction and learning goals

One common form of coursework **assessment** is the writing assignment. In this chapter, writing assignments refer to any written work that requires students to demonstrate critical thinking about course content knowledge, with varying levels of demand for academic writing skills. Examples include individual or group academic papers and essays. In addition, lecturers may assess students' independent learning and reflective thinking through interim reflective writing tasks, such as learning journals, diaries, logbooks, or personal blogs that capture students' experiences and thoughts during the writing process.

Since November 2022, an ongoing debate has focused on the impact of GenAI on writing assessment, particularly whether students should complete writing assignments with or without its use. From a **validity** and authenticity perspective, if a course adopts an integrative approach in which students are expected to learn how to *think and write* with GenAI, then writing assignments should be designed to reflect this goal (Hsiao et al., 2023). In the age of GenAI, meaningful writing assignments in academic settings should prioritize student agency, engagement, and transfer of learning (Girdhar & Khachatryan, 2023).

At the same time, research indicates that students will use GenAI when it is available, whereas learners who are still developing foundational writing skills should avoid becoming overly reliant on it during the learning process (Shen & Tamkin, 2026). Consequently, decisions about the extent to which GenAI is permitted in writing assignments should be made at the program level and aligned with students' progressive development of writing proficiency. The program academic directors should make sure a learning trajectory of academic writing anchored in the program **learning outcomes** is developed with the help of the course coordinators. Building on established Writing Across the Curriculum (WAC) principles, two broad approaches to designing writing assignments have been emerging in response to GenAI (WAC Clearinghouse, 2026a).

Learning-to-write (LTW): The aim of LTW is to use writing to help students express their ideas clearly and communicate them effectively to readers. Therefore, it is essential to connect LTW to real-life situations where the ideas will be used. When students choose their topics, explore different perspectives, and write with real-world goals in mind, their work becomes more personal and relevant. To deepen this engagement, it is important to organize discussions with peers and professors, along with connections to future applications (authenticity) (if possible), so that these discussions trigger students to “actively draw from prior knowledge in order to think in new ways or when

²⁷ This chapter is adapted from the assessment handbooks developed by the Tilburg School of Humanities and Digital Sciences (TSHD) and the Tilburg School of Social and Behavioral Sciences (TSB).

they can see how their work may be applied in future contexts” (Girdhar & Khachatryan, 2023, p.2). GenAI can support the writing process, but the value of writing lies in students’ critical thinking and ownership. It is not just generating texts. One way to implement LTW is through supervised coursework assignments, in which lecturers continuously monitor students’ progress and provide **formative assessment** and feedback throughout the writing process.

Writing-to-learn (WTL): The aim of WTL is learning to think, namely mastery of the content (WAC Clearinghouse, 2026b). In most content-focused courses across a program curriculum (except in programs specifically devoted to composition studies), writing assignments are typically designed as WTL activities. WTL refers to the use of writing as a cognitive tool that helps students deepen their understanding of disciplinary or subject-specific knowledge. Through activities such as clarifying concepts, justifying choices, reflecting on ideas, organizing arguments, or synthesizing information, students use writing not merely to demonstrate knowledge but to actively construct and develop their understanding of course content. In this context, assignments often require only limited academic writing skills and place greater emphasis on critical thinking.

In the WTL approach, students can engage in a wide variety of writing tasks for different purposes, formats, and audiences, such as Analyzing the Process or Event (WAC Clearinghouse, 2026c). The emphasis shifts from evaluating the final written product to assessing the thinking processes it reflects and promotes. As a result, the assessment focuses more directly on evidence of students’ understanding and reasoning, making learning outcomes easier to verify. GenAI can support thinking process, but the verification can be tested, for example, in invigilated exams.

In disciplines such as business and economics, the focus is usually on the *written* product rather than the process of writing itself. Tasks often involve solving problems or answering short, constructed-response questions that assess knowledge, comprehension, and basic application in familiar contexts, such as mathematics or statistics exercises. Consult “Chapter 2.2 Exams With Constructed-Response Question Types” about this situation.

This chapter will focus on both LTW and WTL and offer a brief synthesis of recent evidence-informed suggestions. After applying the instructions of this section to their own assessment practices, lecturers are able to:

- Select appropriate assessment methods, such as supervised coursework (Learning-to-write, to ensure the integrity and validity of the evaluation process or invigilated examinations (Writing-to-learn).
- Design meaningful tasks that assess program learning outcomes (PLOs), course learning goals, writing skills, and character development, while integrating formative assessment and ongoing feedback to support students’ learning.
- Develop clear assessment procedures for writing assignments, including detailed grading instructions and criteria to ensure consistency and **transparency**.
- Incorporate verification methods, such as oral checks or presentations, to confirm students’ understanding and mastery of course content and written work.

LTW – Supervised Coursework

Although GenAI can help produce coherent text with citations, academic writing requires awareness of course contexts, a deep understanding of course content including concepts and theories, familiarity with disciplinary thinking conventions such as using formal modeling to represent economic behavior, markets, and policy effects, critical engagement with the literature, rigorous analysis of data and evidence, methodological precision, and original thought, all of which GenAI cannot yet fully perform independently (Milano et al., 2023). While a single prompt can generate essays or theses on a large scale, these outputs cannot meet these requirements without human orchestration and oversight. When internal cognitive processes such as reasoning, relating, and comparing remain active, GenAI can extend or scaffold cognition. With careful and logical planning, students can strategically delegate certain tasks to GenAI, also referred to as cognitive offloading, and integrate its outputs by verifying and either adopting or

overriding them. Such recursive interactions with GenAI outputs require deeper reflective engagement to produce writing of high quality (Shaw & Nave, 2026).

The so-called cognitive surrender, in which students rely solely on GenAI tools or others to complete an assignment (Shaw & Nave, 2026) and present it as their own work, is considered academic misconduct. Research shows that students who place greater trust in automated systems (Jian et al., 2000) and have lower needs for cognition (Lins de Holanda Coelho et al., 2020) are more likely to exhibit such behavior (Shaw & Nave, 2026). Other examples include a lower preference for complex or challenging tasks, lower enjoyment of thinking or cognitive engagement, or lower enjoyment of thinking or avoidance of cognitive effort. High time pressure is also likely to increase the occurrence of cognitive surrender. Providing intermediate feedback on partial tasks and offering incentives (Shaw & Nave, 2026), for example, by highlighting the added value of engaging with writing assignments (Girdharry & Khachatryan, 2023), can encourage human oversight behaviors such as verifying and overriding GenAI outputs.

Based on these recent findings, several design principles can guide the formative and **summative assessment** of writing assignments in contexts where GenAI use is allowed to a certain extent.

- Review current summative assessment criteria, ensuring that grading shifts from rewarding lower-order thinking skills toward higher-order thinking skills. In addition, performance descriptors should be formulated as qualitative statements of achievement, rather than simple yes/no indicators of whether specific elements are present.
- Distribute deadlines across the course period and align them with the stages of the writing process. You may want students to engage in substantial writing, but this does not necessarily mean that every student needs to submit work every week. One approach is to divide the class into manageable sections and have a different section submit an assignment each week (University of Waterloo, n.d.). Alternatively, if weekly writing practice is desired for all students, everyone can submit assignments while the lecturer provides detailed feedback to only a subset of students each week, while the remaining submissions receive a simple pass-or-fail grade based on completion. This structure helps manage lecturer workload while maintaining regular writing practice. These submissions can also be sequenced across the stages of the writing process, such as brainstorming, drafting, structuring arguments, managing citations and references, proofreading, and formatting (Hayes, 1980). Organizing assignments in this way allows lecturers to embed formative assessment throughout the writing process and provide timely feedback at intermediate stages rather than evaluating only the final product. When possible, these supervision activities can be integrated into interactive teaching formats, such as tutorials. Supervision does *not* always need to be individualized; small group supervision and peer discussion strategies can also be used effectively.
- Second, student characteristics should be taken into account, including factors such as trust in automated systems, level of technology acceptance, and need for cognition. Conducting a pre-measure at the initial stage can help gauge the likelihood of cognitive surrender (Shaw & Nave, 2026). Such a **diagnostic assessment** (e.g., technology acceptance survey [Saif et al., 2024]) enables lecturers to tailor formative feedback and scaffolding strategies to students' learning profiles and their potential reliance on GenAI.
- Third, students should be made aware of the limitations and potential flaws of GenAI at each stage of the writing process, such as hallucinated citations, shallow engagement with the literature, and the homogenization of discourse (Haber et al., 2025). Students can be asked to reflect on how AI-generated outputs are evaluated, adopted, modified, or overridden during their writing process. These reflective activities function as formative feedback loops that encourage metacognitive awareness and responsible AI use in academic writing. See Chapter 2.13 Reflection and Logbook Assessment.
- Fourth, determine whether a follow-up oral process interview is required or whether reflective questions on the writing process will be assessed as part of another invigilated exam. This step ensures that students can articulate and critically reflect on their writing decisions, including how they engaged with GenAI tools and integrated feedback throughout the process.

Transparency

In the syllabus, clearly specify the following regarding the writing assignment:

- The percentage it contributes to the final course grade.
- The relative weight of the draft and final versions, if the draft is graded.
- Formative assessment and feedback, using any logbook or reflection to monitor the process.
- Whether a minimum **pass score** or grade is required, and if resubmission (resit) is allowed. If so, explain the resit process and whether a maximum grade will be capped upon resubmission, in accordance with your institution's policies.
- The permitted uses of GenAI tools, based on your institution's AI policy, and approved usage levels (Hsiao et al., 2025).
- The policy on plagiarism and academic integrity should clearly define expectations. When GenAI use is permitted to a certain extent, it should specify how students are expected to acknowledge and account for their use of GenAI. This accountability document (e.g., logbook, technology statement, acknowledgements) should remain concise. Requiring students to produce extensive documentation, such as hundreds of pages, does not enhance transparency. Accountability becomes meaningful for students' learning only when it is combined with reflection and supported through dialogue in formative or summative assessments.
- State clearly if any further verification is required (see Writing-to-learn of this chapter and "Chapter 2.7 A Follow-Up Oral Process Interview").
- It is essential that the assessment form (outlining assessment criteria, definitions, and how each criterion is evaluated) is provided to students at the start of the course. This helps them understand how their work will be assessed. After grading, the completed assessment form must be returned to students to support learning through the lecturer's summative feedback, which should primarily explain the rationale for the grades awarded.

Reliability

Assessment Form

To grade writing assignments, lecturers must construct an assessment form (e.g., **rubric**) in which they specify several criteria to assess the work. Lecturers can customize their assessment form to ensure that the assessment criteria correspond with the course and assignment learning goals.

Core Criteria of Good Student Writing

The commonly used criteria and their indicators to grade both student critical thinking and writing skills are listed below (Association of American Colleges and Universities, 2009a; Elander et al., 2006).

Please note that these are suggested guidelines and should be adapted to the specific context of your course and course learning objectives. For example, if a course includes the learning outcome "Students will produce clear and academically appropriate writing by using GenAI tools to refine language and style, while ensuring the final work reflects their own understanding and voice," and GenAI tools are permitted for tasks such as language correction, editing, or structuring (i.e., AI index Level 2) (Hsiao et al., 2025), then the relevant criterion should be assessed on a met/not met basis rather than graded.

Critical thinking/critical evaluation

- The material is presented in a critical manner.
- The theory is applied clearly through critical analysis/critical thought of the topic area.
- The evaluation includes conceptual/methodological critique and an appreciation of alternative perspectives and current controversies.

Argument

- The author sustains a well-reasoned and supported argument.
- Logical arguments are clearly present throughout.
- A good development is shown in the summary of arguments based on the theory/literature.

Evidence – Selecting and using information to investigate a point of view or conclusion (VALUE²⁸ – Critical thinking **rubric**) (Association of American Colleges and Universities, 2009a)

- Information is taken from source(s) with enough interpretation/evaluation to develop a comprehensive analysis or synthesis.
- Viewpoints of experts are questioned thoroughly.

Influence of context and assumptions (VALUE – Critical thinking rubric)

- Thoroughly (systematically and methodically) analyzes one's own and others' assumptions and carefully evaluates the relevance of contexts when presenting a position.

Student's position (perspective, thesis/hypothesis) (VALUE – Critical thinking rubric)

- Specific position (perspective, thesis/hypothesis) is imaginative, taking into account the complexities of an issue. Limits of position (perspective, thesis/hypothesis) are acknowledged.
- Others' points of view are synthesized within a position (perspective, thesis/hypothesis).

Structuring

- The writing has a clear, logical, and well-defined structure. (e.g., is there an introduction, middle and conclusion?)
- The essay is well structured with obvious sections.
- The writing is clearly structured with well-organized material.
- The conclusion draws together the various important points made in the main body of the writing.
- The answers are extremely well organized with a structure that reflects the development of the argument.

Use of language/writing style (graded as met/not met)

- The writing is generally clear, readable, and well presented.
- It makes the reader want to read it.
- Spelling and grammar are used correctly.
- The language is fluent, with accurate grammar and spelling.
- The material and arguments are presented clearly and coherently.
- The student's personal voice and style are evident, showing originality and authentic expression.

Criteria and Grading Instructions

All criteria must be properly defined with a set of quality indicators. Furthermore, the assessment form must specify how points will be awarded to each of the criteria and how the final grade for the work will be determined. There are two basic ways to do this:

- A pre-specified number of points can be ascribed to each criterion, and the total number of points adds up to 10 (or 100 points and the total is then divided by 10).
- Each criterion²⁹ can be assessed using a scale, for example a four-point scale: Very good (4), Good (3), Sufficient (2), and Insufficient (1). The lecturer specifies how the criterion scores are converted into a final course grade on a scale of 1–10.

²⁸ Valid Assessment of Learning in Undergraduate Education (VALUE)

²⁹ Except the ones that should simply be present, because GenAI can perfectly do it.

The simplest method is to calculate an average (i.e., the total score divided by the number of criteria) and then define how this average corresponds to a course grade (see Table 2.7.1).

An alternative approach is to use logical rules to determine the grade. In this case, it is not necessary to assign numerical values to the four levels (Very good, Good, Sufficient, Insufficient). For example, an assignment may be assessed based on five criteria using these four levels; Table 2.7.2 presents a set of logical rules for converting the results into numerical grades.

Table 2.7.1.
Converting Aggregated Criterion Scores to a Number Grade

Mean (or Median)	Grade
3.6-4.0	10
3.1-3.5	9
2.6-3.0	8
2.1-2.5	7
2.0	6
1.6-1.9	5
1.0-1.5	4

Table 2.7.2
Using Logical Rules to Convert Criterion Ratings to a Number Grade

Rules	Grade
All 5 criteria are Very good	10
At least 4 Very good, no Insufficient	9
At least 3 Very good and the rest Good (no Insufficient)	8
Mostly Good (e.g., at least 3 Good), no more than 1 Sufficient, no Insufficient	7
Mix of Good and Sufficient, no Insufficient	6
Any Insufficient	5
More than 1 Insufficient	4

Validity

The assessment of a writing assignment depends largely on the instructions given to students. These instructions must explain how the learning goals of a course “translate” into clearly specified assignment goals. Writing assignments are appropriate for assessing higher-order thinking skills (from application to synthesis/creation) or critical thinking skills (see Box 2.7.1) (Hsiao et al., 2023). Lower-level linguistic mechanics such as grammar and spelling are generally not core to the learning goals of content-focused courses and are best treated as threshold requirements rather than components of graded assessment.

There should be clear descriptions of what the task is, such as solving a problem or determining the best strategy integrating multiple perspectives. In addition to the task descriptions, there should be a clear list of required elements so that students know exactly what to include in the written product to demonstrate their performance. Obviously, the more learning goals a writing assignment covers, the more detailed and fine-grained the instructions must be.

Box 2.7.1

Sample tasks for critical thinking (Ennis, 1993; Paul & Elder, 2006; Salido et al., 2025)

- Clarify problems.
- Identify significant similarities and differences.
- Recognize contradictions and inconsistencies, evaluate reasoning, and form justified judgments through reflective questioning.
- Refine generalizations and avoid oversimplifications.
- Clarify issues, conclusions, or beliefs.
- Analyze or evaluate arguments, interpretations, or beliefs.
- Identify unstated assumptions.
- Clarify and analyze the meaning of words or phrases.
- Use sound criteria for evaluation.
- Clarify values and standards while independently evaluating AI-generated information and avoiding superficial understanding.
- Synthesize information while questioning assumptions and identifying bias.
- Distinguish relevant from irrelevant questions, data, claims, or reasons.
- Evaluate the accuracy of information from different sources by verifying claims, reflecting critically, and resisting passive acceptance of AI-generated content.
- Compare analogous situations; transfer insights to new contexts.
- Make well-reasoned inferences and predictions.
- Compare and contrast ideals with actual practice.
- Discover and accurately evaluate the implications and consequences of a proposed action.
- Evaluate one's own reasoning process by analyzing, evaluating, and synthesizing information through dialogic learning.
- Raise and pursue significant questions.
- Make interdisciplinary connections.
- Analyze and evaluate actions or policies.
- Evaluate perspectives or interpretations.
- Demonstrate multidimensional skill involving contextual analysis, evidence evaluation, and methodological reasoning.

Checklist of Designing Supervised Coursework

Transparency (Include in syllabus)

- Percentage of final grade the assignment counts for is clearly stated.
- Weight of draft vs. final version (if applicable) is specified.
- Formative assessment approaches (e.g., logbooks, reflections) are described.
- Minimum passing score/grade requirement is stated*.
- Resubmission (resit) policy is explained (including grade cap, if any)*.
- GenAI tool use policy is stated*.
- Plagiarism and academic integrity policies are included*.
- Oral follow-up interview requirement is clearly mentioned (if applicable).
- Assessment form (rubric) is shared at the start of the course.
- Completed assessment form and summative feedback are shared with students after grading.

Reliability (Assessment form / rubric)

- Rubric includes clearly defined criteria with quality indicators.
- Scoring method is specified (e.g., points total or scale average).
- Conversion method from rubric score to course grade is explained.

Validity (Assignment design & instructions)

- Assignment goals are clearly linked to course learning goals.
- Assignment prompt includes a specific task (e.g., solve a problem, integrate perspectives).
- Required elements for the written product are listed (e.g., structure, evidence, reflection).
- Higher-order or critical thinking skills are targeted (e.g., analysis, synthesis).
- Instructions are sufficiently detailed to support performance across all learning goals.

*Aligned with your institutional guidelines

WTL – Invigilated Exams

As noted earlier, the thinking outcomes in the WTL approach can be evaluated using invigilated exams. When the course intends to use writing assignments for students to master the course content, then the focus is more on thinking about the course content and less on writing skills. In this case, the learning goals can be tested in invigilated exams using so-called modified essay questions (MEQs) (Elsamanoudy et al., 2024; Hayes et al., 1981). MEQs are one type of constructed-response questions.

An exam with MEQs is an assessment format designed to bridge the gap between multiple-choice questions and traditional long-essay questions. Its defining feature is an evolving situational scenario that unfolds in stages (for example, in a computer-based format where students progress through a sequence of linked questions), requiring students to apply theoretical knowledge to progressively presented circumstances. This format is specifically intended to assess higher-order cognitive skills such as reasoning, clinical or analytical decision-making, and problem-solving, rather than simple factual recall. MEQs assessing know-how have a higher level of validity and have a high impact on learning.

A MEQ typically includes:

- Scenario / Context – A realistic situation in which the respondent can apply their knowledge.
- Background / Relevant History – Key context or constraints necessary to solve the problem.
- Current Situation / Problem – Explains why the usual method cannot be used and what adjustment is required.
- Reference / Conversion Data – Provides any equivalence or conversion needed for calculations.
- Question / Task – Clearly asks for a specific answer, often numerical or actionable.

Just like constructed-response questions, MEQs can require either shorter (see Box 2.7.2) or longer answers (see Box 2.7.3).

Box 2.7.2*Example of a Shorter MEQ***Scenario:**

A 40-year-old woman manages a family-owned café that has been operating for 30 years. The café values honesty, community involvement, and sustainability.

Problem:

New staff are unfamiliar with these values, and some incidents of non-compliance occurred during a busy week.

Question:

List three strategies she could use to promote the family's values among new staff.

As for a longer MEQ, structure a MEQ as a series of sequential stages, typically following the natural progression of a consultation or project.

Box 2.7.3*Example of a Longer MEQ***Scenario:**

A 40-year-old woman manages a family-owned café that has been operating for 30 years. The family prides itself on honesty, community involvement, and sustainability. Daily operations include customer service, inventory management, and engagement with local suppliers.

Business history:

- Founded by the woman's parents and now managed by the next generation
- Strong reputation for customer relationships and community engagement
- Staff are mostly long-term employees familiar with family values, but several new employees have recently joined

Problem:

During a particularly busy week, new staff failed to follow sustainability practices, and a customer reported poor communication from a new employee. The manager wants to ensure that family values are consistently practiced across all staff.

Questions:

1. Identify three specific strategies she could implement to promote the family's core values among all staff.
2. Propose a simple method to track adherence to these values daily, and explain how it would help improve staff compliance.
3. Discuss one potential challenge the café might face in maintaining family values as it grows, and suggest a way to address it.

Transparency

Alongside the general principles of transparency, consider the provided examples of MEQ exam instructions for students (Hayes et al., 1981) (see Box 2.7.4)

Box 2.7.4*Instructions for MEQ Examination***Purpose:**

The MEQ assesses your practical approach to a developing scenario. While clear technique and structure can help maximize your score, they are no substitute for broad professional experience and competence in general practice.

Structure:

There are three papers and each paper has 10 questions. In total, there are 30 questions.

Answering:

- Write your answers briefly, clearly, and legibly in the spaces provided. Writing outside spaces is not graded.
- Follow the number of answers and words requested; do not provide more than required (i.e., if you need to give three examples, only the first three are graded).

Timing:

You have two hours to complete this exam. Work steadily and avoid spending too much time on any one question.

Scoring:

- Each page is scored independently, so answer each question fully even if it repeats information from an earlier answer.
- Scores are allocated based on both problem-solving skills and understanding theoretical knowledge. In many cases, how you handle the scenario is more important than simply recalling facts.

Progress guide:

Rough guides will indicate when you have completed approximately one-third and two-thirds of the paper to help you pace yourself.

Strategies:

- Once an answer is completed, avoid altering it unless it is absolutely necessary.
- Focus on what you would actually do in practice in the given situation. Your reasoning, decision-making, and use of theoretical knowledge are crucial to scoring well.

Reliability

One of the primary challenges of MEQs is maintaining grading consistency and minimizing subjectivity, especially in computer-based testing; however, they offer advantages for lecturers such as anonymous scoring, randomized distribution of students' work, and equal readability. To enhance **reliability**, examiners should develop highly specific model answers and detailed grading rubrics for each scoring stage. Limiting response length (for example, to 50–60 words) further promotes objective assessment and reduces variability in grading, while also decreasing the time required for manual grading. In addition, estimating the time required to complete the examination prior to administration is an essential condition for reliability, as it helps ensure alignment between question complexity and expected student performance.

Computer-based testing offers several advantages for MEQ administration, including improved security, automatic word count, enhanced image quality, and flexible pacing that allows candidates to progress at their own pace.

Validity

To be effective, MEQs should move beyond factual recall (Bloom's level 1) to assess comprehension, data interpretation (level 2), and critical problem-solving in new circumstances (level 3). Avoid the common pitfall where MEQs inadvertently become simple list-making exercises. Successful questions challenge students to justify their decisions based on the specific data provided in the scenario.

Checklist for Invigilated Exams with MEQs

Transparency

- Communicate the purpose of the MEQ to students.
- Provide clear instructions on answer format, length, and legibility.
- State total time allowed and pacing guidance (e.g., one-third/two-thirds of the paper).
- Clarify that each question/page is scored independently.
- Explain assessment criteria.

Reliability

- Develop clear model answers or grading rubrics.
- Limit expected response length to maintain scoring objectivity.
- Estimate expected completion time to align question complexity with duration.
- Use computer-based tools where possible for improved security and flexibility.

Validity

- Ensure questions assess higher-order cognitive skills: reasoning, problem-solving, and application.
- Avoid simple list-making or factual recall-only questions.
- Include tasks requiring justification of decisions based on scenarios or data.

2.8

Group Assignments³⁰

Ya-Ping (Amy) Hsiao

Introduction and Learning goals

“Groupwork” involves students working collaboratively on a particular project, assignment, or task. It entails that the **assessment** is based on the output of the groupwork and/or the groupwork process (i.e., assessment of participation in the group). Groupwork is particularly well suited to didactic approaches such as cooperative and collaborative learning. It is especially valuable when tasks are too complex for individual students to complete independently, as it enables learners to pool their knowledge and skills. Additionally, groupwork can foster creativity, encourage active listening, and support the development of ideas through interaction with others. From a practical perspective, it can also be beneficial when resources or time are limited (Francis et al., 2022).

Despite these advantages, groupwork is not universally suitable for all learning goals or assessment contexts. It may be less appropriate for activities that require sustained concentration, deep individual reflection, or high levels of perseverance, particularly when these involve complex thinking processes (Biesta, 2015). Groupwork can be used as both an *instructional* method for teaching and learning activities as well as an *assessment* method for formative and summative purposes. This chapter focuses on graded group assignments, namely using groupwork as a **summative assessment** for grading. As for using groupwork for teaching and learning activities, please consult Oakley et al.’s (2004) instructions “Turning student groups into effective teams” (2004).

Group assignments can have several advantages over individual assessment, such as promoting deep, active, collaborative, and cooperative learning. When students are taught and assessed in a manner that incorporates small-group learning activities and tasks, they acquire greater communication and teamwork skills, which they will need to demonstrate as professionals in the workplace (Oakley et al., 2004). In addition, when group assignment includes authentic tasks, students need to make multiple decisions with their group members on complex issues that require them to gain a better understanding of the context and to integrate multiple knowledge and skills from the course(s). However, group assignments also have several potential problems, such as free-riding, difficulty in assigning grades, time-consuming, etc. These problems have possible threats to assessment quality.

The aim of this chapter is to identify these threats and guide lecturers to design/revise their own group assignments. It also explains how to design group assignments in such a way that it meets the quality criteria of **validity**, **reliability**, and **transparency**.

After applying the instructions of this chapter to their own assessment practices, lecturers are able to:

- Identify possible threats of group assignments to assessment quality;

30 This chapter is adapted from the assessment handbooks developed by the Tilburg School of Humanities and Digital Sciences (TSHD).

- Design tasks appropriate for assessing program **learning outcomes** (PLOs), course learning goals, and collaboration among a group of students;
- Create assessment procedures and instructions for grading group assignments.

Transparency

When using group assignments, it is essential to clearly communicate the following in the course syllabus or assignment description:

- What will be graded? Specify whether the assessment focuses on thinking skills related to course content, group learning and process skills, or a combination of both.
- How will it be graded? Clarify whether the evaluation will be based on the group's final product, the individual contributions during the process, or a combination of these factors.
- Who will conduct the grading? Indicate whether grading will be carried out by the lecturer, through self-assessment by the students, peer assessment, or a combination.
- Resit policy: In general, every graded assessment should have an identical resit in terms of assessment type. Always consult your institutional guidelines to ensure compliance.

In addition

- Make it clear how groups are formed. To assess group assignments, it is recommended to let the course teaching team form groups based on certain criteria such as group size or diversity of skills and strengths.
- The group size (3-4 for smaller assignments and 4-5 for larger assignments) should be stated in the syllabus or assignment description.
- Provide students with a Team Policies Statement or provide sample instructions on avoiding free-riding (see Box 2.8.1).
- Do not allow students to re-form groups unless there is an acceptable reason and only with approval by the lecturer, according to the agreement in the Team Policies Statement. Make it clear what to do when any group member drops out of the course.

Box 2.8.1

Example Instructions on Group Assignment: Avoiding Free-Riding

Should any free-riding or social loafing occur during your group assignment, it is crucial to address this issue early on. In such cases, (1) address your concerns with that group member as early as possible via email, and cc the lecturer in the email. (2) Agree upon a way in which that person can contribute to the group and notify the lecturer of your resolution by email. (3) If the problem persists, make an appointment for your group to meet with the lecturer. The course team will attempt to resolve the issue together. (4) If this attempt fails, leave the person's name off the title page of the class project. (5) The course team will inform the group member that s/he will fail the team project. Note that the course team will only give a fail if steps 1-4 have been followed.

Reliability

Possible Threats to Reliability

Meijer et al. (2020) point out the possible threats to reliability when assigning a collective group grade on the group product:

[G]roup assessment might elicit behaviour such as free-riding and adopting a performance-oriented approach, which in turn might lead to, for example, unequal division of tasks, unequal contributions, and unequal responsibilities for the group assignment (Pitt 2000; Macdonald 2003; Knight 2004; Bacon 2005) (p.6).

Individual Assessment of Group Assignments

Group assignments often take more time to complete compared with individual assignments. Although group assignments require collaboration, it does not mean that students always work together either during the process or on the product. It is inevitable that students divide tasks and perform them individually (see Figure 2.8.1). Therefore, individual work/contribution can be assessed and graded. As pointed out by Meijer et al. (2020), assessing individual work in groupwork makes it “less susceptible to low achievers benefiting from high achievers or high achievers suffering from low achievers (Lejk et al., 1996; Almond 2009; Harvey et al. 2019)” (p.6). The weights assigned to individual work and the group product should reflect the time and effort that students have to spend on each.

Peer Assessment

Peer assessment, where students evaluate the work of peers in the same course, can support learning and foster collaboration, including problem solving, reflection, and other higher order skills. It is most effective as a formative tool or as input for students’ self-reflection during groupwork.

When implementing peer assessment, lecturers should train students to apply clear, well-defined criteria aligned with the course learning goals. Peer assessment often focuses on contribution, which is frequently interpreted as effort. Since effort is typically not graded, students should not assess one another on effort alone. Instead, clearly define what contribution entails. Does it refer to skills, content, or both? To the process, the final product, or both?

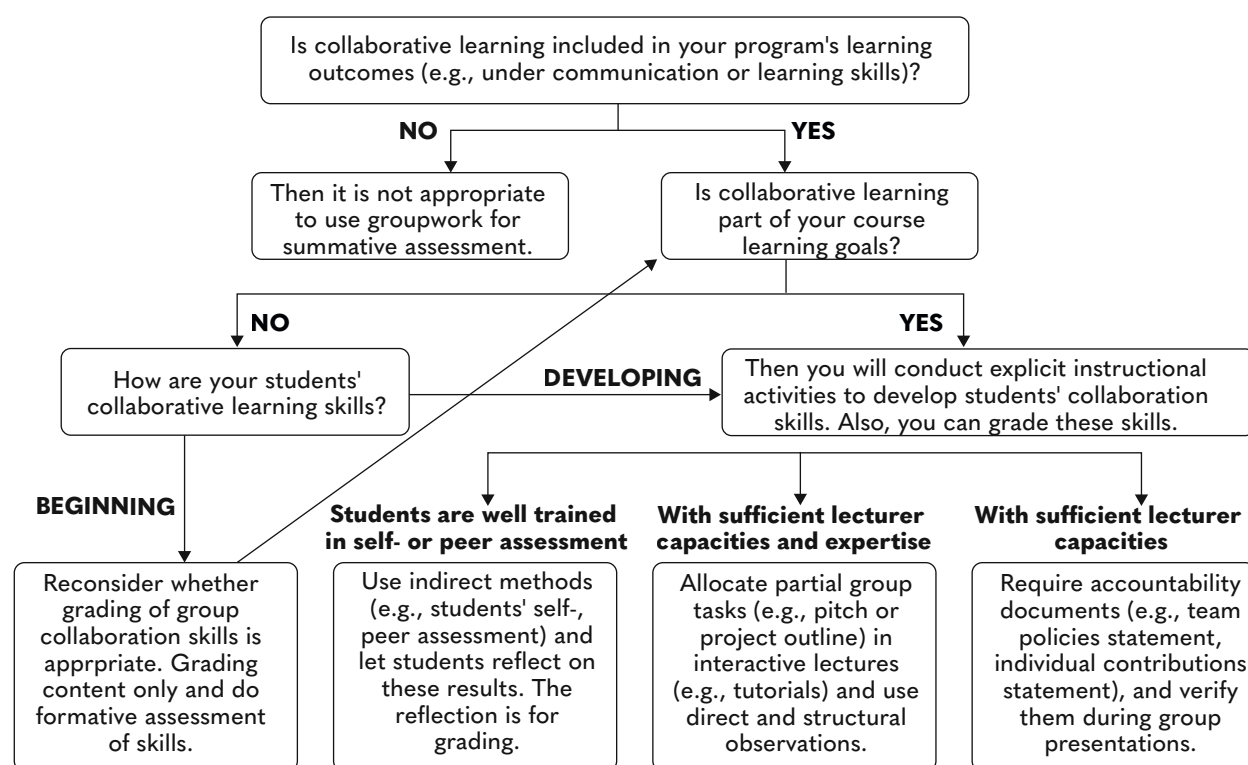
Peer assessment does not necessarily involve peer grading. Only when students are well trained to apply lecturer defined criteria, particularly for evaluating collaboration skills, can peer grades be used to inform adjustments to group grades. Even then, peer input should carry only a marginal weight in the overall course grade, according to your institutional policy.

Self-Assessment

The use of self-assessment in grading is dependent on the accuracy and reliability (or realism) of individuals’ self-judgements (Brown et al., 2015; Joughin et al., 2019), which are often unknown in the course classroom setting. Therefore, a self-assessed grade must *not* contribute *directly* towards the formal assessment determining a course grade.

Self-Reflection

Self-reflection on students’ own learning experience during the groupwork can be assessed or graded (to check individual accountability, as pass-or-fail) if the tasks assess the targeted course learning goals of groupwork. Based on the level of course learning goals, provide students with guiding questions in the self-reflection task (see Table 2.8.1).

**Figure 2.8.1**

Decision-Tree on Whether and How to Grade Group Collaboration Skills

Table 2.8.1

Guiding Questions in the Self-Reflection Task

Level	Guiding questions
Evaluate & Create/ Synthesize	<p>Based on the analysis:</p> <ul style="list-style-type: none"> How specifically might the material (and/or your prior understanding of it) need to be revised? If applicable, what additional questions need to be answered and/or evidence gathered in order for you to make a more informed judgment regarding the adequacy/accuracy/appropriateness of the material (and/or your prior understanding of it)? What should you and/or your organization do differently in the future (or have done differently in the past) AND what are the associated benefits and risks/challenges?
Analyze	<ul style="list-style-type: none"> Compare and contrast the academic material and your experience: In what specific ways are the academic material (and/or your prior understanding of it) and the experience the same and in what specific ways are they different? What are the possible reasons for the difference(s) between the material (and/or your prior understanding of it) and your experience? (e.g., bias/assumptions/agendas/lack of information on the part of the author/scientist or on your part.) In light of this analysis, what complexities (subtleties, nuances, new dimensions) do you now see in the material that were not addressed or that you had not been aware of before?
Apply	<ul style="list-style-type: none"> How does the academic concept apply to/emerge in your groupwork experience? (e.g., How did you or someone else use the material? When did you see it?)

Validity

Starting Points

Whether you should formulate a course learning goal on group collaboration skills depends on your program learning outcomes (i.e., those described in the components of Communication skills, Responsibility and Independence, Learning Skills, see Figure 2.8.1) and the course status in your program curriculum. Keep in mind that first-year bachelor's students need to develop group collaboration skills. To find out your students' teamwork and collaboration skills, consider using a questionnaire for them to self-assess these skills.

Consider student workload (i.e., 1 European Credit Transfer and Accumulation System (ECTS) credit = 28 hours). Compared to individual work, groupwork requires extra time and effort to coordinate with members in the group. Consider whether other courses also require students to do groupwork at the same time.

Course Learning Goals on Group Collaboration Skills

If you decide to develop student group collaboration skills, focus on some sub-skills (see Box 2.8.2 for examples) and design teaching and learning activities to develop them. It is not manageable to develop all sub-skills in one course.

Box 2.8.2

Collaborative Problem Solving Skills Summarized from Hesse et al. (2015)

Social process skills

- Participation skills
 - Action: The general level of participation of an individual, irrespective of whether this action is in any way coordinated with the efforts of other group members.
 - Interaction: Behavior that demonstrates interaction with and responses to others.
 - Task completion: Motivational aspects of participation and consequent perseverance on a task.
- Perspective taking skills
 - Responding skills become apparent when group members manage to integrate the contributions of other members into their own thoughts and actions. For instance, members who rethink a problem based on evidence that was reported by another member exhibit a high degree of responding skill. In contrast, ignoring contributions from others exemplifies a low degree of responding skill.
 - Audience awareness skills are constituted by the ability to tailor one's contributions to others.
- Social regulation skills
 - Meta-memory (Self-evaluation): Recognizing one's own strengths and weaknesses.
 - Trans-active memory: Recognizing strengths and weaknesses of others
 - Negotiation: The skills to achieve a resolution or reach compromise.
 - Responsibility initiative: The responsibility that a member experiences for the progress of the group.

Cognitive process skills

- Task regulation skills
 - Planning: Problem analysis, goal setting, resource management, and complexity.
 - Executing and monitoring processes:
 - Information collection: The ability to identify what information is required and how and when it can be acquired.
 - Systematicity: The level of sophistication of the strategy that a group member's exhibits.
- Learning and knowledge building skills

- Learning is indicated by the ability to identify and represent relationships, understand cause and effect, and develop hypotheses based on generalizations. A low level of learning skills would be evident if the only knowledge that is extracted from a problem solving activity stems from information that was directly provided through instruction.
- Knowledge building is exemplified by the ability to take up ideas from group members to refine problem representations, plans, and monitoring activities. The highest level of knowledge building occurs in those group members who are able to integrate and synthesize the input from other group members in the description and interpretation of a given problem.

Accountability

When accounting for groupwork (e.g., using an accountability table such as course assessment overview), note that it is sufficient to indicate the highest level(s) on which the targeted course learning goals are assessed because coursework is assessed based on a set of assessment criteria, which does not always correspond to a taxonomy. To account for validity, indicate the relationship between the targeted course learning goals and criteria (i.e., the criteria should be relevant to the goals) either in the accountability table or in the assignment descriptions.

If individual work is graded, make it clear in the cells how the targeted course learning goals are assessed as well as whether this individual grade is part of the collective grade or not.

Assignment Descriptions

Groupwork should be assessed at the higher cognitive levels (from application to synthesis/creation).

The task should have the following characteristics (Michaelsen & Sweet, 2008). The task

- ensures individual accountability;
Note: When individual work is graded, it should also assess the higher or highest levels of the targeted course learning goals.
- requires intensive interaction among group members;
- is so complex that collaboration is essential and individual students cannot complete it.

For example, students

- take different roles to discuss an issue;
- use different concepts to solve a discovery-oriented problem;
- are exposed to opposing views on a subject, and then have to resolve the conflicts in the process of making a decision.

Integrate Groupwork with Interactive Lectures

The instructional and assessment design of groupwork typically involves multiple stages (see Figure 2.8.2). A valid and fair design should account for both individual and group contributions, as well as the process and the final product. In higher education, where blended learning is commonly used, it is essential to carefully sequence on-campus and online activities to support learning effectively.

The design should encourage diverse forms of interaction, including:

- Student–content
- Student–student
- Student–lecturer
- Student–student–lecturer

These interactions are essential for promoting higher-order thinking skills and integrating self-assessment, peer assessment, and lecturer assessment.

In the age of GenAI, a key responsibility of students is the critical evaluation of AI-generated content. This demands active lecturer monitoring and constructive feedback to ensure responsible and meaningful learning.

When the group assignment has clearly a clearly defined sub-tasks, these sub-tasks can be assessed for feedback, set for compulsory completion, or graded (i.e., individual contributions), as long as how interim assessment is clearly stated in the course manual or assignment descriptions.

To assess higher-order thinking skills, the group product should require group members to make a specific decision (e.g., on the solution to a problem) supported by the justifications why such a decision is made.

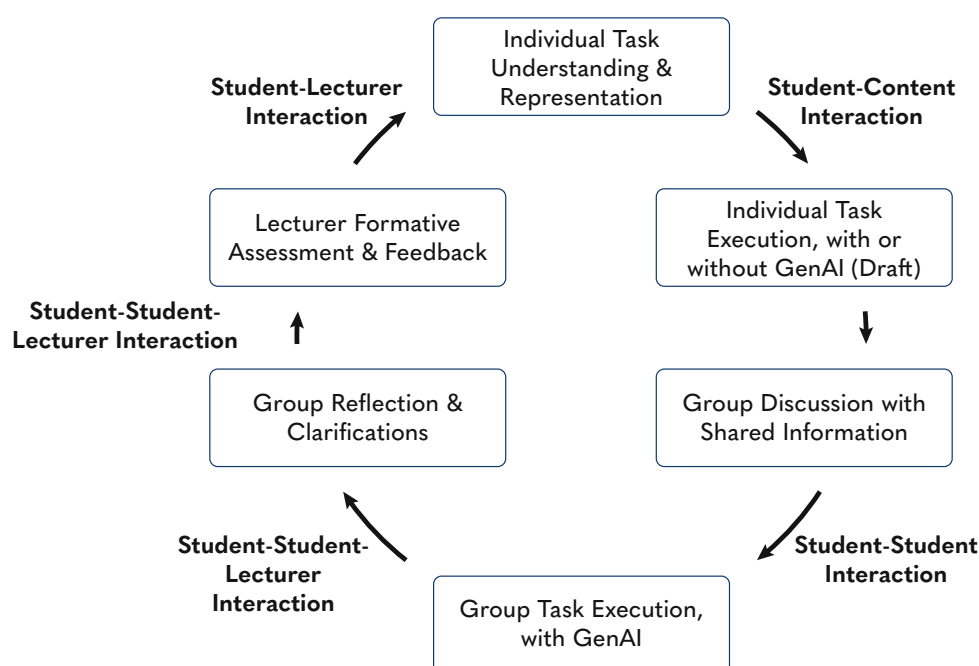


Figure 2.8.2

Multiple Blending Approaches of Groupwork (Bezcioğlu-Göktolga et al., 2025)

Possible Threats to Validity

Meijer et al. (2020) point out the possible threats to validity when assessing groupwork:

Because of the combination of within-group student differences and a lack of individual accountability, students with higher individual grades for individual assignments (e.g., written tests) might receive relatively lower group scores or grades for group assignments, whereas students scoring lower on individual assignments might receive relatively higher group scores or grades for group assignments (Lejk et al, 1999; Almond 2009; Plastow, Spiliotopoulou, and Prior 2010; Moore and Hampton 2015; Harvey et al. 2019). In addition, it has been argued (Kagan 1995; Pitt 2000) and empirically shown (Bacon et al, 1998; Lejk et al, 1999) that between-group differences might lead to situations in which students who are similar in their domain-specific ability might receive a different group assessment. That is, some students are placed within an ‘advantageous’ group with, for example, multiple highly motivated and/or high-achieving students, whereas others are placed within a ‘disadvantageous’ group with, for example, multiple poorly motivated and/or low-achieving students (p.4).

Limited Course Grade on Groupwork Assessment

The policy solution to counter the threats to validity is to limit the weight of groupwork assessment in making up the total grade of a course (Ahern, 2007; Augar et al., 2016). Consult your institutional guidelines on this.

Checklist of Designing Group Assignments

Transparency

Assessment Information

- Activities structured in stages (process + product)
- Clearly state what is graded
- Clearly explain how it is graded
- Clearly indicate who grades
- Criteria and standards shared with students
- Deliverables and sub-tasks clearly described
- Feedback integrated throughout the process
- Interim assessment (if any) clearly explained
- Students self-assess and reflect on their learning
- Weighting of all components specified
- Resit policy defined (i.e., same assessment type as original, institutional rules followed).

Group Organization

- Conditions for excluding a non-contributing member defined
- Conflict resolution steps communicated
- Free-riding policy clearly explained
- Group formation method explained (preferably lecturer-assigned)
- Group size specified (3–4 small / 4–5 large tasks)
- Procedure for member dropout clarified
- Rules for changing groups defined
- Team Policies Statement provided

Reliability

- Assessment applied consistently across groups
- Avoid over-reliance on group grades (i.e., groupwork has limited weight in final course grade)
- Interim feedback ensures alignment during process
- Measures to prevent free-riding
- Measures to avoid unequal task distribution
- Safeguards against social loafing
- Standardized criteria/**rubrics** used
- Measures to reduce group composition bias

Validity

- Assessment methods match learning activities
- Assignment aligned with course/program learning outcomes
- Both process and product are assessed
- Clear link between learning goals and assessment criteria
- Collaboration skills included only if relevant

- Students engage in meaningful interaction
- Task targets higher-order thinking
- Task requires decision-making, justification with evidence, integration of multiple perspectives
- Task requires real collaboration (cannot be done individually)
- Task involves complex, real-world or discipline-relevant problems

2.9

Simulation-Based Assessment

Ya-Ping (Amy) Hsiao

Introduction and Chapter Learning Goals

Simulation is an activity that “imitate[s] a system, entity or process” (Lean et al., 2006, p.228). It is an active and student-centered didactic approach and is featured by three aspects (Ellington et al., 1998; Hertel & Millis, 2023):

- Context-driven decision making: Students make decisions and take actions based on the realistic scenario and professional setting they are placed in. The authenticity of the context allows them to behave as they would in actual practice. For example, in a marketing simulation, students take on the role of brand managers who must decide how to allocate a limited budget across advertising channels based on market research, consumer trends, and competitor activity.
- Consequences of decisions: Each decision leads to specific, traceable consequences that reflect real-world outcomes. This dynamic element makes the simulation responsive and educational. For example, a psychology student incorrectly diagnosing a simulated patient with depression instead of bipolar disorder may lead to inappropriate treatment suggestions, which the simulation highlights through patient feedback and worsening symptoms.
- Reflective adaptation: After observing the consequences of their decisions, students reflect and adapt their future actions. This encourages critical thinking, professional judgment, and continuous learning. For example, students reflect on poor quarterly performance in a business strategy simulation, revise their pricing strategy, and restructure their team for the next round.

Simulation-based **assessment** is a method of evaluation that uses realistic, non-taught scenarios to immerse participants in context-rich situations (so that students build relationships and make connections between different pieces of information) where they perform professional roles with intent and authenticity (opposite to the conventional coursework assignments that students simply complete a task outlined in the rules). By mimicking real-world tasks and decision-making processes, this approach provides examiners with an innovative and authentic means to evaluate how effectively students apply their knowledge, skills, and abilities in practice (dos Santos & Gomide, 2013; Neely & Tucker, 2012; Wilson et al., 2024). Examples include legal courtroom simulations (moot court), teaching simulations (micro-teaching), business management simulations in virtual company environments, and cybersecurity incident response simulations.

After applying the instructions of this course to their own assessment practices, lecturers are able to:

- Differentiate simulation-based assessments from traditional coursework
- Design effective and engaging simulation-based assessments
- Anticipate and address common implementation challenges

Starting Points

Simulation-based assessments are particularly well-suited for evaluating students' development of competence in complex, dynamic systems—those characterized by features such as multiple feedback loops, “time delays, non-linear relationships, and accumulations” (Capelo et al., 2021, p.2). To ensure that these simulations enhance rather than hinder learning, it is essential for lecturers to carefully evaluate and manage this complexity during the design and implementation process. In particular, greater simulation complexity often requires more explanation, more structured guidance, and more time. As such, managing all learning and assessment activities within the allocated study load (based on the number of granted European Credit Transfer and Accumulation System/ECTS credits) is something lecturers **must** carefully consider before adopting this demanding assessment method. In addition to time considerations, the successful implementation of simulation-based instruction and assessment also depends heavily on the lecturer's didactic and assessment expertise. For **novice** lecturers, we recommend seeking guidance through resources like this guide or by collaborating with more experienced colleagues who have prior experience with simulation-based approaches.

Complexity

A critical starting point in deciding whether to use simulation in your course is to evaluate the complexity of the simulation environment and its demands on students' cognitive processing. Research shows that students often struggle at multiple stages of the inquiry process when working within complex simulations. According to Cognitive Load Theory (Sweller et al., 2011), overly complex simulation models can exceed the limits of students' working memory, especially when multiple elements must be processed simultaneously. This is particularly relevant in simulation-based learning environments that involve a high degree of interactivity and dynamic information (Naismith et al., 2019).

Given these challenges, it's important for lecturers to consider whether the intended learning goals can be achieved without overwhelming students. If a simulation is indeed suitable, instructional design should proactively address potential cognitive overload by integrating cognitive supports, such as debriefing, scaffolds, prompts, or embedded guidance, which helps structure students' activities and focus their attention (de Jong et al., 2018).

Given the high complexity of tasks, groupwork is well suited for sharing demanding processes among students (Evans & Kerridge, 2021). However, the coordination involved, often referred to as transactional costs (Kirschner et al., 2009), must remain within a manageable study and cognitive load. If the simulation-based assessment is designed as a simulated group activity with space for individual reflection on the process, it can offer the full range of intended benefits. Still, achieving the right balance in this mix is essential.

Transparency

Simulation activities and assessment information should be standardized in terms of structure, task design, and instructions to ensure consistency and **fairness**. It is advisable to consult with your institution's assessment specialist to review the simulation setup and assessment design for alignment with best practices and quality standards.

A debriefing session, typically conducted as a facilitated or guided reflection within the experiential learning cycle, is structured around targeted questions that encourage students to critically examine their actions, decisions, and outcomes during the simulation. In the context of simulation-based assessment, debriefing is essential for promoting assessment **transparency**. It provides students with a clear understanding of how their performance aligns with the program intended learning outcomes and assessment criteria (see Table 2.9.1).

In many cases, students engage in various combinations of pre-task, in-task, and post-task discussions, all aimed at helping them grasp the structure and behavior of the simulation model (Qudrat-Ullah, 2014). Post-task debriefing

is essential as it makes the lecturer's assessment process more visible by clarifying what is assessed, why certain decisions matter, and how success is measured. This openness not only enhances student learning but also builds trust in the fairness and purpose of the assessment process (Grund & Schelkle, 2020).

Finally, transparency does not mean revealing the full contents of simulation activities in advance. This may have limited didactic value, as it can influence students' decisions and actions in ways that reduce the authenticity of their engagement. When students know the exact challenges or outcomes beforehand, they may tailor their responses to fit expected answers rather than demonstrate real-time application of knowledge and reasoning.

Using the same example in a psychology course, if students are told in advance that the client is in the 'contemplation stage' of change and will resist action-oriented advice, they may prematurely structure their interventions around this information rather than practicing how to assess the client's stage and adjust strategies accordingly. This limits the development of essential skills such as client-centered assessment, critical thinking, and adaptive decision-making.

Table 2.9.1

Debriefing Activities Aligned With Simulation-Based Assessment

Phase	Purpose	Transparency Focus
Pre-Task	Prepare students and clarify expectations	<ul style="list-style-type: none"> When using an online simulation environment, provide students with an introductory information session to help them become familiar with the platform, understand the goals of the simulation, know what to expect from the platform and lecturers (both theory or/and technical support) during the experience, and learn what they need to prepare before the simulation begins. Makes learning goals and assessment criteria clear before assessment begins
In-Task	Guide real-time reflection and decision-making	<ul style="list-style-type: none"> Facilitates and monitors the process, without making decisions for students Prompt students to connect theory to action while being assessed
Post-Task	Reflect on performance and understand lecturer assessment	<ul style="list-style-type: none"> Clarifies how the assignments are judged, linking performance to criteria

Reliability

The participation process in simulation activities should be assessed formatively, with a strong emphasis on providing constructive feedback and addressing errors in thinking and practice.

As with other types of coursework assessments, simulation assignments should be graded based on a clearly defined set of criteria and standards to ensure **reliability**. In addition to meeting the **validity** requirement that the criteria align with the intended course learning goals and the assignment content, multiple examiners must share a common understanding of these criteria. This includes mutual agreements on aspects such as the integration of different types of knowledge (e.g., procedural and metacognitive knowledge), consideration of contextual factors, the complexity of decision-making, and an awareness of the unintended consequences arising from decisions made during simulation activities.

A key challenge to ensuring reliability in simulation-based assessment is observer bias, particularly when examiners are familiar with the participants. This familiarity is often unavoidable and even necessary, because in simulations, lecturers are typically expected to observe and support student engagement throughout the activity. However, this dual role can introduce subjectivity into the assessment process. To address this, it is essential to establish clear, objective assessment criteria and ensure that examiners are trained to apply them consistently. Assessment tools, such as scoring **rubrics**, should be pilot tested with checklists or holistic rating scales on student participation in simulation activities (Alinier & Oriot, 2022). Then, based on providing feedback and student performance in the process, these observational tools can be modified into the final grading rubric.

Validity

If the simulation element is only a relatively small element for learning goals within a course, then it may not be suitable for **summative assessment** (Usherwood, 2018).

A clearly defined learning goal is essential for designing effective simulation-based courses and assessments (Evans & Kerridge, 2021; Wilson et al., 2024). Clear goals help students understand what they need to do to succeed, and they also guide you, as the lecturer, in determining what is being assessed and how to assess it effectively. In particular, clearly defined learning goals are crucial for ensuring assessment validity when selecting appropriate grading methods and tools, such as rubrics, and when deciding whether to use an analytic, holistic, or interactive assessment approach (Cook et al., 2014).

For example, in a psychology course, a learning goal might be for students to apply psychological theories of behavior change (e.g., the Transtheoretical Model, Theory of Planned Behavior) to design and adapt intervention strategies for a simulated client facing motivational barriers to health-related behavior change. This goal clarifies both the purpose of the simulation and the criteria for assessment.

Following the principle of **constructive alignment** (Biggs & Tang, 2011), the design process should begin with a clear articulation of the learning goals what students are expected to understand or be able to do by the end of the activity. Once these goals are established, you should define the specific actions, decisions, or problem-solving processes students will undertake in the simulation to demonstrate their achievement.

However, in practice, lecturers often focus more on what they want the simulation technology to deliver (e.g., recreating a clinical environment or presenting a variety of client profiles) rather than on what they want students to do with the simulation to meet learning goals (Evans & Kerridge, 2021).

Returning to the earlier example, a well-aligned simulation in a psychology course would require students to assess a virtual client's readiness to change, choose and apply appropriate intervention strategies (e.g., motivational interviewing, goal setting), and adjust their approach based on simulated feedback from the client. Students could then reflect on their decision-making process, citing relevant theories to justify their actions.

Simulation-based assessment can help develop a culture of integrity and accountability by emphasizing transparency, honesty, and ethical responsibility in both learning and professional contexts (Buxton et al., 2015; Wilson et al., 2024). By embedding ethical considerations into scenarios and assessment criteria, lecturers can reinforce the expectation that ethical behavior is essential for workplace success. Lecturers/Examiners also play a key ethical role by ensuring fairness, confidentiality, and support while modeling ethical decision-making throughout the simulation-based assessment process.

Determine which assignments students will need to complete and submit after participating in the required simulation activities. The assignment tasks should be *directly linked* to the simulation activities; otherwise, students may be able to complete the tasks without actually engaging in the simulation.

Following the same psychology example: assignments could include a structured intervention plan, where students must analyze specific verbal and non-verbal cues from the virtual client to identify the client's motivational state and design appropriate interventions accordingly. Another option is a reflective journal, which requires direct reference to the student's actions during the simulation and the simulated client's responses. Without participating in the simulation, students would not be able to provide the necessary analysis or insights.

When a simulation is extended, either in duration or in its importance within the course or module, the topics covered in the simulation can also be assessed through a conventional invigilated exam. For example, a psychology simulation running over several weeks allows students to explore behavioral dynamics and relate their experiences to the course's theoretical readings. In this way, simulation serves as a primary method for delivering essential knowledge and skills. The subsequent exam can then assess students on the same theoretical and applied concepts they encountered during the simulation, ensuring they have effectively integrated their experiential learning with academic content.

Checklist for Designing Simulation-Based Assessments

Starting Points

- Is simulation essential for achieving the course's learning goals?
- Is the scope appropriate for the ECTS credits and student workload?
- Do lecturers have the necessary pedagogical and assessment expertise or access to support?

Transparency

- Have students been briefed on goals, expectations, and how they will be assessed?
- Is a post-simulation debrief included to explain assessment decisions and promote reflection?
- Are supports in place to prevent overload (e.g., debriefing, prompts, scaffolds)?
- If using groupwork, is coordination manageable within the study load?

Reliability

- Are the assessment criteria and scoring methods standardized and pilot-tested?
- Have examiners been trained to apply the criteria consistently?
- Is observer bias minimized through structured tools and shared understanding?

Validity

- Are course learning goals clearly defined and observable through simulation tasks?
- Are assessment methods (e.g., rubrics, assignments) aligned with these goals?
- Do students need to engage in the simulation to complete the assignment?
- Does the simulation's complexity match students' cognitive capacity?

2.10

Portfolio Assessment³¹

Ya-Ping (Amy) Hsiao

Introduction and Learning Goals

In general, a portfolio refers to a collection of artifacts (i.e., pieces of evidence, entries for a portfolio). These artifacts can be written, visuals, audios, videos, products, etc. Everyone can make a personal portfolio containing materials for a certain purpose (i.e., administration, entertainment, memories), but professionals make a *purposeful* selection of these artifacts to record their *best* performance for the intended audience of their portfolios (i.e., to showcase) and to reflect on learning experience.

In terms of portfolio **assessment**, we refer to a *professional* portfolio that is made for assessment purposes. A portfolio can be used for both formative or **summative assessments** or for either one. If a portfolio serves both purposes, then it is both a learning and assessment portfolio. If it is formative (to document progress and to get feedback from others), it is a learning portfolio. If it is summative for students to demonstrate their best performance, it is an assessment portfolio and it can be used for grading.

After applying the instructions of this chapter to their own assessment practices, lecturers are able to:

- Explain the differences between conventional coursework and portfolio assessment;
- Define assessment criteria and performance standards based on the targeted course learning goals;
- Design tasks appropriate for the course learning goals and assessment criteria;
- Write clear assignment instructions for students to compile the evidence and to reflect on their own performance.

Why Portfolio Assessment?

Portfolio is defined as “a purposeful collection of student work that exhibits the student’s efforts, progress, and achievements in one or more areas. The collection must include student participation in selecting contents, the criteria for selection, the criteria for judging merit, and evidence of student self-reflection” (Paulson, 1991, p.60). Portfolio assessment can be tailored to each student’s individual needs because it allows students to formulate their personal learning goals, evaluate their own work, select their best work, and justify this choice. This flexibility suits student-centered education (i.e., university didactic framework).

In addition, the conventional forms of assessment often test only a narrow range of knowledge and skills. Portfolio assessment emphasizes authenticity: students perform tasks they encounter in their professional or daily lives, such as solving real-world problems. Solving these problems require students to integrate a *wide* range of knowledge, skills, and attitudes (character building). Moreover, students often need multiple attempts, feedback, corrections, and revisions to be able to deliver satisfactory products or achieve a certain level of proficiency. This authenticity

³¹ This chapter is adapted from the assessment handbooks developed by the Tilburg School of Humanities and Digital Sciences (TSHD).

strengthens students' employability and is in line with the institutional educational vision. The differences between conventional coursework and portfolio assessment are summarized in Table 2.10.1.

Table 2.10.1

Major Differences Between Conventional Coursework and Portfolio Assessment (Biggs & Tang, 1998; Dwyer, 2001)

Conventional coursework (e.g., Chapter 2.7 of this practical guide)	Portfolio assessment
<i>Lecturers ...</i>	<i>Students...</i>
set the course learning goals (CLGs) or relate specific assessment purposes to the CLGs;	formulate personal CLGs curated by lecturers, which are specific to their own topic of interest or context, provided that they relate their personal CLGs to the course CLGs;
define the quality criteria and performance standards for assessing the work;	co-define or actively make sense of the criteria and standards, provided that
list the required evidence that would be relevant for students to submit for assessment against those criteria;	interpret the eligibility criteria defined by lecturers for acceptable evidence or a list of recommended evidence; autonomously select evidence for submission and justify how each piece of evidence is relevant to the CLGs and eligibility criteria;
make a judgment about the extent to which these criteria have been met.	reflect on the process and experience of the selected work; self-assess the quality of their selected contributions using the assessment criteria/ rubrics that will be applied by their examiners to grade their portfolios.

While giving students more autonomy to determine their personal learning goals and select evidence, portfolio assessment requires students to make evaluative judgements, “the ability to assess a piece of work (one’s own or that of others) while attending to the context, quality, standards and criteria built upon previous experience” (Panadero & Broadbent, 2018, p.2). Most portfolio assessment practices require students to include evidence of self-assessment and reflections that are written commentaries accompanying the samples of the work (Baume, 2001). These reflective commentaries focus on what is learned in the process, which steps are carried out, what could be changed in the future, and what the next steps in the process will be (Van Wyk & Carl, 2010). Without self-reflection, “the portfolio is no more than a notebook filled with meaningless pages” (Gillespie et al., 1996, p.48).

What Is Not Portfolio Assessment

From observations of several evaluated courses in our contexts, it can be concluded that portfolio assessment has been regarded as simply assessing a collection of a student’s work. For example, one course used a portfolio of shorter assignments written during the course, but these assignments with short answer questions were meant to assess students’ knowledge and comprehension of the core theories. Although they are aligned with the targeted course learning goals, these standardized assignments neither allowed students a choice in selecting their own work nor engaged students’ personal experience (lacking authenticity).

Formative and Summative Portfolio Assessment

Formative portfolio assessment provides students opportunities to demonstrate their learning growth over a given time and use the portfolio contents as a basis for discussions with the lecturer about their understanding of concepts, principles, and theories that underlie the work. The lecturer guides the discussion so that students use the same thinking strategies and abilities used by professionals in the field or discipline (Baume, 2001; Nitko & Brookhart, 2013). For example, in writing portfolio assessments, students practice the iterative process of writing and compiling a range of artefacts enabling them to summarize what progress has been made and reflect on their learning improvement (Lam, 2018).

Summative portfolio assessment (i.e., showcase portfolios) is used to assess a limited collection of the student's best work(s) (Mertler, 2016; Nitko & Brookhart, 2013).

Both formative and summative portfolio assessment share the following three features in the frequently cited literature (Biggs & Tang, 1998; Fernsten & Fernsten, 2005; Nitko & Brookhart, 2013; Palm, 2008):

1. Design authentic tasks that are directly meaningful to the course learning goals. Students are expected to construct knowledge when working on real-life with activities performed in actual, authentic settings with real consequences or simulation tasks with activities that imitate real-world situations in a controlled or artificial environment (see Chapter 2.9). The work that students include in their portfolio must provide a direct opportunity for them to engage in the types of thinking and skills typically used by those working in the field or discipline.
2. Record conceptual development. Portfolio entries must record students' own explanations, understandings, and conceptual frameworks. Students should engage in the use of theoretical knowledge to go beyond that knowledge, establish relationships between pieces of this knowledge to construct in-depth understanding around a reasonably focused topic, relate to a higher order principle, generalize from principles to new data, apply metacognitive understanding by reflecting on one's own effectiveness in terms of the theories learned, solving problems far from the domain in which learning occurred, and express their conclusions through elaborate communication.
3. Engage in reflective activity. Reflections require students to articulate and review the process and/or products of their portfolio artifacts. Students make time and effort to analyze their work in relation to the assessment criteria and performance standards, determine growth and needs, and engage in self-assessment to what extent they have attained the course or personal CLGs. To cater diverse student needs and learning styles, portfolio assessment should allow students to do written, audio, or video reflections (e.g., using Canvas e-portfolio, metapholio³²).

Summative portfolio assessment

Transparency

The same suggestions on **transparency** of coursework assessments are applicable to portfolio assessment. Include the following aspects in the assignment descriptions:

- State clearly why portfolio assessment is used and the purpose (formative, summative, or both).
- List the criteria with definitions and performance standards for students. Use checklists, rating scales, and rubrics based on the purpose. For example,
 - Ask students to use a checklist to examine the eligibility or completeness of their evidence.
 - Advise students to self-assess their work using a rating scale before writing self-reflections.
 - Use the rubrics to grade the quality of evidence and self-reflections.
- Instruct students to organize evidence based on learning goals or criteria. Ask students to make an overview that shows cross references among learning goals, criteria, and the title of artifact entries (with page numbers) (see Table 2.10.2 for an example).

Table 2.10.2

Template Overview Cross-References Among Course Learning Goals/Criteria and Artifact Entries

Title artifact entries (with page number)	CLG1/ Criterion 1	CLG2/ Criterion 2	CLG3/ Criterion 3	CLG4/ Criterion 4	CLG5/ Criterion 5

32 <https://metapholio.ch/>

- Specify what counts as artifacts, whether students can include both primary artifacts produced by the students themselves and secondary artifacts produced by others (e.g., peer assessment based on peer review), and whether both types of artifacts are graded and in what way.
- A portfolio can quickly become a mess of materials and papers that are difficult to assess. To improve the situation, each entry should have an appropriate entry caption containing the information stated in Box 2.10.1:

Box 2.10.1*Contents of An Entry Caption*

- Title of the entry
- Date and context of the entry
- Who created the entry? Students themselves or others?
- Indication of the learning goals or criteria
- Justifications on why this entry is included. Provide students with the following prompt questions:
 - Why is this artifact better than other artifacts that I could have chosen?
 - Does this artifact provide evidence of my success or growth against one or more learning goals?
 - Can I rationalize the importance of this artifact to those viewing it out of context?
 - Is this artifact a result of my professional growth and accomplishments?
 - How does this artifact reflect who I am as an independent learner?
- Specify the weighting, if
 - some learning goals are more important than others;
 - reflections are graded (and should be weighted less than evidence);
 - secondary artifacts (e.g., peer assessment) are graded (and should be weighted less than primary artifacts).
- Specify a range of size limits:
 - minimum-maximum numbers or pages of evidence;
 - minimum-maximum number of words in self-reflection.

Reliability

The requirements of **reliability** are the same as coursework assessment types. There should be a set of well-defined criteria or indicators, performance standards/levels (e.g., excellent, good, sufficient, and insufficient), grading instructions for examiners to grade student work consistently (see Box 2.10.2).

- The performance levels/standards should be manageable for examiners. Using more than 4 levels makes it difficult to classify student work. Use a logical rule to convert criterion scores (e.g., 1-4) to a number grade (i.e., 1-10).
- For portfolio assessment at the graduation level, two examiners are required for grading to prevent rater effects. It is also important to have a shared understanding of the meaning of the criteria and performance standards. Consider a calibration session to scrutinize the examiners' grading practices or establish and share anchor exemplars of past student work on each level.

Box 2.10.2*Additional Information for Examiners to Assess Portfolios*

Portfolios are examined based on an interactive approach. Examiners first check whether bottom-up application of individual judgements on outcomes “adds up to” their overall, holistic, judgement. If they match, then all is well. If not, examiners may modify their holistic judgement to match the judgements of the bottom-up approach; or they may alter individual judgements on outcomes to match their holistic judgement or make a combination of these two kinds of changes. This suggests that a process of negotiation between analytic and holistic assessment could, over time, lead to two very different but compatible accounts of assessment criteria for the portfolio, one account holistic, the other based on the assessment of individual outcomes.

The examiners first review the evidence, then the justifications, and finally validate evidence and justifications based on the placement advisory report.

Validity

It is important to apply the didactic principle of **constructive alignment** (Biggs & Tang, 2011) to ensure the **validity** of portfolio assessment.

Constructive

“Constructive” refers to constructivist learning theories that emphasize students’ active roles in constructing their own knowledge.

Producing a portfolio requires students to relate the evidence to the CLGs, collect or produce the evidence, and then annotate, label, and organize it. They then need to review the evidence critically, to discover and explain what the evidence shows about their attainment or/and development. Thus, students need to actively engage in production, selection, critical judgement, and reflection (Baume, 2001; Venkatesh et al., 2013).

Alignment

“Alignment” refers to the consistency that optimizes the likelihood of students engaging appropriately with teaching and learning activities as well as assessment tasks.

- Is portfolio assessment an appropriate assessment type for your course learning goals?
Based on the three features, portfolio assessment is appropriate for assessing higher-order thinking skills. It is less appropriate for assessing lower-order thinking skills that can be assessed by **formative assessment** types (e.g., quizzes, summaries) or the examination method (as interim).
- To what extent are the following aspects aligned with each other?
CLGs, criteria, required evidence, and reflection tasks should be aligned with each other. See Table 2.10.3 for an example.
 - CLGs state what should be assessed.
 - Criteria define the level of performance on the learning goals should look like. One CLG can be assessed by multiple criteria.
 - Required evidence should provide students with sufficient opportunities to demonstrate their performance. It is thus important to determine a range of artifacts (e.g., 2–4 pieces/types) needed to judge student performance on each learning goal and criterion.

- Reflection tasks should require students to critically demonstrate analytical, evaluative, and synthetic thinking skills. Descriptive reflections are not critical. Consider using a reflective model to guide students towards critical reflections³³.
- What should be graded? Evidence, reflections, or both?
Assess the evidence that shows whether students have attained the CLGs of the course. The evidence should be graded with substantial weight. Assign a limited weight for grading the quality of the reflections or use the reflections to check the authenticity of the evidence without assigning a grade.

Table 2.10.3*An Aligned Example*

Aspects	Example
CLG	To demonstrate domain-specific content knowledge.
Criterion	The student intern demonstrates current and accurate knowledge of subject matter covered in the curriculum.
Evidence	Recommended evidence <ul style="list-style-type: none"> • Course syllabus • Unit lesson plans • Integration of multiple sources on a content subject
Reflection	Reflective commentary based on Gibbs' Reflective Cycle (Coffey & Gibbs, 2001) <ul style="list-style-type: none"> • Description of the experience • Feelings and thoughts about the experience • Evaluation of the experience, both good and bad • Analysis to make sense of the situation • Conclusion about what the student learned and what could have done differently • Action plan for how the student would deal with similar situations in the future, or general changes you might find appropriate.

Checklist of Designing Portfolio Assessment

Transparency

- Is the purpose of portfolio assessment clearly stated (e.g., formative, summative, or both)?
- Are criteria, performance standards, and definitions shared with students in advance?
- Are students instructed to organize evidence by learning goals or criteria?
- Is a clear overview structure required (e.g., cross-referencing table showing learning goals, criteria, artifact titles, and page numbers)?
- Are students guided on what counts as primary and secondary artifacts?
- Is the grading weight of evidence, reflections, and secondary artifacts clearly specified?
- Are students advised to self-assess using a checklist or rating scale before reflection?
- Are rubrics used to assess evidence and (if graded) reflections?
- Are there clear size/length limits for evidence and reflections (e.g., min-max pages or word counts)?
- Does each entry require a caption/summary with relevant metadata (as described in Box 2.10.1)?

Reliability

- Are there well-defined assessment criteria and performance standards/levels (e.g., excellent, good, sufficient, insufficient)?

33 <https://www.ed.ac.uk/reflection/reflectors-toolkit/all-tools>

- Are the number of performance levels manageable (ideally 3–4)?
- Is there a grading instruction/rubric to ensure consistent evaluation by examiners?
- Is a logical rule used to convert criterion scores to a final grade (e.g., 1–4 to 1–10)?
- Are two independent examiners assigned for graduation-level assessments to reduce bias/rater effects?
- Has a calibration session or discussion occurred among examiners to standardize grading interpretations?
- Are anchor exemplars of student work provided for each performance level?

Validity

- Are the CLGs clearly defined?
- Does the portfolio require students to actively engage in evidence selection, annotation, organization, and reflection?
- Is the portfolio assessment appropriate for the cognitive level of the CLGs (e.g., higher-order thinking skills)?
- Are learning goals, assessment criteria, evidence, and reflection tasks all aligned with one another?
- Are reflection tasks designed to encourage critical, rather than descriptive, thinking (analytical, evaluative, synthetic)?
- Is there guidance on using a reflective model to support student reflection?
- Are multiple artifacts (2–4) specified per learning goal or criterion to demonstrate achievement?
- Are reflection tasks aligned with the evidence and course outcomes?

2.11

Reflection and Logbook Assessment

Ya-Ping (Amy) Hsiao

Introduction and Chapter Learning Goals

Reflection is “the conscious examination of past experiences, thoughts and ways of doing things. Its goal is to surface learning about oneself and the situation, and to bring meaning to it in order to inform the present and the future. It challenges the status quo of practice, thoughts and assumptions and may therefore inform our decisions, actions, attitudes, beliefs and understanding about ourselves” (University of Edinburgh, 2024).

Reflection (Branch & Paranjape, 2002) includes “consideration of the larger context, the meaning, and the implications of an experience or action” (p.1185) and it “allows assimilation and reordering or reworking of concepts, skills, knowledge, and/or values into pre-existing knowledge structures” (p.1185).

Reflective thinking is a widely recognized graduate attribute, included in frameworks such as the Tilburg Educational Profile (TEP) and the Dutch Qualifications Framework (NLQF, 2025). It is also considered one of the essential critical thinking skills. While critical thinking emphasizes reasoning, reflective thinking focuses on “deciding what to believe or do” (Norris & Ennis, 1989, p.3). It is widely accepted that reflective thinking can be taught and transferred to real-world contexts. However, despite the existence of many didactic strategies for developing reflective thinking, there is relatively little guidance on how to assess these skills within domain-specific courses.

One tool commonly used to facilitate reflection is the logbook. A logbook is a structured, ongoing record in which students document their experiences, thoughts, and reflections as they engage with learning activities (Hudson et al., 2012). Based on the reflection model chosen (e.g., the 5R framework, Gibbs’ reflective cycle) (Bain et al., 2002; Gibbs & Andrew, 2001), the logbook can vary in structure and content, ranging from descriptive observations to deeper critical analysis. When aligned with learning goals and assessment criteria, it can serve as an effective tool for both formative and **summative assessment**.

After applying the instructions of this chapter to their own assessment practices, lecturers are able to:

- Analyze the program context to determine the suitability and value of integrating reflective thinking into course design.
- Formulate course learning goals (CLGs) and design appropriate reflective tasks and logbook.
- Identify appropriate conditions and strategies for assessing reflective thinking skills, distinguishing between formative (ungraded, feedback-focused) and summative (graded, numeric or pass-or-fail) assessment purposes.

Starting points

When deciding whether or not to assess reflections in your course, consider:

In most academic programs, reflective thinking skills are embedded within the overall program **learning outcomes**. Your institution may have already defined how reflective thinking is categorized within the curriculum. For example, a business education program might classify all reflective thinking goals under the domain of learning and development, while another program may group them under a broader character-building pillar (see Box 2.11.1). These classifications can shape how reflection is emphasized and assessed in your course. To ensure alignment, it is important to consult your institutional or program-specific guidelines to understand how reflective thinking is positioned within your context.

Box 2.11.1

Example Program Learning Outcomes That Focus on Reflection³⁴

CHARACTER

Students can...

- *judge and evaluate* on formal grounds and in terms of content their *own* research and other products (e.g., DIGGIT articles) and those of fellow scholars in Culture Studies.
- *interpret* (the importance of) cultural expressions from a variety of perspectives in diverse cultural environments (offline and online, local and global).
- *employ* reasoning and arguing in the process of learning, to put forward *relevant points of view* and to understand and appreciate points of views of others, to share knowledge and contribute to discussions and the process of argumentation in a multidisciplinary group.
- *reflect* on their own qualities and skills, draw conclusions regarding their development, and take action to optimize their learning trajectory in light of functioning as a Culture Studies scholar within and beyond.
- independently *orient* themselves to relevant employment sectors (see the BA exit profiles) or to possible further studies.

When considering how to develop and assess students' reflective thinking skills in your course, it is important to ask whether there are any existing agreements or expectations at the program level. Courses often play different roles within a curriculum. Some courses are primarily content-driven, others blend content with skill development and character education, while professional practice courses may focus heavily on applied reflection and personal growth. These differences influence how much emphasis should be placed on reflective thinking and how it should be assessed.

When designing reflection activities in your course, it is important to first ask: Which reflection model is used (or recommended) in your program? What does your discipline consider as "reflection"? And what are the sub-skills that make up reflective thinking in your field?

Depending on your discipline's characteristics and conventions, some reflection models may be more suitable than others (see Table 2.11.1).

34 Made by Dr. Martin Hoondert. With permission to use.

Table 2.11.1*Reflection Models in the Three Disciplines*

Disciplines	Reflection Models
Teacher Education	<ul style="list-style-type: none"> Action, Looking back on the action, Awareness of essential aspects, Creating alternative methods of action, and Trial (The ALACT model) (Korthagen & Vasalos, 2005; Korthagen, 2014) The 5R framework: Reporting, Responding, Relating, Reasoning, and Reconstructing (Bain et al., 2002)
Business Education	Gibbs' Reflective Cycle: Description, Feelings, Evaluation, Analysis, Conclusion, and Action plan (Gibbs & Andrew, 2001)
Psychology Education	<ul style="list-style-type: none"> Thought records (Persons et al., 2001) Teaching and Assessing Reflective Learning (TARL) model (O'Connor et al., 2015)

Transparency

Sharing assessment criteria, grading standards, and clear grading instructions is essential to help students understand how their final grade is determined. For reflective assignments, it is important to state explicitly what is required and what aspects will be assessed, to ensure **transparency**. Well-designed reflective prompts, especially those that support self-monitoring, such as planning and reviewing one's own actions, are effective in helping students develop deeper, more critical reflections.

Do not assume students know how to reflect critically. Many students lack experience with structured reflection, especially when programs do not offer a consistent or scaffolded approach to developing reflective thinking skills. Teaching reflection is essential. Model reflective thinking, providing feedback, and sharing quality work of critical reflection during class is the best way to ensure transparency.

In addition, lecturers should follow their institution's Generative artificial intelligence (GenAI) and assessment policies (e.g., by stating permitted uses of AI in the syllabus). It is also important to clarify the role of GenAI tools in reflective work. Students should be informed about which tools are appropriate for supporting reflective thinking, as some GenAI tools are specifically designed to facilitate personal reflection by offering reflective prompts only without answering or generating reflections (Wolff et al., 2024).

Reliability

When course goals emphasize student development and progress over time, a pass-or-fail grading system is most appropriate, as such growth is difficult to quantify reliably. In contrast, when the goals target the attainment of a specific level of reflective thinking—particularly as part of broader critical thinking skills (Pavlovich et al., 2009; Ryan & Ryan, 2013), then quantitative grading (e.g., a 1–10 scale) may be suitable. However, even in such cases, it is strongly recommended to use a limited number of performance levels per criterion. This simplifies grading, improves **inter-rater reliability**, and reduces the burden on examiners. A conversion method can then be used to calculate a final numerical score across multiple criteria (see Chapter 2.8).

To ensure assessment **reliability**, there must be clear, observable evidence of reflection and well-designed instruments or **rubrics** to evaluate both the levels of reflection and students' attitudes toward reflective practice (Kember et al., 2000; Rogers et al., 2019; Wald et al., 2012). The assessment design should include a carefully selected set of criteria that are relevant, significant, and aligned with the intended course learning goals (Kember et al., 2000; Wald et al., 2012) (see Box 2.11.2). These criteria should be mapped to levels based on the chosen reflective framework or model (see Table 2.11.2). Avoiding excessive levels is important to maintain clarity, consistency, and efficiency in grading.

Box 2.11.2

Example Criteria for Grading Student Reflective Products (Grierson et al., 2020; Wald et al., 2012)

The REFLECT rubric is a tool designed to assess the depth and quality of reflective writing. It includes five essential criteria: (1) the range and depth of exploration, (2) the writer's presence and voice, (3) the clarity and relevance of the issue described, (4) the writer's engagement with emotions (their own and others'), and (5) the overall meaning derived from the experience. Each criterion and the overall piece is rated on four levels: non-reflective, introspective, reflective, or critically reflective, supporting consistent and nuanced evaluation of reflective work.

(Rogers et al., 2019)

The reflection rubric criteria contained five evaluation criteria: (1) presentation, (2) perspective taking, (3) connection, (4) understanding-cognition, and (5) understanding-emotion.

Validity

Define different levels of reflective thinking skills by integrating Bloom's Taxonomy (see Table 2.11.2) and established reflective models. These models often span multiple dimensions (i.e., personal, interpersonal, contextual, system) and each representing progressively deeper levels of cognitive engagement and reflection (Ryan, 2012), spanning across a timeline, from past, current to future³⁵. Although some argue that there is a difference between reflection and reflexivity (Thompson & Pascal, 2012), it is beyond the scope of this chapter and thus we only refer to the critical level and how students' own background, beliefs, assumptions, and social context influence their thoughts, decisions, and actions.

Once the intended CLGs related to reflective thinking are clearly defined, it is important to select a reflective task (see Box 2.11.3) that aligns with those goals and engages students in the appropriate level of cognitive processing. This includes choosing a suitable reflective genre (i.e., a recognized academic format that structures how reflection is demonstrated) (see Box 2.11.4). A logbook, for example, enables students to track their experiences and development over time, draw connections between events, and gradually engage in deeper and more complex levels of reflective thinking.

To effectively support student learning, the CLGs, reflective task, required logbook entries, and the delivery method in the chosen reflective genre must be constructively aligned. This alignment ensures that every element of the course design works together to guide students toward the program intended learning outcomes. Equally important is the role of lecturer supervision and feedback throughout the process. Lecturers should provide ongoing guidance and monitoring, helping students develop reflective thinking skills.

Although grading reflection is often debated, a balanced approach is most appropriate, as suggested by Creme (2005). Use logbooks as a formative tool to support reflective tasks and the development of a final product in a reflective genre, without compromising authenticity. Then, assess only the final product summatively, allowing students to synthesize their learning while preserving the integrity of the reflective process. Also, using reflection as a standalone formative assignment may not motivate students to do it (Hudson et al., 2012). Strategically aligning formative with summative assessment is essential to motivate students to work towards intended CLGs on reflective thinking.

35 Inspired by personal communication with Dr. Krijn Pansters.

Table 2.11.2*Example Course Learning Goals on Reflective Thinking, Based on Bloom's Taxonomy*

Dimensions	Level	Example Learning goals
System (integration of personal, interpersonal, contextual)	Evaluate & Create/ Synthesize	<ul style="list-style-type: none"> Evaluate the adequacy or limitations of academic material and/or prior understanding, proposing specific revisions based on critical reflection and experiential evidence. Formulate well-reasoned questions and identify additional sources or types of evidence needed to make more informed and nuanced judgments about the accuracy or relevance of the material. Design thoughtful, evidence-informed recommendations for future action or decision-making at the individual or organizational level while clearly articulating the potential benefits, risks, and challenges involved.
Interpersonal & Contextual	Analyze	<ul style="list-style-type: none"> Critically compare and contrast academic concepts with personal or observed experiences, identifying specific similarities and differences. Analyze potential reasons for discrepancies between theory and experience, considering factors such as personal bias, assumptions, contextual limitations, or gaps in the original material. Identify and articulate new insights, complexities, or nuances in the academic material that emerge through reflective analysis of real-world application or experience.
Interpersonal or Contextual	Apply	<ul style="list-style-type: none"> Apply relevant academic concepts to analyze and interpret personal learning experiences, demonstrating how theoretical knowledge is used or observed in real-world or course-related contexts.
Personal (stance and epistemological understanding, 'disciplinary assumptions about the nature of knowledge') (Bruce, 2008, p.131)	Identify & Describe (Knowledge & Comprehension)	<ul style="list-style-type: none"> Identify an academic concept that relates to learning experience. Describe the academic concept that relates to learning experience.

The two short design examples illustrate this research-informed suggestion:

Example 1

As a summative assessment graded as pass-or-fail, students will write a “Dear Past Me” letter, a reflective genre that encourages them to revisit and engage with their earlier logbook entries and reflections. The logbook entries document the practice of targeted skills (e.g., study skills) and include weekly self-evaluations, noting what went well and what could be improved (e.g., attending all lectures on time versus delaying study course materials until the week before the exam). Lecturers can identify common issues across groups and allocate class time to discuss them, offering targeted group feedback to support collective learning and improvement. The “Dear Past Me” letter serves as a reflection of reflections: it requires students to synthesize their weekly reflections, identify patterns in their performance across different topics or contexts, and analyze how their understanding and skills have evolved over time.

Example 2

Reflection can also be integrated into groupwork, particularly when CLGs emphasize collaboration, communication, reflection, and responsibility. One effective approach is to require students to submit a Critical Incident Report alongside their group project (Nylén & Isomöttönen, 2017). This report involves a deeper analysis of key moments, such as how the group managed conflict, made decisions, acted, and dealt with outcomes. Each group member records their perspective on these incidents in their individual logbooks, capturing the context and their personal response. These entries foster both collective insight and individual reflection on group dynamics and learning processes. Lecturers can integrate groupwork into tutorial activities to allow for observation of group processes, particularly moments of conflict. In addition, lecturers can use students' logbook entries to verify the authenticity of the group's interactions and dynamics. The Critical Incident Report can then serve as one part for summative assessment, providing evidence of both reflective thinking and group collaboration.

Box 2.11.3

Examples of Reflective Tasks (modified from the critical thinking tasks of Ennis, 1993; Facione, 1989)

- Appraise own skills and abilities.
- Set well-defined goals, monitor progress, and motivate self.
- Examine personal beliefs and measure them against the beliefs of others.
- Interpret self as part of the larger community.
- Justify own actions and impact on others.
- Show respect to individual rights and the dignity of others.
- Analyze new ideas through interaction with others.
- Seek and acquire information and knowledge to achieve academic, career, and personal success as a lifelong learner.
- Apply critical analysis and reflection to ethical research, synthesize and evaluate complex information, problems, concepts, theories and interpretations to demonstrate cognitive and technical skills in a body of knowledge or practice.
- Evaluate their own and others' performance in teaching, training, or tutoring.

Box 2.11.4

Examples of Reflective Genres³⁶

- Critical incident reports (Bruster & Peterson, 2013; Thomas, 2016)
- Fictional narratives: Dialogue with a theorist (Brokerhof et al., 2018)
- Dear past me: Reviewing past written reflections (Leyland et al., 2024)
- Letter to future self (Haase, 2019)
- Metaphor reflection (Maddamsetti, 2025; Shakoury & Boers, 2025)
- Portfolio reflection (Lyons, 2009)
- Reflective essays (Bennion et al., 2020)
- Reflective case study (MacLellan, 2008)

³⁶ Note that these reflective genres should be designed as either Learning-to-write (LTW) or Writing-to-learn (WTL) activities, depending on their purpose and intended course learning outcomes (see Chapter 2.7).

Checklist for Designing Reflective Assignments

Starting points

- Align the assignment with your program's learning outcomes on reflection.
- Clarify the course's role (e.g., theory-focused, practice-based) and adapt reflection emphasis accordingly.
- Use a reflection model that fits your discipline (e.g., Gibbs for Business, ALACT for Education).
- Define what "reflection" means in your context to guide student expectations.

Transparency

- Clearly communicate expectations, assessment criteria, and how grades are calculated.
- Share examples of high-quality partial reflection reports with students.
- State permitted uses of GenAI tools in the syllabus.
- Encourage only those tools that support reflective thinking (e.g., prompt-only tools³⁷, limit AI uses to Level 1-3 based on AI-index when AI is allowed).

Reliability

- Modify criteria from validated rubrics (e.g., REFLECT) with no more than four levels per criterion.
- Grade only the final product summatively to preserve an authentic learning experience in the process.

Validity

- Write clear, measurable goals using Bloom's taxonomy (e.g., analyze, evaluate, create).
- Ensure goals cover both cognitive and emotional aspects of reflective thinking.
- Choose a genre that matches learning goals and student level (e.g., "Dear Past Me" letter, case reflection).
- Design tasks that require students to synthesize experiences, not just describe them.
- Include weekly prompts that support planning, self-monitoring, and reflection.
- Emphasize the formative purpose of logbooks that support development; however, these are not directly graded.
- Align learning goals, instructional activities, task type, and assessment criteria.
- Ensure the logbook contributes meaningfully to the final reflective product.
- Schedule regular check-ins or feedback opportunities on logbook progress.
- Use classroom time to discuss common issues or patterns in reflection.

Acknowledgement

We would like to thank Dr. Astrid Kramer for her review and feedback, which helped improve the quality of this chapter.

37 Tailor Tilly bot to prompt student reflections. See <https://chatbot.tilburg.ai/>

2.12

Internship Assessment³⁸

Ya-Ping (Amy) Hsiao

Introduction and Learning Goals

In recent years, the themes of employability and authenticity have received more and more attention in research universities in the United States (Association of American Colleges and Universities, 2009b), the United Kingdom (Knight & Yorke, 2003), and the Netherlands³⁹. As this is being recognized as a high-impact educational practice (Kuh, 2008), more study programs include internships in the curriculum to develop work-ready graduates (Jackson et al., 2017). Note that the terms internship, work-based or work-integrated learning, and cooperative education are often used interchangeably in practice, but internship is the most used term in the Dutch research universities, such as Tilburg University.

There are various types of internships. This chapter focuses on an *academic internship* that is part of the program curriculum (not an extra-curricular activity) and graded (either with a number grade or pass-or-fail). An academic internship provides students with professional learning experience in which they integrate knowledge and theory learned in the classroom with practical applications and skills development. An academic internship is a well-planned educational activity through which students make connections between theories and practices in professional fields. When working on projects, students not only need to contextualize their academic knowledge into placement but also mediate different criteria and quality standards between academic and placement supervisors. Therefore, an academic internship provides students with the experience to deal with multiplicity that is authentic and close to professionals' real-life practices.

This chapter focuses on the most essential issues that are specific to internship **assessments**, such as the role of the external supervisor and student ownership in assessment.

After applying the instructions of this chapter to their own assessment practices, lecturers are able to:

- Define the internship and explain why an internship fits the program curriculum;
- Apply the guidelines for ensuring **validity**, **reliability**, and **transparency** in internship supervision and assessment;
- Use reasoned arguments based on the program's published assessment instructions to justify their assessment results when discussing them with colleagues.

³⁸ This chapter is adapted from the assessment handbooks developed by the Tilburg School of Humanities and Digital Sciences (TSHD).

³⁹ Doing an internship | Study in Holland

Starting Points

Define the internship and clarify the connection between internship experiences and the student's academic program curriculum.

In the Netherlands, university internship supervisors have dual role tasks of supervision and assessment. Supervision is always a shared responsibility between the university and placement supervisors (thus it is not only the responsibility of external supervisors at the placement). University supervisors can design and organize the following four prototypes of learning activities (Cheng et al., 2011; Hsiao et al., 2020) (see Table 2.12.1):

Table 2.12.1
Prototypes of Internship Supervision Activities

Dimension	Individual Reflection	Peer Learning Group	University-Facilitated Group	Collaborative Supervision
Participants	Student	Students and peers	Students, peers, and university supervisor	Student, university supervisor, and external supervisor
Main Activities	Self-reflection	Problem discussion, peer review, self-reflection, and peer reflection	Problem discussion, peer review, case studies, self-reflection, and peer reflection	Reflection, feedback, and assessment
Delivery Mode	Online	Online	On campus	Online or on campus

These activities can be designed to serve both formative and summative functions, and it is essential to make clear which functions each activity form serves.

Transparency

State the grading policy clearly in the internship manual. It should be clear to students whether they get a pass-or-fail or a conventional number grade. When using a pass-or-fail for the entire internship grade, it is sometimes required to consult your program academic director and/or Examination Board for approval.

In the internship assignment descriptions, clearly state the required deliverables for both formative and **summative assessments**. The common deliverables for academic internships are (Grady, 2017, p.356):

- daily or weekly diary, journal, or calendar of events;
- a description or discussion of formal and informal tasks completed;
- reports on books or readings pertinent to employer or type of work;
- a portfolio of work completed during the internship;
- papers based on interviews with significant personnel in the company or organization;
- a final report or reflection paper;
- a debriefing session with the university supervisor;
- an external supervisor's evaluation;
- a student evaluation of the worksite and internship experience.

When different deliverables count towards the final grade, it is important to specify how they are weighted.

Based on observations made over the past few years across various internship assessment (re)designs, the most common problem in internship assessment is the lack of clear policy and assessment information for stakeholders and key roles involved in the assessment organization and process. This is why we provide the following texts on policy and assessment as *examples* for internships undertaken as *graduation projects*. Your institution may already

provide similar information. If not, you may consult and adapt the example texts below to fit your program context. These materials are provided as examples only and should be adapted to align with institutional requirements and practices. Once the information has been officially approved by the relevant stakeholders such as the educational management team, program curriculum team, and/or Examination Board in accordance with your institution's assessment governance it should be shared with students. The information covers the following ten aspects:

1. The requirements for the first and second examiner if internships are a *graduation* project;
 - The university examiners are appointed by the institutional or program's Examination Board.
 - External supervisors are usually not appointed by the Examination Board, and therefore they are not examiners. They can only provide feedback (no grading) for university examiners to grade.
2. The guarantee of independence of both examiners;
 - 2nd examiner is not involved in any supervision activities.
 - Independent grading: The university supervisor and 2nd examiner fill out the assessment form separately without consulting each other. and they send out the form independently.
3. The determination of the mandatory use of the assessment form and the way in which it should be used;
 - The assessment form is co-developed by the coordinator, academic directors, internship coordinators, and the assessment specialist. It is approved by the stakeholders. Its use is mandatory.
 - for communicating shared quality expectations and standards (transparency);
 - for assessing student internship progress and outcomes (validity and reliability).
 - After approval of the internship package by the Examination Board (i.e., Rules and Procedures, assessment forms), the academic director shares the Rules & Procedures Internship Assessment and assessment form with the students and the university and external supervisors. In addition, they elaborate on the relationships between the internship learning goals (= program **learning outcomes**) and assessment criteria as well as how criteria should be used during the supervision process.
 - Formative continuous assessment: The external supervisors use the criteria to formatively assess student progress on the assigned tasks and give feedback. The feedback moments (when, frequency, and on which deliverables) should be clear for students.
 - Formative interim evaluation (if feasible): About halfway through the internship, a (digital) conversation takes place between external supervisor, internship coordinator/university supervisor (can be the same person), and the student. In this meeting, they discuss student developments so far and set goals for the period thereafter. The student writes a report of that meeting and sends it to the site and university supervisors for approval. The student adds this report to the final deliverable (e.g., portfolio).
 - Final summative assessment:
 - At the end of the internship, the external supervisor draws up an advisory report by filling out qualitative evaluations based on university criteria (without grading). In this report, students' developments are evaluated in terms of the internship learning goals and assessment criteria.
 - University examiners use the assessment form with the criteria to assess the internship outcomes. They also consult the advisory report to validate the evidence and justifications of the student's portfolio (to show credibility).
4. Guidelines on how to determine the final grade for the final assignments;
 - Grading instructions by using a **rubric** (see Chapter 2.7 for an example):
 - The assessment of student deliverable (e.g., portfolio) is based on a mixed approach of both analytical and holistic grading.
 - For each outcome, the examiner will first hold a holistic judgement based on all criteria to review the evidence, justifications, and advisory report. Then (s)he will carry out analytical judgments to examine deeply whether the documents meet the criteria of each outcome.

- Based on a set of logical rules, the analytical grades on all outcomes of each outcome are converted into an outcome grade. Then all outcome grades are converted into the final grade of the internship.
- Each set of logical rules contains the performance standards translated from the collective expectations of examiners.
- Comments

For each competence, there is a comment space for each examiner i) for feedback: justify why a certain grade is given and ii) for feedforward: provide suggestions on improvement.
- 5. The way in which the first and second examiners arrive at a final grade;

After independent grading, two university examiners (i.e., the supervisor and second reader) discuss their results with each other and the university supervisor fills out the agreed grades per outcome, the final grade, and integrated comments in a collective assessment form. With the consent of the second examiner, the university supervisor shares the collective assessment form with the student.
- 6. The way in which it is made clear to the student how the final grade was arrived at;

The student receives the collective assessment form within 15 working days after portfolio submission.
- 7. Rules on how to deal with differences in assessment between first and second examiners;

When two university examiners differ on an outcome grade, they need to re-examine the student's portfolio documents during the discussion and reconcile the differences.
- 8. Rules indicating when a second or third examiner is involved and the weight of his/her opinion;

When both examiners cannot agree on the grades.
- 9. Rules indicating how a resit is arranged by answering the following questions (and these rules can be set by the Examination Board):
 - What counts as first opportunity and resit?
 - What do “pass-or-fail” and “repairable” mean? Note that the practical part of the internship can often only be repaired by doing another internship.
 - What grades can students receive on the re-submission? Only 6 or higher?
 - Can a student with a grade that is already higher than 6 resit? What additional requirements do you want to impose on this situation?
 - When students fail on the first opportunity and it is not repairable within a short period (e.g., one or two weeks), what can students do for the resit? To revise the internship deliverables and/or do extra/comparable (if possible) assignments? Or to re-do the internship at another organization?
 - When the resit is failed again, what can students do? Note that some parts (like written reports can be improved and a resit is possible), but the internship itself there is often no resit and students will have to do another internship next year).

Be aware that resits for internships are not always feasible due to their extended duration (e.g., 9–12 months), making it impossible to offer a second opportunity within the same academic year. For example, in Psychology, internships are assessed using competence-based rubric. If a student fails to demonstrate the required competences, the result is a fail. This is not a matter of revising a product, as the assessment is based on performance in practice rather than on a written output.

- 10. Archiving and storage periods of final assignments and assessment forms.
 - The School is responsible for the proper archiving of matters relating to the internship laid down in the university's Thematic Privacy & Personal Data Protection Policy – Processing Personal Data. The School may delegate this task, for example to the educational support team or the secretariat of a Department.

- The following must be stored together: a. the portfolio; b. the assessment form of the university supervisor; c. the assessment form of the second examiner; d. the final grade; e. in case a third examiner is involved, the assessment form of the third examiner is also kept.
- The retention period of the documents mentioned in the University's Thematic Privacy & Personal Data Protection Policy – Processing Personal Data is seven years.
- The retention of portfolios and the associated documents involve personal data. This data processing is necessary for compliance with a legal obligation and is recorded in the university's Thematic Privacy & Personal Data Protection Policy – Processing Personal Data.

Reliability

External supervisors are able to directly observe student performance across a wide variety of professional situations over an extended period of time. It is thus important to use their feedback to validate student evidence, justifications, and reflections. However, external supervisors cannot directly grade internships because, in most disciplines, they are not qualified examiners, appointed by the Examination Board, and they often have not received training in educational assessment. In addition, research in clinical psychology has shown the leniency bias in the judgements made by external supervisors (Gonsalvez & Freestone, 2007). In addition to the leniency bias, halo effects⁴⁰ have also led to higher grades in the public relations internship programs (Wolf, 2015). Therefore, external supervisors focus mainly on daily supervision activities, **formative assessment** based on their observations and student interns' ongoing progress, and formative feedback that prompt student interns' improvement.

To fulfill supervision and assessment tasks, both academic and external supervisors should use the same criteria and performance standards. They should have a shared understanding of these quality criteria so that they can feedback students to what extent student performance has met the criteria and feedforward suggestions for improvement in the follow-up and final products. Using a template based on assessment criteria but allowing some level of personalization can help both academic and external supervisors to give feedback efficiently and to assess student performance consistently.

As internships apply a complex assessment design, it is important to organize sessions for examiners to regrade past student internships.

Validity

- Formulate the internship learning goals (Walsh, 2007). See Box 2.12.1.
 - Relate each goal to the corresponding program learning outcome (i.e., PLOXX).
 - For the internship assessed at the exit level, assess the highest level of the targeted program learning outcomes. The cognitive levels should focus on critical analysis, evaluative judgment, reflection, integration, and transfer from academic theories into practical professional contexts (Eraut, 2009). Internship learning goals (as the graduation project) should cover *all* of the components or categories of the program learning outcomes. For smaller internships with fewer European Credit Transfer and Accumulation System (ECTS), this is not applicable.

40 Hanging a halo over the actual performance of a student, hence giving her or him a higher grade than deserved.

Box 2.12.1*Example Internship Learning Goals*

- Apply knowledge gained in the classroom (or major) to real-world challenges in an internship environment (PLOXX).
 - Develop and enact a compelling professional vision that values diversity and inclusion in the workplace (PLOXX).
 - Engage in responsible conduct while working as an intern and allow decisions to be informed by a value-centered life (PLOXX).
 - Develop professional connections and identify a strategy for maintaining those connections (PLOXX).
-
- Provide students with clear instructions on formulating their personal career development goals⁴¹. If student attainment of their personal development goals is graded through a reflection report, then these goals must be derived from the program's internship learning goals.
 - Distinguish between the academic program's internship learning goals and the student personal development goals (if applicable) and include both.
 - Choose appropriate assessment types for both the internship and the students' personal development goals. Assessment types should be aligned with the supervision activities. Authentic assessment tasks are appropriate for assessing student performance on both the process and products (Ajjawi et al., 2020). The examples of authentic assessment tasks in psychology are case-based assignments, graphic organizers, and analogical reasoning (Mayo, 2010).
 - Determine whether personal development goals are assessed by formative or/and summative assessments.
 - To design internship assignments, it is necessary to communicate the expectations and requirements between external organizations and the program. The internship assignments should be relevant and important for the internship learning goals that aim to advance students' capabilities (M'Baya et al., 2016). The internship coordinators should establish a matching system to examine the appropriateness of the internship posting descriptions for the internship learning goals (M'Baya et al., 2016).
 - Case-based assignments (such as anamnesis patient x, formulate hypothesis, ... intervention, Basic Certificate in Psychodiagnostics /BAPD): Apply diverse life-adjustment principles over a broad range of contexts in the lives of self and others.
 - Graphic organizers: State and relate connections between diverse facts and theoretical concepts in psychology.
 - Analogical reasoning: Relate and apply connections between facts and developmental theories over a wide range of analogical contexts.
 - In addition, authentic assessment tasks should include student engagement in a workplace context/ audience, high quality cognitive engagement, and reflexive evaluation performance (Bosco & Ferns, 2014).
 - Determine when, how, and what kind of feedback (formative or summative) students can expect from peers, the university, and external supervisors (based on cross-organizational agreements). Also, students should be able to make an evaluative judgement of their own overall performance based on the synthesis of a series of self-assessments on the internship assignments they have carried out during the internship process (Jackson et al., 2017).

41 Consult 6.o Formulating goals (p.9-10) of personal development plan (PDP) guidelines 1920.pdf (tue.nl).

Checklist of Designing Internship Assessment

Transparency

- Grading policy clearly stated (pass-or-fail or numerical grade)
- All deliverables clearly defined (formative + summative)
- Weighting of each deliverable specified
- Students informed about assessment procedures
- Students informed about criteria and expectations
- Matching process for internship placements implemented

Formative Assessment Process

- Continuous feedback moments scheduled
- External supervisor uses criteria for feedback
- Midterm assessment meeting conducted
- Midterm report written and approved (and included in student portfolio)

Roles & Responsibilities

- Examiners appointed by Examination Board
- External supervisor role clarified (feedback only, no grading)
- University supervisor/examiner role clarified (supervision and grading responsibility)

Resit Policy

- First attempt vs resit clearly defined
- “Pass-or-fail” and “repairable” clarified
- Rules for resubmission grades defined
- Conditions for retaking internship specified
- Policy for failing resit included
- Feasibility of resits considered (duration constraints)

Reliability

Assessment Form Use

- Mandatory assessment form approved and shared
- Criteria aligned with internship learning goals
- The same form used consistently by all supervisors

Summative Assessment Process

- External supervisor provides advisory report to validate student evidence
- University examiners assess portfolio using rubric
- Advisory report used to validate student evidence
- Final grade calculation method explained
- Summative Feedback (comments + feedforward in case failing) provided
- External supervisor input used for validation
- Awareness of bias (leniency, halo effect)
- Examiner calibration sessions organized
- Shared understanding of criteria ensured

Independence & Objectivity

(for graduation internships)

- Second examiner not involved in supervision
- Independent grading completed by both examiners
- Assessment forms submitted separately
- Third examiner appointed if needed
- Final decision process clearly documented

Validity

- Internship learning goals defined and aligned with program outcomes
- Exit-level cognitive skills included (analysis, reflection, transfer)
- Personal development goals, if included in the assessment, defined separately
- Assessment methods aligned with goals
- Authentic tasks included (real-world relevance)
- Internship assignments aligned with learning goals

Archiving & Data Management

- Portfolio, including all assessment forms, stored according to legislation GDPR, policy
- Final grade recorded
- Third examiner form (if applicable) stored
- Retention period (e.g., 7 years) defined
- Data protection compliance ensured

2.13

Thesis Assessment⁴²

Ya-Ping (Amy) Hsiao

Introduction and Learning Goals

The Oxford English Dictionary defines ‘thesis’ as “a long essay or dissertation involving personal research, written by a candidate for a university degree,”⁴³ and it is an “in-depth study of a topic that contributes novel information in the field of research” (Parija & Kate, 2018, p.4). The focus of this chapter is thesis **assessment** used as a *final project* near the end of the program curriculum. As stated in the NVAO assessment framework, a final project refers to the assessment with which students complete their study program. Note that it is sometimes not possible to assess all program **learning outcomes** in the graduation thesis, but it is important to assess all categories of program learning outcomes in the thesis (e.g., Tilburg Educational Profile/TEP three pillars, Dublin Descriptors, The Dutch Qualifications Framework (NLQF), Teacher Competences). Therefore, a final project should assess the achievement of all or essential program learning outcomes (especially when thesis learning goals are formulated).

A misunderstanding is that a final project must be a thesis while in fact a final project can be “the final thesis, a portfolio, a professional product, an interim exam or series of interim exams, a paper, an artistic achievement, or a combination thereof” (NVAO, 2018, p.21). Therefore, each program should consider carefully whether a thesis is the most appropriate method for assessing the final attainment (highest levels) of the program learning outcomes. For example, for a competence-based program like the Teacher Training program (*Universitaire Lerarenopleidingen*), an internship is a more appropriate assessment type than a thesis for a final project.

At the same time, the growing emphasis on student employability has prompted many Dutch research universities to reflect on the long-standing tradition of the thesis as the standard graduation project. Although stakeholders increasingly recognize the need for alternative assessment methods, the thesis remains the most commonly used format. Nevertheless, several initiatives have emerged to improve current practices, which will be discussed in this chapter.

After applying the instructions of this chapter to their own assessment practices, lecturers are able to:

- Explain the research nature of bachelor’s and master’s theses and relate these characteristics to the program learning goals and thesis learning goals.
- Apply the guidelines of ensuring **validity**, **reliability**, and **transparency** in thesis supervision and examination.
- Use reasoned arguments based on program published assessment instructions to justify their assessment results when discussing them with colleagues.

⁴² This chapter is adapted from the assessment handbooks developed by the Tilburg School of Humanities and Digital Sciences (TSHD).

⁴³ https://www.oxfordlearnersdictionaries.com/definition/american_english/thesis?q=thesis

What is a Thesis?

A thesis is a formal academic work that presents a clear proposition (e.g., an idea, hypothesis, or recommendation) and develops it into a structured argument. Rather than merely stating a viewpoint, it provides a reasoned rationale supported by appropriate logic and sufficient scientific evidence. Throughout the work, the argument is consistently developed and maintained to convincingly address a specific research question or issue. In essence, a thesis makes a focused claim, proposes an answer to a problem, and systematically seeks to justify and validate that answer through critical analysis and evidence-based reasoning (Field & de Neufville, 2021).

Bachelor's Thesis

The bachelor's thesis focuses on developing core research skills through an independent, in-depth project (see Box 2.14.1). Students learn to formulate research questions, choose and apply appropriate methods, and collect or analyze data (can be either primary or secondary data). The process emphasizes critical thinking, problem-solving, and the ability to evaluate findings within a disciplinary context. In addition, students develop project management skills by planning and conducting a substantial piece of work overtime, as well as communication skills through written and oral presentation. Overall, the thesis serves as a transitional experience, preparing students for graduate-level research or professional practice by fostering independence and systematic inquiry (Boyer Commission on Educating Undergraduates, 1998; Todd et al., 2004).

Box 2.14.1

Example of Short Introduction to a Bachelor's Thesis

The bachelor's thesis is the culminating component of the bachelor's program, consisting of XX European Credit Transfer and Accumulation System (ECTS) (XX-hour) research project conducted within a department. It provides students with the opportunity to apply the knowledge and skills from the program in an independent project. The thesis serves to assess students' initiative and their ability to plan, execute, report, and present their work. Its level of complexity is aligned with the program's intended learning outcomes and the modules completed prior to the thesis. Students complete the thesis or an individual assignment independently, under the supervision of a designated supervisor (University of Twente, 2026).

Master's Thesis

A master's thesis is a substantial piece of academic research that demonstrates a student's ability to conduct independent, original, and rigorous investigation within a specific field (see Box 2.14.2). It is characterized by originality, depth of theoretical understanding, and breadth of scholarly engagement, enabling the student to contribute meaningfully to existing knowledge.

As a form of postgraduate research, a master's thesis emphasizes a high degree of independence, requiring students to design, execute, and present their research with limited supervision. It is typically structured around key components such as an introduction, literature review, methodology, results and analysis, and conclusion, though variations may occur depending on the discipline. In many academic contexts, it is a substantial work completed within a relatively limited timeframe, and it reflects the student's ability to engage critically with existing literature, apply appropriate research methods, and draw well-supported conclusions with broader academic implications (Felix & Smith, 2019).

Box 2.14.2*Example of Short Introduction to a Master's Thesis*

In the second semester of your MSc program, you write your MSc Thesis. The main goal is that you apply the knowledge and insights from your marketing (analytics) courses to a real-life marketing problem. The topic of your MSc Thesis can be inspired by a practical problem of a company (company-defined topic) or by a research problem provided by one of the supervisors (supervisor-defined topic). The result is an academic paper in which you present a conceptual framework, elaborate on related literature, describe the results of an empirical study, and discuss the academic and managerial implications of your findings (Bosmans, 2026).

Transparency

One way each program safeguards the proper assessment of final projects is to ensure transparency. Transparency concerns how well students know what is expected of them and how their thesis will be judged. At the program level, the thesis assessment information should be clearly stated in the thesis manual. Supervisors not only use the assessment form for thesis assessment but also discuss the assessment criteria and performance standards during the supervising activities. The supervisors should clearly explain the meaning of these criteria and score levels (excellent, good, sufficient/pass, and insufficient/pass) to their students, using anchor theses or past student work as illustration (Hsiao et al., 2023).

The following is a list of attention points for transparency, derived from an analysis of multiple thesis manuals across different faculties at Tilburg University.

- For most students, this may be their first experience completing a complex, multi-step assessment. It is therefore essential to guide them carefully through each stage of the process. A clear visual representation can help students understand the overall structure and gain a better overview of the entire workflow (see Figure 2.13.1 for an example).
- It must be explicitly stated whether the use of Generative artificial intelligence (GenAI) is permitted in the thesis process, and if so, the extent to which it is allowed. Where applicable, an AI usage classification (AI index) should be introduced to clarify acceptable levels of AI involvement (e.g., from no use to supervised assistance). The thesis manual should clearly define responsible and ethical use of GenAI, particularly in relation to supervision, academic integrity, and independent student work. If AI use is permitted, it should align with the program intended learning outcomes of the program and the established learning trajectory. Students are expected to progressively develop competencies, including those related to AI literacy. Therefore, it is unrealistic to assume that students will not engage with GenAI tools during their studies. Instead, by the time they write their thesis, they should be able to use GenAI in a controlled and appropriate manner that reflects their level of academic development.
- If any form of GenAI justification documentation is required (e.g., an AI logbook, usage record, or technology statement), the manual should clearly specify what information must be included, such as the type of AI tools used, their purpose, the extent of their contribution, and how outputs were verified, critically evaluated, and integrated into the work. It should also explain how this documentation will be used in **summative assessment** to ensure transparency and consistency. Importantly, such documentation should not be purely bureaucratic but should support the development of students' understanding of responsible and ethical AI use. Transparent disclosure of GenAI use is essential for fostering a trusting academic environment, as it helps supervisors understand how these tools are integrated into student workflows and addresses academic integrity and **fairness** concerns, ensuring students are neither unfairly advantaged nor disadvantaged (Sabbaghan & Eaton, 2025). When such documentation is required (see Appendix 1 Technology statement), the consequences of non-compliance should be clearly communicated.

- In cases where a thesis proposal is required, the assessment format must be clearly defined, including whether it is graded on a pass-or-fail basis, whether a resit opportunity exists, and the consequences of failure (e.g., revision requirements or restarting the thesis process). If the proposal is formative only, this must be explicitly stated, along with its purpose as a developmental feedback instrument rather than a summative assessment.
- When intermediate deliverables are required during the thesis process, their purpose must be clearly explained, including how each submission contributes to the overall research process, supports supervision and progress monitoring, and informs feedback (Hsiao et al., 2025). It should also be specified whether feedback is provided by supervisors, peers, or both, and how it is intended to guide improvement toward the final thesis submission. Importantly, AI justification documents should be discussed during supervision meetings alongside these submissions, enabling AI literacy to be gradually developed through formative dialogue and allowing supervisors to identify potential misuse early and provide timely guidance. Clear consequences for non-compliance with submission requirements should also be communicated to students.

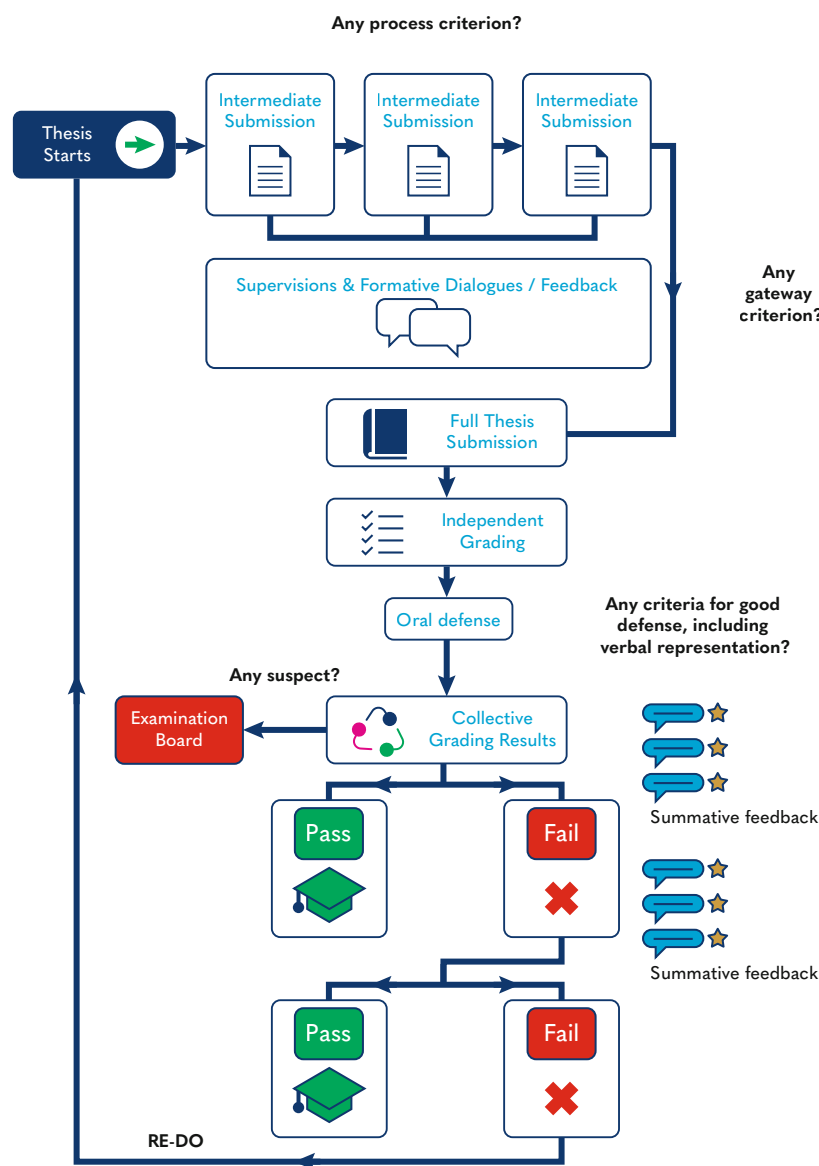


Figure 2.13.1
Thesis Assessment Procedure Prototype

Reliability

Reliability concerns the degree to which we can use grades to accurately determine students' true performance on the thesis. There should be a certain degree of consistency in the grades for one single examiner, so-called "**intra-rater reliability**," or across examiners, known as "**inter-rater reliability**." This also is important across exams. Ideally, exams should be parallel tests and therefore have a more or less equal distribution of **item difficulties**.

Intra-rater reliability refers to situations in which the scoring process of a given examiner changes over time. Inconsistencies in the scoring process result from influences that are internal to the examiner rather than true differences in student performance. To ensure a satisfactory degree of intra-reliability, the definition of each criterion in the grading form should be clearly stated. In addition, throughout the scoring process, the examiner should revisit the established criteria to ensure that consistency is maintained.

Inter-rater reliability refers to the degree of consistency between examiners. To ensure a satisfactory degree of inter-rater reliability, examiners should make consistent judgments about the criteria and score level. A problem is that even with pre-defined criteria in the grading form, examiners might interpret the criteria and each score level differently. There are two ways to remedy this⁴⁴. First, find anchor theses. Anchor theses provide a set of evaluated criteria and grades from the previous years that illustrate the nuances of the grading form. Examiners can then refer to the anchor theses throughout the scoring process to illuminate the differences between the score levels. Second, organize an information or calibration session. Prior to grading, examiners need to have a shared understanding about the expectations for each criterion and each score level so that a score level on a particular criterion is awarded to student responses with the same characteristics, regardless of who grades them.

Validity

The starting point of validity is to clearly define what the assessment intends to measure. In general, the three aspects, thesis learning goals, thesis activities, and thesis examination, should be aligned constructively. Validity focuses on alignment between the thesis learning goals and the thesis examination.

Every study program should establish a program assessment plan that shows how program learning outcomes are assessed by both the courses and the thesis. According to the plan, the graduation project or thesis measures all program learning outcomes. However, program learning outcomes are often global in character. Thus, they are often perceived as too general, sometimes ambiguous, and of little assistance to students' understanding of what is expected in the thesis examination. To make clear how a bachelor's or master's thesis contributes to the program learning outcomes, it is necessary to formulate thesis learning goals.

The thesis type should be appropriate for assessing thesis learning goals. Two common thesis types for bachelor's programs are

- original research using primary sources: Relate theory to empirical observations, e.g., by using existing data⁴⁵, by collecting data using questionnaires or experiments;
- review studies synthesize or analyze research already conducted in primary sources: This type of thesis builds on a theoretical model or a generic problem. Often a theoretical thesis is based on existing literature studies in which a theoretical problem is analyzed.

The assessment criteria should also be aligned with the thesis learning goals. Each study program should examine whether the criteria are appropriate for assessing its thesis learning goals. Literature on thesis assessment criteria (Brookhart, 2018; Hand & Clewes, 2000) has suggested using a limited number of criteria because using too many

44 The institution should establish quality assurance system to monitor inter-rater reliability, such as periodic checks of grade differences among examiners within and across programs, re-grading a sample of past theses by the quality assurance committee.

45 It is possible that a group of students use the shared data, but each student should write own thesis (not a group thesis).

criteria to assess theses could result in grading fatigue, which itself can cause much variability found in assessment quality. The criteria should be definable, observable, distinct from one another, complete, and able to support descriptions along a continuum of quality (Brookhart, 2018; Brookhart, 2003; Walvoord & Anderson, 2011). Pay special attention to the criteria on originality (Guetzkow et al., 2004) and independence (Magnusson & Zackariasson, 2019) because these two criteria are often defined in a vague way that leads to subjective interpretations among examiners.

- Originality criteria include: identifying or formulating new questions, topics, or areas of inquiry; applying new ideas, methods, approaches, or analyses to existing questions, problems, texts, or theories; developing or applying new theories, theoretical descriptions, or conceptual frameworks; inventing or using new methods, techniques, computational approaches, or technologies; generating or using new data, datasets, archives, materials, or sources; applying existing ideas, methods, or frameworks to new data or sources; developing new analytical approaches, models, frameworks, or statistical procedures; producing new insights, interpretations, connections, perspectives, or observations; drawing new conclusions, findings, answers, or proofs; and combining or synthesizing knowledge, methods, or concepts across disciplines or fields (Lovitts, 2023).
- Independence criteria include: autonomy in topic selection, decision-making in the learning process, management of the learning structure, control of research focus, independent use of learning resources, reflection on independent learning, and overall demonstration of independent learning (Todd et al., 2004).

Whether the criteria are considered as holistic or analytic, a balance should be established between professional autonomy (holistic) and the need for public accountability (analytical). For novice examiners, assessment criteria are regarded as an important tool to assess with confidence and make references in the assessment process (Hand & Clewes, 2000). It became clear that in various discussions with many new lecturers at our university they feel uncertain when assigning grades and anxious when participating in the thesis moderation/reconciliation process with a senior examiner.

For quality assurance purposes, we recommend program academic director or thesis coordinator to make an overview to show cross references among the program learning outcomes, thesis learning goals, and assessment criteria.

Thesis Learning Goals

Akin to course design, thesis learning goals are fundamental for the thesis design. Each program should have a set of thesis learning goals for all supervisors to follow. Mandated by institutional guidelines, the academic director and thesis coordinator are responsible for the thesis learning goals. When formulating thesis learning goals, it is important to consider the nature of thesis research for different degrees. In addition, each thesis learning goal should contribute to the attainment of the *highest* level in the targeted program learning outcomes.

Advice on Examining a Thesis for Novice Examiners

- Read the thesis manual and assessment form (with grading instructions) carefully. Consult your thesis coordinator or academic director when encountering any muddy points. Note that informal information from your close colleagues is *not always* accurate or up to date.
- Remember that this is NOT your thesis.
- Set realistic expectations based on the nature of the thesis research and thesis learning goals (Lipson, 2007).
- Time management: Schedule the examination activities based on estimated hours (see Table 2.13.1) in your agenda.

Table 2.13.1
*Estimated Hours on Examining a Thesis*⁴⁶

	Reading Twice	Taking notes	Filling out the assessment form	Discussion with the other examiner	Total (hrs.)
Bachelor's theses 6000-8000 words	60 min	10 min	30 min	20 min	2
Master's theses 10,000-15,000 words	90 min	10 min	40 min	40 min	3

- Although each examiner uses different reading strategies, Mullins and Kiley describe the commonalities among thesis examiners (Mullins & Kiley, 2002, p.376):
 Most examiners begin by reading the abstract, introduction, and conclusion to gauge the scope of the work, and by looking at the references to see what sources have been used and whether they need to follow up on any of them. They then read from cover to cover, taking detailed notes, and finally go back over the thesis to check on whether their questions have been answered or whether their criticisms are justified.
 While reading, consider the following questions (Mullins & Kiley, 2002, p.377):
 - How would the examiners have approached the problem outlined in the abstract and title?
 - What questions would the examiners expect to be answered?
 - Do the conclusions follow logically from the introduction?
 - How clearly are the procedures explained?
 - Is the bibliography current and sufficiently comprehensive?
 - Are the results meaningful and worthwhile?
 - How much work has actually been completed?
 - What is the intellectual depth and rigor of the thesis?
 - Does this qualify as genuine research—is there a clear argument?
- Read a thesis twice (Pearce, 2004). For example, once before filling in the assessment form independently, and again before the discussion with the other examiner.
- Make notes on key points/problems with each chapter and how the thesis could be restructured/revised during the course of the reading (Pearce, 2004).
- Keep a running list of typos and minor corrections (Pearce, 2004) or use a checklist to classify and record students' grammar and mechanics errors (Nelson et al., 2012).

Formulating Summative Feedback

The chapter of **formative assessment** in Section 1 has described the feedback loop: feed-up, feed-back, and feed-forward and formative feedback (Hattie & Timperley, 2007):

- Feed-up: Focus on the quality of the student's thesis work in the process or as interim products.
- Feed-back: Identify what the student has done well and not done well based on the learning intentions and assessment criteria.
- Feed-forward: Motivate and challenge the student to improve their work in the follow-up tasks.

At the end of the thesis course, it is mostly mandatory to grade the student's completed thesis using the standard assessment form defined by the study program. At this moment, supervisors need to switch to the role of examiners to do summative assessment and provide summative feedback, namely, to write "qualitative comments" in the form.

⁴⁶ The estimated hours are based on the average reading speed of most adults around 300 words per minute.

Although examiners also need to follow assessment criteria and identify strengths and weaknesses in student thesis work, the focus of summative feedback is to justify *why a certain grade is given*.

In addition, when a thesis is graded as fail, examiners may be asked to provide suggestions (feed-forward) based on the guidelines provided by the program, such as providing restricted suggestions so that students are only able to get a pass (6.0) or to provide extensive suggestions so that students can get a higher grade (7-10). Consult your thesis assessment rules and guidelines for specific information.

Quality of Assessment Form is Crucial

Thesis assessment quality is determined by the degree to which the assigned grades are aligned with the criteria and their sub-criteria/indicators. The only way to demonstrate this is to justify the grades by providing quality feedback/comments. Examine the following criteria of your assessment form:

Completeness

Fill in the basic information (student name, thesis title, your names, assessment date, etc.), grades and comments on each criterion and a total grade which is provisional in the initial independent grading phase or final after two examiners have completed the collective grading.

Alignment Between the Criteria and Comments

This alignment is a requisite for validity. Examine whether your comments are relevant to the criteria and sub-criteria/indicators. This may sound straightforward, but comments on the criteria are often irrelevant or address them only partially. Sometimes, the comments focus mainly on one or two sub-criteria, and it is not clear what they mean: Are other sub-criteria ignored or are they all sufficient/good?

Alignment Between the Grades and Comments

This alignment is a requisite for validity and reliability (intra-rater consistency). Mostly, the standard assessment form has defined what performance looks like at each grade level, in relation to the sub-criteria of each criterion. It is important that examiners follow these definitions. In addition, the comments and the grade should be congruent. For example, if the comments are mostly positive and affirmative that student thesis meets most of the sub-criteria at a high level, then assigning a 6 does not seem right.

Quality of Comments

In addition to validity (i.e., abovementioned alignments), the quality criteria for summative feedback are similar to formative feedback. They are:

- **Constructive:** Start with explicit statements about good performance and then point out poor performance and why it does not meet the criteria.
- **Specific:** In the statements about good and poor performance, provide clear explanations (i.e., examiners' interpretation of the criterion, what has been observed in student work, pointing out student work that meets or does not meet the criterion) and examples (i.e., referring to the thesis sections or page numbers).
- **Provide students with a glossary of the evaluative words used in the comments** (Bastola & Hu, 2024; Winstone et al., 2017) (see Box 2.14.3). Use these words in the formative assessment and feedback already so that students are familiar with the meanings of these words when they read the summative feedback in the thesis assessment form (i.e., transparency).

Box 2.14.3*Glossary of Common Confusing Words and What Examiners Actually Mean with These Words***Concise**

In your work, you need to explain ideas clearly but with fewer words – if you have a word limit, make effective use of it! The marker may think you are waffling. Be succinct and avoid needlessly complicated words and phrases.

Clarity

Make sure the reader can easily understand what points you have made by writing clearly and explaining why you have made these points. Sometimes it's just a case of writing straightforwardly, and not assuming the reader will automatically know what you were thinking.

Critically Evaluate/ Critically Analyze

Show that you have actively thought about and questioned the claims you are describing or making. Even if the claims are completely valid, show that you haven't just accepted them at face value.

Depth/Elaborate

Make sure you explain your arguments in detail, using examples where appropriate and working through your ideas rather than simply glossing over them.

Originality

Demonstrating your own thinking, perhaps by drawing upon research beyond the ones you learned about in class, to make an argument that not every student would have thought of.

Specific

Give a more precise and detailed account of what is being described, drawing on examples.

Structure

A way of presenting your work so the reader can follow the argument. Make sure your paragraphs are in a logical order, that you show the connections between different paragraphs, and that each section has good beginning and ending sentences.

Synthesis/Integration

Show how different sources and theories go together to make a good argument. A lack of synthesis could mean your essay reads more like a list of research than an argument.

Checklist of Ensuring Thesis Assessment Quality

Transparency

- Thesis learning goals, expectations, assessment forms, and the overall thesis process (including clear guidance and visual structure) are communicated transparently to students.
- The use of GenAI is clearly regulated, including whether it is allowed, to what extent, and under defined ethical and academic integrity guidelines.
- The **rubric** is aligned with the permitted level of AI use. For example, if spelling and language accuracy are explicitly assessed within the marking criteria, this implies that the student is expected to demonstrate independent writing ability, and therefore the use of GenAI for generating or correcting text may be restricted or not permitted for those aspects of the assessment.
- AI documentation requirements (e.g., tools used, purpose, verification, and integration) are clearly defined, including how they are assessed and the consequences of non-compliance.
- Thesis components such as proposals and intermediate deliverables are clearly explained in terms of their purpose, assessment format, feedback, and role in the overall process.
- Supervision practices are transparent, including ongoing discussion of criteria, feedback processes, GenAI use, and clear consequences for failing to meet requirements.

Reliability

- Assessment criteria and standards (i.e., performance levels, from excellent to insufficient) are consistently applied to support **intra-rater reliability**.
- Examiners regularly revisit and adhere to established criteria and standards during the grading process to maintain scoring consistency over time.
- Consistency across examiners (**inter-rater reliability**) is ensured through shared understanding of criteria and standards.
- Anchor theses or benchmark examples are used to illustrate assessment criteria and support consistent interpretation of standards.
- Calibration or alignment sessions are conducted to ensure examiners apply criteria and standards uniformly.

Validity

- Thesis learning goals are clearly defined at the program level, aligned with program learning outcomes, and appropriate to the level and nature of the degree.
- Assessment forms are complete, including all required information, grades, and comments for each criterion, as well as a clear overall grade.
- Comments are aligned with the assessment criteria and sub-criteria, addressing all relevant aspects of performance.
- Grades are consistent with the comments and reflect the defined performance levels for each criterion.
- Feedback is constructive and specific, providing clear explanations, examples, and consistent use of evaluative language to support student understanding.

Acknowledgement

We would like to thank Dr. Astrid Kramer for her review and feedback, which helped improve the quality of this chapter.

2.14

Quick Guide to Thesis Oral Defenses⁴⁷

Ya-Ping (Amy) Hsiao, Brenda den Oudsten, Jorna Leenheer, Peter de Goeij,
and Marijn van Wingerden

Introduction and Learning Goals

There has been renewed interest in oral **assessments** in higher education (Bayley et al., 2024; Fenton, 2025). Initially reintroduced during the COVID-19 pandemic to enhance student engagement and enable valid examinations (Gaines & Burrows, 2024), they are now also promoted as a response to Generative artificial intelligence (GenAI) tools such as ChatGPT (Moorhouse et al., 2023; Rudolph et al., 2023; Sotiriadou et al., 2020).

Oral examinations are a standard component in master's programs and are increasingly being adopted in bachelor's programs. In general, there is growing interest in their use, partly driven by concerns related to GenAI tools (Wagner et al., 2022) and their implications for written assessments. While institutional guidelines are often still absent, it is essential to clarify the purpose, design, and assessment of oral examinations within higher education. Moreover, as oral defenses are increasingly being adopted in bachelor's programs, questions arise about what should be adapted, developed, and assessed. This motivates us to offer guidance for programs, supervisors, and examiners.

The thesis oral defense, when conducted prior to grade determination, serves primarily as an examination tool to assess a student's mastery of research, subject knowledge, and scholarly communication. This chapter therefore frames the oral defense in terms of assessment criteria and procedures, rather than as a ceremonial or celebratory event. That is, a celebratory element can only be included after the formal assessment has taken place.

The purpose of this chapter is to provide guidance for program directors, thesis coordinators, and thesis supervisors on integrating the oral defense into thesis assessment. It highlights key considerations for ensuring **transparency**, **reliability**, and **validity** by conducting defenses fairly, consistently, and in accordance with academic standards.

The chapter is structured as follows. For thesis examiners, it first discusses the format of the oral defense and its relationship to the written thesis. It then examines the purposes and assessment domains of the defense, outlines the key design dimensions, and presents specific implementation considerations related to grading.

As graduation theses are assessed at the program level, the chapter also provides guidance for developing formal policies or regulations, such as a thesis manual or thesis assessment guidelines (see Appendix 2.14.1). Because oral defenses require careful preparation, Appendix 2.14.2 offers a checklist to help examiners cover all essential points.

47 This chapter focuses on thesis oral defenses at the graduation level while Chapter 2.4 focuses on oral exams at the course level.

To ensure fair and equitable assessment, students also need adequate preparation for this intensive evaluation; Appendix 2.14.3 provides a checklist to guide student preparation.

Overall, this chapter aims to support institutions in designing and implementing oral defenses that are consistent, transparent, and aligned with academic standards. After applying the guidance in this chapter to their own assessment practices, thesis supervisors and examiners will be able to:

- Describe their thesis defense in terms of its design, assessment criteria, and procedures.
- Communicate assessment information to students effectively.
- Prepare students throughout the thesis process, including a near-end mock defense.

Oral defense: Format, Purpose, and Connection to the Thesis

Public –Private Format

The institution defines the format and determines the options for programs to choose in terms of guidelines and procedures. There are several elements important to consider when defining the format, such as: relationship between oral defense and thesis product, purpose of the defense, assessment domains, and private – public format of the defense.

At the bachelor's and master's levels, thesis defenses are mostly private at Tilburg University. Public defenses highlight a student's ability to present and discuss their work openly, and they are more commonly required for PhD candidates. Private defenses take place behind closed doors, involving only examiners and the student. The examiners are, in many cases, the supervisor and a "second reader". Private defenses provide space for in-depth academic examinations without the need for ceremonial or politeness restrictions.

But note that Bachelor and Master defenses can be public in case a program allows it. A public defense is in principle open to anyone waiting to attend, though questioning can be restricted to examiners only.

Relationship Between the Written thesis and Oral Defense

The written thesis generally carries greater weight in the overall assessment than the oral defense (Golding et al., 2014). But depending on the goal of the defense, its result can lead to an adaptation of the grade or even failure of the project. Though failure is not expected under normal circumstances, it can take place if fraud, such as unacceptable use of GenAI, has taken place.

Purposes of Defenses

Any quality assessment starts with well-articulated purposes, that is, clearly defined thesis learning goals, including both disciplinary thinking and communication skills. The most common purposes of thesis defenses include (Watts, 2012; Wellington, 2010a):

1. Verify authorship and integrity of the thesis.
2. Assess the student's ability to defend and contextualize their research within the broader field.
3. Evaluate the ability to think on their feet, communicate effectively, and the capacity to explain the implications of the work.

When an oral defense is included, these purposes should be explicitly integrated into the thesis learning goals. For example:

- Demonstrate authorship and academic integrity of the thesis.
- Defend and contextualize the research within the relevant field.
- Communicate research clearly and effectively in an oral format.
- Respond thoughtfully to questions and critiques.
- Explain the implications and limitations of the research.

Design Dimensions of Oral Defenses

Assessment Domains

A graduation thesis at the exit level is designed to test whether students can apply and integrate the knowledge and skills acquired throughout the program. It also evaluates their ability to identify and address gaps in their learning, apply their work to slightly different contexts, as well as their responsibility and independence. Consequently, the thesis, through both the written product and the oral defense, should assess the majority of the program's **learning outcomes**.

In addition to the knowledge and skills evaluated in the written thesis, programs should clarify what is assessed during the oral defense. For example, oral defenses typically begin with a presentation or pitch, and it should be determined whether this component is graded. Grading is particularly important if communication skills have *not* been formally assessed in prior program courses.

Assessment Form

Using a standardized assessment form based on a pre-defined **rubric** promotes transparency by making expectations and criteria clear to both students and examiners (Fenton, 2025; Joughin, 2010). It enhances reliability and **fairness** by supporting consistent judgments across candidates and examiners, while reducing subjectivity. In addition, it strengthens the validity of the oral defense by ensuring that assessment is aligned with the intended learning goals and provides a clear record of decision-making.

A standardized assessment form can include the following elements:

- Criteria (linked to each purpose/domain)⁴⁸
- Questions (prepared in advance)
- Examiner responsible (who asks each question)
- Notes on student answers
- Grade
- Overall comments

Prior to the defense, examiners must prepare tentative questions for each criterion, using the official question list provided by the program (see Table 2.14.1 for example).

Table 2.14.1

Suggested Format of Assessment Form for an Oral Defense

Criteria	Questions	Asked by which examiner	Notes on student answers	Grade
	Prior to the defense, the examiners should fill out tentative questions per criteria, based on the list of questions provided by the program. Asking different or alternative questions is allowed, provided it complies with the guidelines of your program.			
Overall comments				

⁴⁸ Open Canvas Commons. In the search bar, type “rubric”. Browse results to find oral assessment rubrics.

Be Aware of Examiner Bias

Bias may occur when factors unrelated to academic performance, such as gender, age, ethnicity, educational background, language proficiency, or social class, consciously or unconsciously influence an examiner's judgment. To ensure fairness and equity in oral defenses, examiners should be aware of the common bias (e.g., stereotype, halo or horn, leniency or severity effect) and apply the corresponding mitigation strategies (i.e., using a standardized assessment form) (Hurford, 2020; Joughin, 2010).

Grading Procedure

Programs should clearly communicate the following information to examiners and students both at the beginning of the thesis course and *again* at an appropriate time before the defense.

- Clearly state the time allocated for grading and writing comments after the defense.
- Indicate whether the final grade will be shared with the student immediately after the defense (most common practice), particularly when adjusting the written thesis grade by small increments (e.g., ± 0.5 points) or a preliminary grade (if the grade cannot be lowered further).
- Define the procedure for cases where examiners cannot reach consensus on the collective grade.
- Define possible outcomes of the oral defense.
 - Pass: All criteria are met; no further action is required.
 - Pass with conditions: The student passes, but the examiners require minor clarifications, corrections, or an additional written response. No second defense is required.
 - Fail: The student does not meet oral defense criteria and standards (e.g., insufficient mastery, inability to defend findings, inadequate presentation, or unresolved authorship concerns).
 - Share institutional policies in cases where students object to the assessment outcomes of the oral defense, so that both examiners and students are aware of and can follow the appropriate procedures.
- Resit possibilities of the oral defense. Students who fail the oral defense **may be granted with** one resit opportunity.
 - The resit must be scheduled within the same academic year. Note: Consult your institutional policy for students with special circumstances.
 - The resit will assess the same criteria: presentation, content mastery, and responses.
 - If the student fails the resit, the thesis requirement is not completed, and the degree cannot be awarded.

Summative Feedback to Students

Clearly specify how summative feedback will be given to students (oral or/and written) and what information will be shared with students after the defense: either the full assessment form (with detailed notes and grades), or only the overall comments and final grade. Consult your institution's Examination Board for further advice.

Good Preparation is the Key

In addition to clear policy, guidelines and procedures, preparing thesis examiners and students for the oral defense is essential (Joughin, 2010).

Beyond **constructive alignment**, it is important for supervisors to integrate preparation for the defense into the supervision activities themselves. Students need structured learning activities that support their progression toward the oral defense (and thus the intended thesis learning goals). Incorporating practice oral defense during supervision activities helps make assessment expectations explicit and allows students to experience how learning goals will be graded. For example, in thesis circles (small-group supervision) (Hsiao et al., 2025; Rajagopal et al., 2021), students can first present their intermediate work in pairs and then share it with the full circle group. For programs with large student cohorts, an online workshop can be an effective way to address topics such as presentation skills, adherence

to presentation time, use of audiovisual media, non-verbal communication skills, answering examiners' questions, practical aspects of the presentation, and managing anxiety prior to the oral defense (Chust-Hernández et al., 2025).

These experiential opportunities for trial, error, and feedback support skill development and can significantly reduce student anxiety (Hurford, 2020). Research indicates that some student anxiety is linked to unfamiliarity with the oral defense format (Fenton, 2025). When students have a clearer understanding of what to expect, their anxiety tends to decrease in subsequent oral assessments (Roecker, 2007). Appendix 2.14.2 outlines the preparation requirements for examiners involved in the oral defense.

To help students gain confidence, reduce anxiety, and demonstrate their competence in an oral defense, we can support them in several ways (Rawls et al., 2015; Wellington, 2010b):

- **Skills:** Students can develop communication and debate skills through exposure to academic research cultures, such as attending conferences, seminars, and engaging with scholarly networks. Mock defenses and guided practice with peers or supervisors can also help refine these skills.
- **Content:** Students can prepare by reviewing their research thoroughly and anticipating likely questions. Supervisors can guide them on key topics and potential lines of inquiry.
- **Conduct:** Although the dynamics of the viva are unpredictable, students benefit from understanding its structure and ground rules, including participants, duration, purposes, possible outcomes, and their rights (e.g., consultation on examiners, supervisor attendance, and examination procedures) (see Appendix 2.14.1).
- **Clarifying misconceptions:** Students should be informed about what to expect, correcting hearsay or extreme anecdotes, and emphasizing the formative (i.e., knowledge, skills, and character strengths developed in the preparation process) as well as summative value (i.e., confirmation of developed capability) of the defense.

Appendix 2.14.3 provides additional guidance and resources to support student preparation for the oral defense.

Acknowledgements

This chapter is based on an analysis of oral defenses in several thesis manuals from TSB, TSHD, and TiSEM.

Appendix 2.14.1

Guidelines and Procedures: Examples and Instructions

To ensure a structured, fair, and supportive environment for the thesis defense, the program should establish clear guidelines and procedures in accordance with the institution's policy. These guidelines should ensure consistency in **assessment** while supporting the evaluation of the student's knowledge, reasoning abilities, and communication skills.

The sections below contain two types of material:

- **Example formulations**, which illustrate how guidelines and procedures could be written. These are presented in *italics*.
- **Instructions for programs**, which indicate what issues must be decided, specified, and regulated locally. These are presented in **bold**.

The example formulations are illustrative only and not prescriptive. Once the program's own guidelines have been formalized, they should be shared with both examiners and students.

Grading policies

Sequential assessment: The written thesis is assessed first, and only if it is passed does the student proceed to the oral defense.

[[Instructions for programs]]

If a sequential model is used, the program must specify:

- **Whether the written thesis must be passed before the student is allowed to proceed to the oral defense**
- **What counts as a “pass” for eligibility (e.g., minimum grade, rounding rules, compensations)**
- **Whether and how a resubmission opportunity is offered if the written thesis is failed**
- **The timeframe for revision and resubmission**
- **How many resubmission attempts are allowed**
- **Whether a failed written thesis automatically blocks or postpones the defense**
- **How the decision is formally recorded and communicated to the student**

These rules must be consistent with institution-level and Examination Board's regulations and must be applied uniformly to all students.

[[Example formulation]]

- *The written thesis is graded first.*
- *Only students whose thesis is graded as Pass are eligible to proceed to the oral defense.*
- *If the thesis is graded as Fail, the student may revise and resubmit within a defined timeframe before being reconsidered.*

Interim assessments

[[Instructions for programs]]

When a program includes compulsory interim assessments, the program must define and document the following aspects. These rules must be communicated in advance and applied consistently:

- **Clarify the status:** The program must state whether interim submissions and/or oral discussions are purely formative or have summative consequences.
- **Define what counts as completion:** It must be specified whether submission of the written parts alone is sufficient, or whether satisfactory participation in discussions and/or oral checks during supervision meetings is also required.
- **Specify consequences in advance:** The program must decide and state clearly
 1. **What happens if a student does not submit all required parts.**
 2. **What happens if a student submits the parts but cannot demonstrate sufficient authorship or understanding.**
 3. **Whether and how such situations affect the right and/or timing of submission of the final version (for example, whether the oral defense can only take place after interim assessments have been successfully completed as part of the first submission opportunity).**

Weighting

[[Instructions for programs]]

The program must define and document the weighting of the written thesis and oral defense in the final grade. This includes:

- **Proportion of each component:** Specify what percentage of the final grade is assigned to the written thesis and to the oral defense.
- **Role of the oral defense:** Decide whether the oral defense serves as:
 - A separate, fixed component of the grade, or
 - A moderating/adjusting component that can slightly increase or decrease the final grade based on performance.

[[Example formulation]]

The written thesis contributes XX% of the final grade. The oral defense contributes XX% of the final grade.

or

The oral defense may function as a moderating or adjusting component (for example, an excellent defense may raise the final grade by 0.5, or a very weak defense may lower it by 0.5).

Balanced assessment criteria between written thesis and oral defense

[[Instructions for programs]]

Balanced assessment criteria aim to assess both the quality of the scholarly work as documented in the written thesis and the student's demonstrated competence in defending and communicating the research in the oral defense. The written thesis primarily assesses the work in terms of depth, rigor, and originality, while the oral defense evaluates the student's ability to articulate, justify, and contextualize the research within the broader field. Together, these components provide a comprehensive assessment of academic quality at the exit level. To ensure transparency and consistency, the program must define and document clear assessment criteria for both components:

For the written thesis, the criteria should normally cover:

- **Major sections (e.g., introduction, literature review, methods, results, discussion, conclusion)**
- **Overall qualities such as originality, rigor, coherence, and clarity**

For the oral defense, the criteria should normally cover:

- **Ability to explain, justify, and contextualize the research**
- **Clarity and precision of communication**
- **Ability to respond critically and reflectively to questions**

Professional and scholarly conduct (including, where relevant, performance in the presentation part)

The program must ensure that these criteria are communicated to students and examiners in advance and applied consistently.

Preliminary submission of written thesis

[[Example formulation]]

- *The student must submit the full version of the written thesis via an authorized platform (e.g., Canvas) before the specified deadline.*
- *A plagiarism check must be performed upon submission.*
- *The student must follow the program's procedure (e.g., sending an email to XX or completing an application form) to register for the defense, including:*
 - *Scheduling the date and time with all relevant participants (e.g., supervisor, second reader, company supervisor, and, where applicable, an audience in the case of a public defense). For theses based on a company internship, the student is responsible for inviting the company representative.*
 - *Reserving a room for the presentation and defense (unless this is arranged by the department or program).*
- *If the thesis receives a passing grade (e.g., 5.5 rounded up to 6.0, or higher) for the thesis as a whole, the student becomes eligible for the defense.*

Resit of the written thesis

[[Example formulation]]

If the written thesis does not receive a passing grade:

- *The student is granted an extension for the period specified in the institutional rules, guidelines, or thesis manual to revise the thesis. During this period, the defense application is marked as "not eligible."*
- *The revisions must address the feedback provided by the supervisor and second reader.*
- *The revised thesis must be submitted via the learning management system (e.g., Canvas) and is subject to a new plagiarism check.*

Redo after failing the resit of the written thesis

[[Example formulation]]

If both examiners assess the revised thesis as insufficient:

- *The student must start a new thesis project.*
- *The student must work with a different supervisor.*
- *The student must choose a different subject.*

The defense application is withdrawn from the system.

Planning the defense

[[Instructions for programs]]

The program must explicitly define who is responsible for organizing and coordinating the defense (e.g., the thesis coordinator, program administrative staff, or the student). This responsibility must be documented and communicated to students and staff in advance.

[[Example formulation]]

Once the written thesis has received a passing grade:

- *The defense application has been formally registered and approved.*
- *The responsible party confirms the date and time with all participants (e.g., supervisor, second reader, company supervisor, and, where applicable, an audience in the case of a public defense).*
- *The room for the presentation and defense is confirmed or reserved.*

Duration

[[Instructions for programs]]

The program must define the purpose(s) of the oral defense and specify:

- **The total duration of the defense**
- **Any minimum and/or maximum duration**
- **The time allocation between presentation, questioning, and deliberation (if applicable)**
- **Consequences when students are late for the defense**

These rules must be documented and communicated in advance.

[[Example formulation]]

Depending on the purpose and format of the oral defense, the total duration of the defense is XX minutes (minimum: XX minutes; maximum: XX minutes).

During the defense

Consent for Recording

[[Instructions for programs]]

The program must decide and document whether the defense may or must be recorded, under what conditions, and for what purpose (e.g., inspection or appeals). The rules must specify:

- **Whether recording is optional or standard practice**
- **How the student's consent is requested and documented**
- **What happens if the student does not grant permission**
- **How recordings are stored, accessed, and deleted**

These rules must comply with institutional and legal regulations on privacy and data protection.

[[Example formulation]]

When two examiners are present, the student is asked at the start of the defense for permission to record the session for inspection purposes. If the student does not grant permission, the defense may still take place, but the student accepts that inspection is limited (e.g., based on the examiners' written notes, if available).

Oral presentation

[[Instructions for programs]]

The program must specify:

- **Whether an oral presentation is required**
- **The language(s) in which the presentation may be given**
- **The required or allowed duration**
- **Whether the defense is public or closed, and what this implies for the intended audience**

[[Example formulation]]

- *The student delivers an oral presentation in English or Dutch lasting XX-XX minutes.*
- *In the case of a closed (private) defense, the presentation is scientifically oriented and suitable for a peer-level audience (i.e., understandable to fellow students).*
- *In the case of a public defense, the presentation is open to all interested attendees and is oriented toward a general, non-specialist audience.*

Defense session (questioning)

[[Instructions for programs]]

The program must define:

- **Whether the questioning session is public or closed**
- **Its approximate duration**
- **Who may ask questions (e.g., only examiners, or also the audience in public defenses)**
- **The scope of the questions (e.g., thesis content, methods, theoretical background, implications)**

[[Example formulation]]

Immediately following the oral presentation, a closed session (for closed defenses) or a public session (for public defenses) of approximately XX minutes takes place.

During this session, the examiners ask questions about the content, methods, and conclusions of the student's thesis.

Grading session

[[Instructions for programs]]

The program must define and document:

- **Whether examiners first assess independently before deliberation**
- **Which assessment instruments are used (e.g., rubric, assessment form)**
- **How disagreements between examiners are resolved (e.g., arithmetic averaging, decisive vote of the second reader, involvement of a third examiner)**
- **Who determines and formally records the final grade**

These procedures must be communicated in advance and applied consistently.

[[Example formulation]]

After the defense session, the examiners first complete the program's assessment form or rubric independently, covering the presentation and/or defense as applicable. They then hold a closed meeting to discuss their assessments and determine the final grade in accordance with the program's procedures (for example, by calculating an arithmetic average or by giving decisive weight to the second reader's assessment).

Announcement of the grade

[[Instructions for programs]]

The program must specify:

- **Whether the grade is normally announced immediately or at a later time**

- **Whether the announcement may be public or must be private**
- **Who is present at the announcement**
- **Who formally communicates and records the result**

[[Example formulation]]

In consultation with the thesis supervisor, the student may decide whether the announcement of the thesis grade takes place in public or in private.

Graduation requirement

[[Instructions for programs]]

The program must clearly define and communicate the conditions under which a student is eligible to graduate. This includes:

- **Confirmation of program completion:** Ensure that students have successfully completed all curriculum components as specified in the Education and Examination Regulations (EER) or equivalent program rules.
- **Verification by the Examination Board:** Define the procedure for the Examination Board to formally verify that all requirements for awarding the degree have been fulfilled before graduation.
- **Clarification to students:** Communicate to students that passing the thesis and oral defense alone does not automatically result in graduation, and outline any additional steps or checks required.
- **Documentation and record-keeping:** Ensure that records of completion, defense results, and Examination Board approvals are properly documented and stored in accordance with institutional policies.

[[Example formulation]]

Passing the thesis and oral defense does not automatically mean graduation. To graduate, students must complete all curriculum components of the program, as specified in the Education and Examination Regulations (EER). The Examination Board formally verifies whether all requirements for awarding the degree have been fulfilled.

Role and tasks of the examiners (Supervisor and Second Reader)

[[Instructions for programs]]

To ensure fairness, consistency, and transparency in assessment, the program must define, organize, and document the following:

- **Information session:** The academic director or thesis coordinator organizes an information session for examiners to ensure a shared understanding of the guidelines and procedures for the thesis defense.
- **Calibration session:** Conduct a calibration session to train examiners on how to apply the grading rubric consistently for the oral defense.
- **Instructions for examiners and students:** Provide clear written instructions for both examiners and students. Checklists are available in Appendix 2.14.2 (for examiners) and Appendix 2.14.3 (for students).

Appendix 2.14.2

Checklist for Examiners – Preparing the Oral Defense

(Murray, 2015; Wisker, 2012)

[[Please modify the texts to suit your program contexts.]]

Before the defense

Preparation and communication

- Ensure students have access to departmental or program procedures, regulations, manuals, or handbooks.
- Clarify the expected duration, format, and composition of the examining committee (e.g., two examiners), as well as proper etiquette (e.g., how to address examiners, dress code, and expressing thanks at the end of the defense).
- Explain expectations regarding questions during the defense, including content mastery, presentation, and critical discussion.
- Inform students what materials (e.g., notes and pen) they may bring to the defense and how to mark important or likely questions in the thesis using post-its.
- Organize workshops (Chust-Hernández et al., 2025) or mock defenses for students to rehearse. Offer constructive feedback for students to improve.
- Be aware of any rater effects (e.g., halo effect, horns effect, leniency or severity bias, personal bias) which may influence your judgment and compromise the **fairness** or accuracy of the **assessment** (Van Berkel et al., 2023).

Planning the defense

- Allocate time for each component: student presentation, questions by each examiner, discussion, and deliberation.
- Share the schedule with the student in advance so they can prepare accordingly.

During the defense

Professional conduct and logistics

- Ensure the examination room is properly set up with all required equipment and a professional environment.
- Clearly explain the oral defense process to the student at the start, including format, timing, and rules.
- Chair the defense to maintain fairness, adherence to regulations, and smooth progression of the session.
- Monitor the student's presentation to keep within the allotted time, using polite signals if necessary.
- Coordinate questioning: take turns asking questions, keep them focused, and avoid unnecessary repetition.
- Encourage concise, clear answers, guiding the student back on track if they stray from the topic.

Questioning

- Phrase questions clearly and avoid leading questions.
- Organize questions by thesis section and stage them in a logical sequence.
- Ask one question at a time; avoid combining multiple questions simultaneously.
- Use adequate body language (e.g., maintain an open posture, use calm, steady eye contact, controlled gestures, friendly expressions, avoid interrupting physically, demonstrate active listening).

Assessment focus

- Verify authorship and originality of the thesis.
- Grade the student's presentation based on the criteria and standards.
- Assess understanding of theoretical frameworks and conceptual models.
- Examine the rationale for research questions, methodology, and analysis.
- Check ability to respond to critical questions and defend conclusions.
- Record decisions, justifications, and any recommendations for resits or revisions in a formal report.

Handling responses and difficulties

- Allow students to acknowledge uncertainty and reasoning gaps without penalizing honesty.
- Ensure students understand the boundaries of the research and stay within scope.
- Provide fair guidance if a student struggles during the defense.

Feedback and follow-up

- Communicate results clearly (i.e., pass, conditional pass, or resit requirements).
- Provide constructive feedback for improvements, future research, or publication.
- For immediate graduation after the oral defense: Verify that all curriculum and degree requirements are met before graduation recommendation.

List of questions for examiners

Formulate generic questions based on the aspects outlined by (Wisker, 2012) and organize them according to the three phases of the defense (Trafford, 2003):

Opening

Focus on conceptualization and the rationale behind the research (why the study was conducted).

- Research rationale and context
 - Why did you choose this topic, and why is it relevant in today's context?
 - What motivated you to pursue this research?
 - How does your research fit within the broader field of study?
- Research questions and scope
 - How did you formulate your research questions?
 - What boundaries or limitations did you set for your study? Why?
 - What gap in existing knowledge does your research address?
- Conceptual framework
 - What conceptual framework guided your research?
 - Which theories or models did you rely on, and why were they appropriate?
 - Did you consider alternative theoretical approaches?

Consolidation

Assess the student's understanding of the thesis content, methodology, and arguments.

- Methodology and methods
 - How does your chosen research design align with your conceptual framework/concepts/theories and address your research questions?
 - How did your sampling strategy reflect the requirements of your research questions and the concepts/theories within your conceptual framework?
 - How does your data collection process support the concepts and variables defined in your framework, and what steps did you take to ensure **reliability** and **validity**?
 - How do your data analysis methods link back to your conceptual framework and research questions, and why were these techniques appropriate?
- Challenges and unexpected findings
 - What problems or difficulties did you encounter during your research, and how did you address them?
 - Were there any findings that surprised you? How did you interpret them?
- Conclusions
 - What are your main factual conclusions?
 - What conceptual or theoretical conclusions can be drawn from your study?
 - How do your conclusions relate back to your research questions?

Closing

Encourage the student to reflect on the findings and discuss their implications or applications.

- Contribution to knowledge
 - What is the contribution of your work to the academic field?
 - How might your findings be applied in practice, policy, or future research?
- Academic Level
 - In what ways does your work reflect the requirements of a bachelor's/master's thesis?
 - How does the scope and depth of your research align with the level of your program?
- Future Work
 - What would you do differently if you had more time or resources?
 - What are the most promising directions for further research building on your study?
- Closure
 - Would you like to clarify or expand on any of your earlier answers?
 - Is there anything we have not asked that you would like to add or comment on now?

Appendix 2.14.3

Student Checklist – Thesis Submission and Oral Defense Preparation

[[Please modify the texts to suit your program contexts.]]

Preliminary submission of written thesis

Once the full version of the written thesis is completed.

Submit on the Learning Management System (e.g., CANVAS)

- Upload your thesis under the assignment “Plagiarism scan thesis”. Be aware that submitting your thesis does not guarantee a passing grade.
- Failure to submit will be considered a fail.

Send to faculty

Email the submitted version to your supervisor, with cc to the thesis coordinator and the department secretariat.

Share with second reader

As soon as the second reader is assigned, send them your thesis. Begin planning your defense date with all required participants.

Inform stakeholders

Notify all involved parties (e.g., supervisor, second reader, thesis coordinator, company supervisor) that you have submitted your thesis. Provide them with the planned defense date.

Checklist for students – Preparing the oral defense (Foote, 2016; Murray, 2015; Wisker, 2012)

Preparation and procedures

- Review department or program oral defense procedures and relevant regulations, manual or handbook.
- Recall the expected duration and format of the defense.
- Recall who will attend (examiners, chair, observers).
- Remember to bring pen and paper to take notes during the defense.

Research and content

- Be prepared to explain why you chose your topic and research questions.
- Be able to situate your research within the broader academic field.
- Be able to clearly describe your conceptual framework, theories, and key contributions in your own words.
- Be able to justify your choice of methodology, data collection, and analysis.

- Be aware of the limitations and scope of your study.
- Practice answering questions suggested by your program or supervisors (see below).

Handling questions

- Practice answering questions clearly, logically, and confidently. Rehearse the defense with your peers or attend the mock sessions/workshops organized by your program or supervisors.
- Make sure that your answers are closely related to the examiners' questions.
- Ask for clarifications when you do not understand any questions. Do not try to answer questions you do not understand.
- Stick to what is written in the thesis. Do not introduce new information or ideas which you do not really know.
- Be able to use strategies if you cannot fully answer a question (acknowledge uncertainty, outline reasoning, propose next steps).

Practical and professional

- Recall where to wait before the defense and logistics of entering the room.
- Confirm what materials or equipment you may need.
- Dress appropriately according to academic standards.

Professional dispositions

- Maintain calm and professional (know how to behave in a stressful situation).
- Focus on discussion of your work, not personal judgment.
- Take note of feedback for future revisions or research directions.

List of questions for students to prepare for the defense (Foote, 2016; Möwes, 2015)

[[Please decide whether to share the full list or just a selection of these questions with students.]]

General/Introductory questions

- Summarize your thesis/dissertation in a few sentences.
- Discuss your major findings.
- Explain your plan for further research.
- Suggest ideas for publication of the thesis/dissertation (whole or part).
- Discuss the implications of your study for the work of other scholars or research groups.

Contribution and originality

- What are the major weaknesses of your study?
- Why were a particular method of analysis chosen?
- Suggest a research question for someone extending your study.
- Express opinions on broader issues beyond the study's findings, clearly labelled as personal opinion.
- How does your thesis/dissertation qualify as an original contribution to knowledge? (for master's program)

Critical reflection and growth

- Are there scholars who might disagree with your conclusions?
- In what ways has this work contributed to your growth as a scholar?
- If you were to start again, would you do anything differently?
- Are there courses or program components you think should be added to better prepare students for this research?

2.15

Quality Assurance of Course-Level Assessment⁴⁹

Ya-Ping (Amy) Hsiao, Astrid Kramer, and Cor Sluijter

Quality Assurance in Higher Education

In higher education, each institution establishes an overarching assessment framework that sets the principles and standards for assessment quality. Based on this institutional framework, each institution develops its own assessment policy, which specifies criteria and procedures for ensuring assessment quality. Institutions typically operationalize these policies through a structured quality assurance cycle to ensure that assessment practices consistently align with established standards.

Quality assurance is defined as “the collections of policies, procedures, systems and practices internal or external to the organization designed to achieve, maintain and enhance quality” (Harvey, 2004-2026). The Accreditation Organization of the Netherlands and Flanders (NVAO, 2018) describe an effective quality assurance system as “simple, univocal, and verifiable aims and objectives; procedures to safeguard quality; embedding of the Plan-Do-Check-Act (PDCA) cycle in the organization; ‘hard controls’; periodic evaluations; and systematic monitoring of improvements” (NVAO, 2018, p.6).

The Plan-Do-Check-Act cycle (PDCA cycle) is a widely used model for continuous improvement in organizational and educational contexts (Deming, 2023). In the planning phase, goals are set, and strategies are developed to improve existing processes. The doing phase involves putting these plans into action. During the checking phase, results are compared with the initial plans and strategies to see what worked well and what needs adjustment. The acting phase focuses on making successful changes becoming part of regular practice and preparing for the next round of improvement. While it enables systematic, stepwise quality enhancement, effective implementation requires substantial time, attention, and commitment from stakeholders, as each phase depends on careful planning, observation, and evaluation. Without sustained engagement, improvements may remain superficial and fail to achieve meaningful progress.

Within the institution’s quality assurance system, tasks, roles, and responsibilities are divided among the vice-dean of education, the Examination Board, policy and assessment teams, academic directors, and course coordinators. Examination boards have the statutory duty to determine that students have achieved the intended final qualifications and to safeguard the quality of the examination (Articles 7.12, paragraph 2, and 7.12b, paragraph 1, section a, of the higher education and Research Act (Ministry of Education, Culture and Science, 1992)). The Examination Board

49 An extended version of this chapter will be submitted to a journal for publication.

holds responsibility for ensuring the quality assurance of **assessments** and examinations, including both the quality of the assessments (i.e., **validity**, **reliability** and **transparency**) and of the organizational processes and procedures (Dutch Education Inspectorate, 2025). Its dual role involves both monitoring and evaluating assessment standards and providing advice, regulations, and guidance to lecturers, students, and administrators.

The internal quality assurance system aims to both care for (in Dutch: *zorg*) and assure (in Dutch: *borg*) the quality of course-level assessments. To achieve this, most Dutch higher education institutions have installed assessment committees (van de Veen et al., 2018). These committees are composed of independent academic staff with expertise in assessment. They carry out quality assurance tasks assigned by the Examination Board, which oversees assessment quality, and the vice-dean of education, who is responsible for overall educational quality within the institution.

Design Dimensions in Quality Assurance of Assessment

This section identifies four quality dimensions derived from the assessment pyramid (Sluijsmans et al., 2012) and institutional policies: vision of assessment quality, assessment competences, assessment organization, and assessment program. Their design for effective quality assurance is interpreted using three key sources: Van de Veen et al. (2018) on assessment committees, Williams (2016) on quality assurance and enhancement, and Deming's (2023) PDCA cycle.

Vision of Assessment Quality

The vision, developed by key players (e.g., stakeholders of the institution) within the quality assurance system, sets the strategic direction of quality assurance practices. Both quality assurance and quality enhancement are essential, yet often treated as separate trajectories (Gosling & D'Andrea, 2001; Williams, 2016). Overemphasis on external monitoring can narrow the focus to compliance rather than continuous improvement. As Van de Veen et al. (2018) noted, effective quality assurance involves not only verifying but also facilitating processes that enhance quality. Integrating **constructive alignment** (Biggs & Tang, 2011), a didactic framework to education, ensures that assessment is embedded in the broader educational process.

Assessment Literacy: Examiners

Quality assurance is intended to strengthen examiners' competencies in course assessment design, test administration, and grading practices. Following a complete quality assurance cycle, examiners should ideally gain a deeper understanding of the organization's assessment quality standards and be better equipped to enhance their future assessment practices. Achieving this, however, requires a structured system of feed-up, feedback, and feed-forward that is closely aligned with the evaluation process. Such a system informs examiners about the expected standards (feed-up), how they should respond to feedback, and support the application of insights to improve future assessment practices (feed-forward). For feedback to be effective, it must be dialogical, enabling examiners to understand both the intent and content of feedback provided (Ajjawi & Boud, 2018). Consequently, the integration of quality assurance mechanisms with carefully designed feedback processes is essential for the ongoing development of examiners' assessment competencies.

Assessment Organization: Roles and Responsibilities

Assuring assessment quality is a collaborative process involving multiple stakeholders with roles and responsibilities, including the vice-dean of education, Examination Board, academic directors, and coordinators. These stakeholders are supported by education and assessment specialists. The assessment organization ensures quality assessment that meets national and external frameworks. Clear definition of roles and responsibilities in a context of constructive and continuous dialogue are critical for integrating quality assurance and enhancement. When these stakeholders operate in isolation, for example when assessment committees evaluate course assessment quality while academic

directors monitor improvements, the coordination and application of findings to enhance quality can be compromised. Continuous monitoring ensures that enhancement follows from quality assurance activities.

Assessment Program

Although this chapter focuses on quality assurance of course-level assessment, each course must be considered within the broader program context and its role in the overall assessment program. Program policies and guidelines determine assessment types, weighting, sequencing, and alignment with program **learning outcomes**. For example, knowledge outcomes should be organized according to a specific learning taxonomy, like Bloom's taxonomy (Anderson & Krathwohl, 2001) or Miller's pyramid (Miller, 1990), based on the discipline's requirements. Skills outcomes should progress through stages, including introduction, development, intermediate, and advanced levels. Additionally, the assessment program outlines how assessments should be sequenced and developed. For instance, a first-year bachelor's course might give more attention to assessing lower cognitive skills like knowledge and comprehension, while a third-year course should focus more on higher cognitive skills such as evaluation and synthesis. Program-level oversight ensures coherence and alignment across courses. Neglecting this context can lead to misaligned assessment practices that undermine program learning outcomes.

Three Quality Assurance Methods and Their Analyses

This section describes three quality assurance methods taking place at Tilburg University and analyzes their strengths and weaknesses which are based on the design dimensions of quality assurance in assessment. Strengths enhance the likelihood of achieving quality assurance objectives, while weaknesses hinder the effectiveness of the quality assurance system. The analysis is based on the four design dimensions of quality assurance of assessment described in the previous section.

Post-hoc Evaluation

Post-hoc evaluation focuses on courses completed in the previous educational unit (e.g., course or module). As shown in Figure 2.15.1, the process starts with selecting courses through random sampling, predefined criteria, or signals from program management. In the next step, relevant course documentation, including syllabi, assessments, grading instructions, and test analyses, is compiled into an assessment package. Assessment committee members then evaluate the package using a standardized form, approved by the Examination Board. Findings are synthesized into a report highlighting areas for improvement and providing actionable recommendations for course coordinators. The report informs the Examination Board and initiates the next PDCA cycle and is shared with stakeholders like the vice-dean of education, academic directors, and course coordinators.

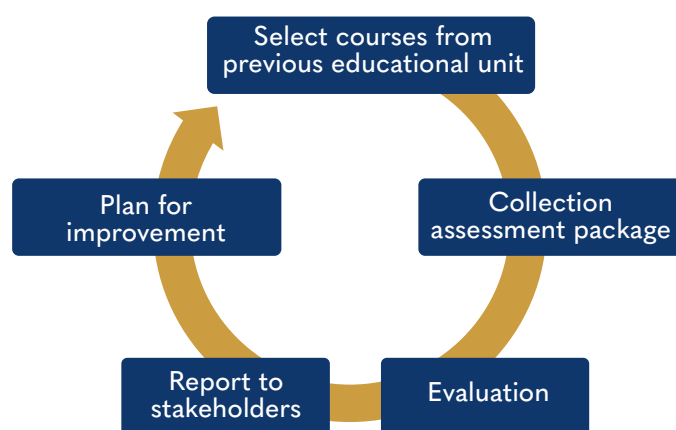


Figure 2.15.1

Post-Hoc Method to Evaluate Course Assessment Quality

Strengths and Weaknesses

Post-hoc evaluation allows quality assurance to be implemented without immediate pressure and can support lecturers in improving future courses. However, it focuses on the Check phase of the PDCA cycle, which is about reviewing documents and providing delayed feedback. It limits both its effectiveness and contribution to quality enhancement. Assessment quality is evaluated in isolation, without consideration of instructional activities or program-level assignments, restricting holistic improvement. The role of education specialists is limited, reducing interdisciplinary input while a substantial portion of committee effort is devoted to reporting than developing examiners' competencies. Consequently, the development of lecturer assessment competencies occurs in subsequent cycles, whereas resources of assessment committee members and specialists remain allocated to past-course evaluations. In addition, evaluating course assessments without considering their role in the overall program limits the translation of course-level insights into program-level improvements. As a result, this method cannot fully ensure that students achieve the intended final qualifications, and therefore the Examination Board cannot fulfill its statutory duty.

Concurrent Evaluation

The concurrent evaluation method assesses courses during the semester in which they are delivered, shifting the focus from compliance alone to proactive, continuous support. It employs a three-part feedback system: feed-up, feedback, and feed-forward. Feed-up involves a needs analysis conducted prior to the semester, feedback provided ongoing guidance to improve assessments and exam questions, and feed-forward includes group discussions at the end of the semester to address common issues and enhance program coherence.

As illustrated in Figure 2.15.2, two to three weeks before the semester starts, a program is selected for evaluation, encompassing all scheduled courses for the upcoming educational unit. The chair of the assessment committee informs course coordinators about the upcoming evaluation, after which an assessment specialist meets with each course coordinator or teaching team to explain the evaluation's purpose, procedures, and criteria, and to conduct a needs analysis. This initial interaction allows lecturers to clarify policies, rules, and guidelines.

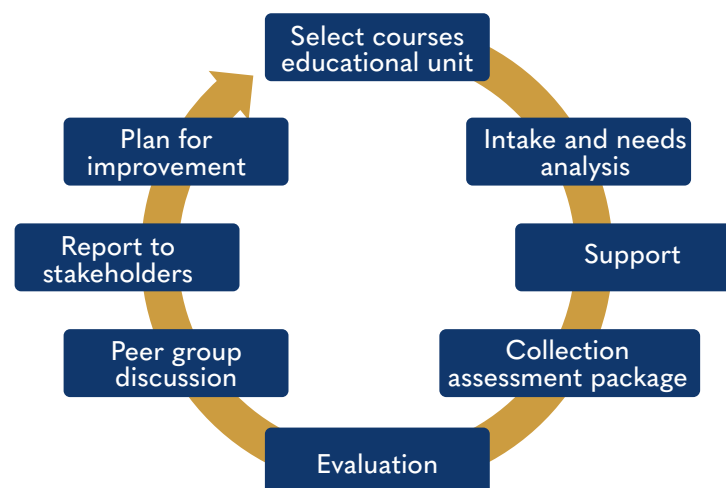


Figure 2.15.2

Concurrent Method to Evaluate Course Assessment Quality

Throughout the semester, the assessment specialist provides ongoing support tailored to identified needs, and lecturers can request guidance as necessary. Periodic meetings with the assessment specialist and the assessment committee ensure progress is monitored, and emerging issues are addressed.

At the end of the semester, assessment documents are collected and reviewed using an approved standardized form. The assurance committee identifies common themes across courses within the same program, which are discussed during a group discussion with course coordinators and committee members. Meeting minutes and individual course assessment forms are shared with the Examination Board and the vice-dean of education.

Strengths and Weaknesses

This concurrent evaluation enhances transparency, supports ongoing improvement, and fosters a collaborative environment. By involving stakeholders (i.e., course coordinators and program academic director) throughout the educational unit, it promotes both care and quality enhancement. However, the process remains heavily focused on assessment quality and relies primarily on assessment specialists. Not involving education specialists limits pedagogical insights, reducing the integration of instructional activities with assessment practices. While courses within the same program can be evaluated collectively, the lack of comprehensive assessment program design (e.g., how to assess process and attitudinal aspects) confines improvements at the course level, and recurring coordination issues, such as assessment scheduling and sequencing, persist across cycles.

Evaluation of Future Courses: The SMART Cycle

To enhance strengths and address weaknesses in current educational quality assurance methods, we have introduced a SMART cycle, consisting of five steps: Strategic sequencing, Meetings, Asking and acting, Review and reflecting, and Thematic adaptations (Figure 2.15.3). It can also be regarded as a course readiness review. Unlike approaches that focus solely on assessment, the SMART cycle emphasizes overall educational quality and program-level evaluation and therefore perfectly fits with the principle of structural alignment.

The cycle shifts responsibility for quality assurance to the program, promoting collaboration among stakeholders. It fosters a reflective, self-monitoring community focused on curriculum and assessment design. A curriculum team, led by the academic director and including members from the assessment committee and education support team (i.e., assessment and educational specialists), coordinates the process. Assessment committee members support monitoring and advising assessment quality, ensuring constructive alignment at both course and program levels.

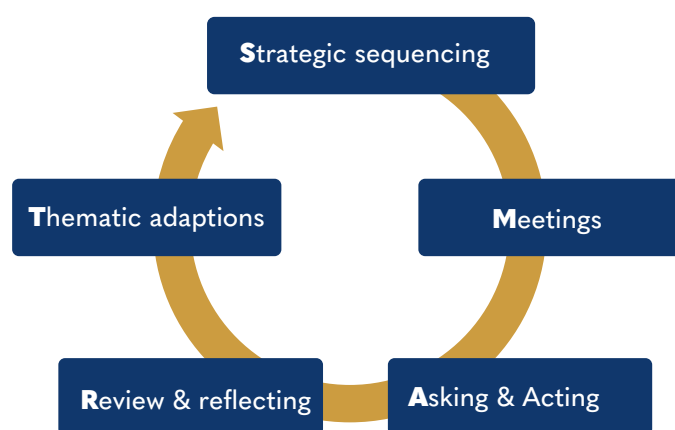


Figure 2.15.3

SMART Cycle of Quality Assurance

Given the resource constraints, this setup prioritizes the use of existing data, thus minimizing the need for labor-intensive collection of course documents. Courses within the same educational unit are evaluated together to optimize efficiency. Strategic sequencing begins before and during the semester, enabling course coordinators to make timely adjustments. By synthesizing program assessment plans, previous course evaluations, and administrative data, the curriculum team obtains a comprehensive overview of learning goals, assessment types, weighting, and

passing rates. These insights inform curriculum design, including sequencing learning trajectories, balancing assessment workload, and designing **formative assessments** that encourage deep learning.

The curriculum team meetings are organized to focus on clusters of courses, with the academic director clarifying program assessment policies (e.g., 100% exam, formative feedback, a maximum of two assessments per 6 European Credit Transfer and Accumulation System (ECTS) course, and one assessment for 3 ECTS courses). Course coordinators discuss learning goals, instructional strategies, and assessment design with education and assessment specialists. These discussions are formalized in the publicly available course catalog, a course assessment overview, curriculum map and program assessment plan, ensuring alignment with program and institutional quality criteria of assessment.

The asking and acting step enables coordinators to request support on demand, with the education support team providing tailored guidance and timely feedback during the course. Review and reflection involve ongoing monitoring by specialists, identifying issues such as unclear **rubrics** or ineffective formative assessments. Findings are systematically recorded and used by the curriculum team to collaborate with coordinators, reflecting on improvements for both exams and coursework.

Finally, thematic adaptations group review findings into prioritized themes (e.g., assessment workload, setting minimum grades) to inform the next cycle, ensuring continuous improvement.

Strengths and Weaknesses

The SMART cycle effectively integrates quality assurance and enhancement, strengthens examiners' competencies through structured feedback, and promotes program-level coherence. Coordination among stakeholders clarifies responsibilities and fosters an educational community. Its main limitation is the need for professional development and a learning-by-doing approach, requiring dedicated time for reflection and capacity-building among stakeholders. Nevertheless, the cycle offers a sustainable, program-focused community-building approach for continuous improvement in higher education.

Concluding Remarks

This chapter examines the three evaluation methods for course-level assessment quality, focusing on design dimensions for quality assurance and quality enhancement. Transitioning from post-hoc and concurrent methods to the SMART cycle necessitates strong support from the management team and Examination Board. In particular, the SMART cycle addresses the central questions regarding long-term support for quality enhancement. Across these methods, our experiences reflect a learning community committed to the continuous improvement of both quality assurance and quality enhancement practices. This development has been guided by reflective observation and a deliberate shift from focusing solely on compliance and external accreditation toward prioritizing educational quality, assessment practices, and the well-being of academic and support staff as well as students (i.e., a perspective from care and relational pedagogy). Implementing such a system is initially challenging, particularly when assessment expertise is limited. Over time, as the competencies of stakeholders within the assessment organization have developed, greater confidence has been achieved in designing a quality assurance system that effectively integrates enhancement alongside compliance.

Our experience highlights the importance of structured cycles, timely feedback, and collaborative engagement among program directors, assessment specialists, and education support teams. These practices ensure that course-level assessments align with program-level goals, promote constructive alignment, and support the professional development of examiners. We hope that readers can draw practical insights from both our current and past practices and consider how a care-based, integrated approach to assessment quality and enhancement might be adapted to their own institutional contexts. Ultimately, our experience suggests that investing in structured, collaborative, and reflective quality assurance processes can strengthen both educational outcomes and the professional growth of academic staff.

SECTION 3

DESIGN CASES

A Note on Section 3

Introduction

Section 2 presented design principles and practical suggestions for ensuring key assessment quality criteria, including **transparency**, **reliability**, and **validity**. Building on this foundation, Section 3 illustrates how both conventional and alternative forms of **assessment** can be developed and adapted to address these challenges.

As assessment always takes place within a specific institutional context, its design must take into account external accreditation frameworks, the institution's educational vision, and the policies and priorities of institutions, and programs. At the same time, educational institutions continuously develop and adapt their assessment and feedback practices in response to ongoing developments and the increasing complexity of societal and technological change. Section 3 is therefore situated within this dynamic interplay between institutional frameworks and evolving educational practices.

In Section 3, the chapters are written by academic teaching staff in collaboration with educational and assessment specialists. Each chapter follows a similar structure, beginning with a brief introduction to the course, including its context, overall design, learning activities, and assessment approach. While each chapter highlights a particular assessment type, it also explains how this assessment relates to others within the course, enabling readers to develop a clear understanding of the overall course-level assessment design.

Across these chapters, the authors share and justify the reasoning behind their assessment decisions. Throughout this book project, we observe how they gradually become more familiar with theoretical perspectives from the field of assessment and deepen their understanding of validity, reliability, and transparency in their own educational contexts. Misconceptions are identified, discussed, and clarified. Most importantly, the authors critically reflect on their own practices and, through dialogue and written feedback, create the necessary distance to examine and improve their work. As several contributors note, this project functions as an alternative form of professional development for their assessment expertise, and the insights gained from this reflective process can be carried forward into future improvements.

From the perspective of teaching professionals, we believe that through writing and constructive, critical dialogue, lecturers can further develop their assessment expertise, which is an additional and important benefit of this book project. At the same time, these reflective practices also provide us, as assessment specialists, with the opportunity to examine our own disciplinary assumptions from cross-disciplinary perspectives. This reciprocal process deepens our understanding of assessment as an integral part of the educational process and reinforces the need for a more holistic and integrated approach that balances measurement, assessment, and evaluation.

A well-designed educational system must therefore maintain a balance among these three domains rather than allowing any single one to dominate. A balanced assessment culture therefore treats measurement as a tool, not a master; assessment as interpretation, not mere scoring; and evaluation as judgment, not mere calculation. They should be understood as mutually constraining rather than independent.

A system that pursues transparency through excessive formalization (e.g., example, through overly detailed **rubrics**) may distort learning by encouraging performative compliance (Bloxham et al., 2016; Sadler, 2009), where students adopt a “tell me what to do and I will do it” attitude (Balloo et al., 2018), rather than developing genuine understanding. In such cases, transparency becomes a technical display rather than a support for meaningful learning. Returning to the perspective introduced in this book, from a Taoist point of view, transparency in educational

assessment is not merely a matter of making criteria, procedures, or results visible. Rather, it concerns *not* obscuring the natural process of learning through excessive control, technical complexity, or artificial standards. In Taoism, true transparency means allowing reality to reveal itself as it is. Applied to assessment, this suggests that assessment practices should allow students' learning to emerge in its own forms and rhythms, rather than forcing it into rigid, pre-designed measurement frameworks (Brookhart, 2003).

A system that maximizes reliability through rigid standardization may undermine validity by narrowing what can meaningfully be assessed; as a result, essential **learning outcomes** may remain invisible, particularly complex skills (e.g., oral communication and creative thinking) and outcomes in the affective domain (e.g., character strengths) (VanderWeele, 2022).

Interrater reliability is therefore essential, but it should not be understood as a requirement that all examiners assign exactly the same score (Sadler, 2005). Rather, it concerns whether examiners' judgments are educationally and professionally defensible, grounded in shared criteria and standards, and coherent in their reasoning. The goal is not mechanical agreement, but *interpretive alignment*: different examiners may legitimately emphasize different aspects of performance when forming their evaluative judgments, yet their conclusions should be comparable in meaning and level. In this sense, reliability in complex assessment is not about eliminating professional judgment, but about making such judgment trustworthy through calibration, dialogue, and a shared understanding of quality (Sadler, 2013).

Finally, in terms of validity, the contributors to Section 3 use their course assessment designs to demonstrate the principle that “we assess what we value,” rather than “we value what we assess.” This is reflected in their emphasis on formative activities with supervision on process monitoring and feedback, which supports the development of students' knowledge, skills, and character, while also enhancing their understanding of course learning goals, assessment criteria, and standards. In this way, validity is more meaningfully achieved.

Our Assessment Vision

To better understand how chapters of Section 3 are formed, it is essential to know our assessment vision at Tilburg University.

Most of us in academia have extensive experience with assessment and have developed so-called test-wise strategies to survive and succeed in a competitive academic environment. Reflecting on our own assessment experiences, we may ask whether we would want our young generations of students to be shaped by the same assessment practices. More importantly, we should question whether these past practices are sufficiently resilient to sustain students' motivation for lifelong learning, support their ability to function independently in personal and professional contexts, and can care for themselves, others, and the world around them (Bologna Working Group on Qualifications Frameworks, 2005; NLQF, 2025).

These concerns form the starting point for our assessment. Any assessment shares responsibility for realizing an institution's educational vision and therefore must be aligned with the institution-wide educational vision and didactic framework. At Tilburg University, this vision is grounded in a commitment to serving society through education and research based on truth-seeking, solidarity, empathy, and responsibility. Inspired by the tradition of Martinus Cobbenhagen, the University embraces the decompartmentalization of science and academia by connecting disciplines, perspectives, and practice. Its goal is to educate thinkers of character with an international outlook and a strong focus on the future. This mission is pursued through slow and reflective thinking, professional and dedicated lecturers, responsible and engaged students, supportive alumni, and a strong emphasis on the human dimension of education. In doing so, Tilburg University consciously integrates knowledge, skills, and character as the core pillars of its educational mission (Tilburg University, 2017a).

Constructive alignment is the most widely adopted didactic framework in Western higher education (Biggs & Tang, 2011) and is also prevalent in Dutch higher education. Figure 3.0 illustrates its application and the translation of Tilburg University's educational vision into practice.

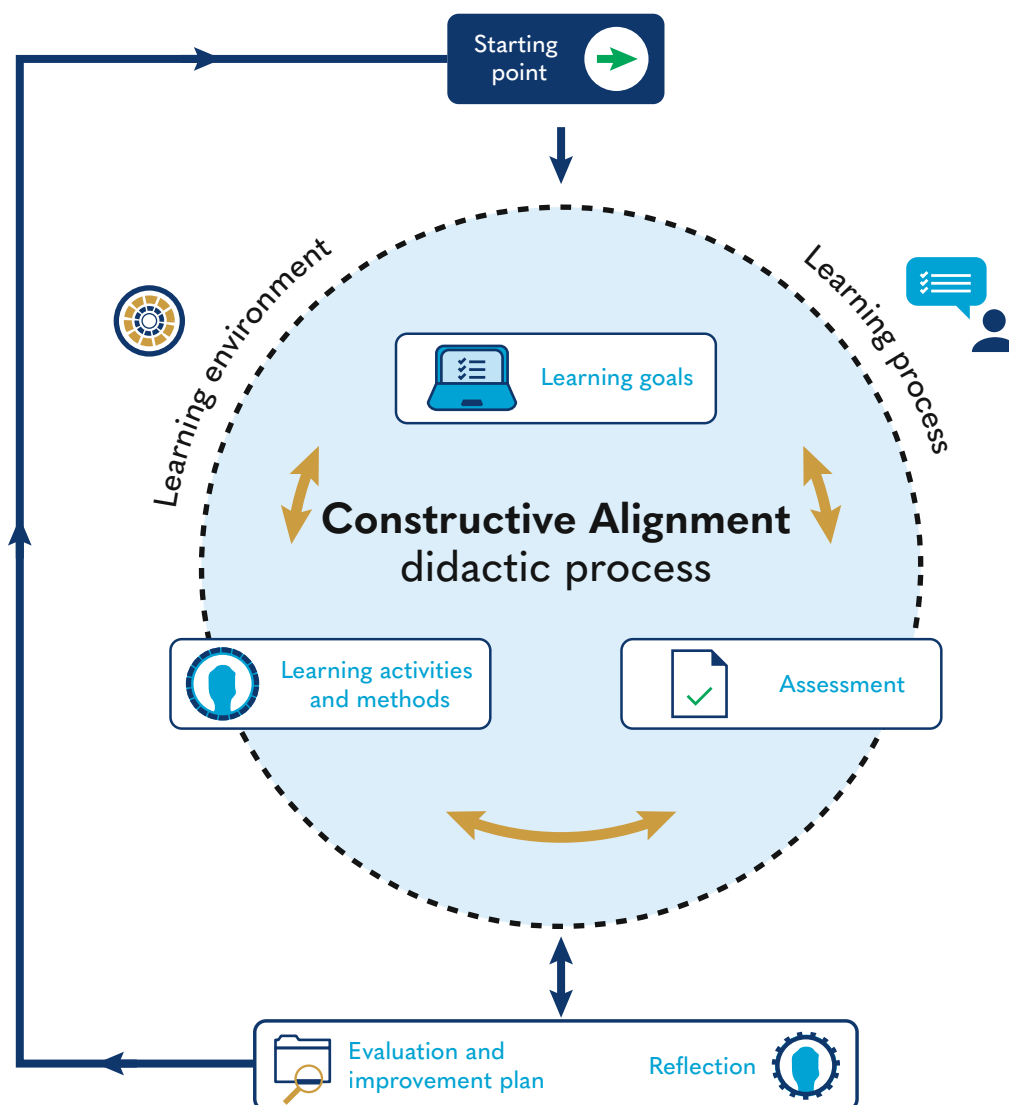


Figure 3.0
The Didactic Framework at Tilburg University (Teaching and Learning Center, 2025)

Our assessment vision is grounded in the principle that assessment should be embedded within the constructive alignment framework. Effective instructional design requires both alignment and constructiveness, but these serve distinct functions. Alignment provides structural coherence by ensuring consistency among learning goals, learning activities and methods, and assessment. In contrast, constructiveness emphasizes the active role of students in learning and assessment, how learners make sense of knowledge, and transform explicit and tacit understanding into skills and competences through experiential and reflective processes. Impacted by GenAI and a rapidly evolving global landscape, assessment must therefore do more than measure outcomes; it must motivate students, deepen engagement, and remain responsive to the changing demands of the educational ecosystem.

Within this framework, alignment functions as a cornerstone of assessment validity and plays a central role in program-level curriculum mapping and assessment planning (Jankowski & Slotnick, 2015). It offers a concrete structure for organizing learning goals, learning activities and methods, and assessment tasks across courses and

programs. However, while alignment is often carefully designed, the constructive dimension of assessment is frequently underdeveloped in practice. Constructiveness is commonly treated as a *conceptual* principle (i.e., consistency among the three educational processes) rather than being operationalized in assessment design at the program and course levels. As a result, opportunities to actively engage students in integrating explicit and tacit knowledge through interaction with lecturers, peers, and real-world contexts are often missed. This leads to the first proposition for quality assessment: alignment and constructiveness must be intentionally and operationally integrated to inform curriculum development and foster deeper student engagement at both the program and course levels.

Traditionally, higher education has prioritized to transfer explicit, codified knowledge (Eraut, 2000), typically articulated in written documents, while tacit forms of understanding, developed through socialization and contextual experiences, have received less attention. Assessment practices have emphasized external assessment and feedback, often neglecting the cultivation of students' own evaluative capacities. Moreover, assessments tend to focus on end-products rather than capturing learning as it unfolds over time through meaningful milestones. These prevailing approaches have led educators to focus primarily on aligning goals or outcomes with assessment of explicit knowledge, rather than fostering student engagement through self-regulatory practices such as self-assessment, self-feedback, and the development of evaluative judgment. Given the growing need for lifelong learning and adaptive expertise, integrating self-regulatory didactics and assessment into the curriculum is urgent. This leads to the second proposition: assessment must develop students' capacity to monitor and guide their own learning in order to function effectively in complex, dynamic environments beyond formal education. This emphasizes the evaluative function of assessment for multiple stakeholders, including educators and students, across different levels of the assessment sub-system, such as the faculty, program, and course levels, within the broader educational ecosystem. In addition, promoting student agency in line with student-centered approaches and co-assessment has the potential to bridge the divide between the formative and summative functions of assessment.

We seek to highlight the increasing importance of tacit knowledge, particularly in the context of criterion knowledge, which is the ability to make informed evaluative judgments (Bearman et al., 2024). Tacit knowledge closely relates to *knowing how*, defined as a "student is ready to detect and correct lapses in his operations, to repeat and improve upon successes, to profit from the examples of others and so forth" (Ryle & Tanney, 2009, p.17). This ability is essential for both educators and students, especially as AI technologies become increasingly integrated into education. Accordingly, student accountability, defined as students taking responsibility for their own learning, actions, and performance, is more important than ever to maintain academic integrity and the credibility of qualifications. As a key component of authentic assessment defined as building "authentic knowledge" from engaging in real-world or workplace relevant tasks (McArthur, 2023), tacit knowledge requires not only active participation from students but also purposeful facilitation by educators. The relationship between explicit and tacit knowledge significantly influences assessment practices, particularly in task design and criteria. Some argue that focus on explicit knowledge could eventually encapsulate tacit knowledge, leading to overly detailed tasks and rubrics that fail to capture quality (Sadler, 2007). However, as explicit knowledge grows, so does tacit knowledge, challenging outcomes-based education assumptions. Holistic assessment, where quality is evaluated interactively, blending explicit elements with the nuanced, non-codified aspects of tacit knowledge (Tomas et al., 2019) is therefore a promising approach. This leads to the third proposition: assessment design should motivate students to integrate explicit and tacit knowledge throughout the learning process through **formative assessment**, thereby ensuring that the three core assessment quality criteria of validity, reliability, and transparency are achieved in **summative assessment**. Emphasizing the assessment of integrated knowledge types is essential for addressing both the cognitive and affective domains, including attitudes and character strengths. At the same time, this holistic and integrative approach helps avoid the pitfalls of assessing learning outcomes in isolation, over-assessment, and unnecessarily complex assessment designs.

Vision Statements

Together, these three propositions lead to the conception of *assessment as learning*: assessment is truly meaningful only when it actively promotes and enhances learning. Rather than serving merely as a tool for measurement or certification, assessment should function as a catalyst for growth, supporting students in developing the capacities required for lifelong learning.

Contributors of Section 3

Based on these vision statements, we invited contributors whose perspectives align with ours, drawing on our ongoing collaborations. In addressing multiple, concurrent challenges, these contributors demonstrate a strong commitment to maximizing student learning through assessment and to gradually advancing towards assessment as learning.

The work presented in Section 3 should not be seen as final; rather, it represents work in progress. We encourage readers to view these chapters as evolving contributions that will be continuously refined and updated through our interactive book platform.

3.1

Formative Testing in an Interdisciplinary Minor Course

Pytrik Schafraad and Frea Dooremaal

Motivation

Currently, the course Communicating Sustainability⁵⁰, as a part in an interdisciplinary minor focusing on sustainability, is running in its third year. From the beginning, considering the interdisciplinarity of the student population has been a didactic challenge requiring creativity and flexibility. This interdisciplinarity involves not only a wider range of interests and foci of the students, but also a greater diversity of skills and knowledge. We quickly learned, for example, that one cannot expect such a diverse group of third- year bachelor students to have the same basic methodological skills as one can expect from first year Communication & Information Sciences students after the first semester. Similarly, their motivations for taking this minor go from general interest, to integrating sustainability knowledge into their main studies, or aspiring to work in the field of sustainability after graduation.

These observations led us to start searching for ways to better connect their experiences, knowledge and skills with the learning goals of the course. One of these ways we have been trying out since, is formative testing.

Working from what we know

Formative assessment is a widely researched phenomenon in all types of education and ranges from how to implement an ungraded assignment into a course, to building relationships through dialogic feedback. For this course, we interpreted the concept of “formative” as assignments that are ungraded and contain some form of reflection. This reflection can be on how different elements of the course relate to each other and to students’ disciplinary experiences, reflection through critique from others, or reflection through some form of interaction.

For the design of the formative elements of the course, we took several things into consideration:

- **Social learning:** Learning is a social process (Bandura, 2006). But for this course, we looked at more than learning through observation. Students also take on an active role in their interaction with the course content and each other. Students are asked to annotate course reading to deepen their understanding of core concepts (Marissa, 2021). Additionally, students participate in a peer feedback assignment, which not only provides them with feedback, but also gives them the opportunity to learn by giving feedback (Nicol, 2011).

⁵⁰ For this course the course coordinator partnered with an Instructional Designer connected to his program. The design of the formative assessments as well as this chapter are products of that collaboration.

- **Study crafting.** Study crafting allows students to bring in their own needs, skills, and preferences and align their education with their strengths, interests, and goals (Duchi et al., 2025). An element that might come into play in study crafting is 'agentic engagement', defined by Reeve and Tseng (2011) as the fourth aspect of student engagement.
- **Feedback cycle.** Instead of one directional feedback (usually from a lecturer), we have attempted to introduce a feedback cycle in the course. A feedback cycle (feed-up, feed-back, and feed-forward) can help feedback to be adopted by a receiver and make improvements in their work (Hattie & Timperley, 2007).

The Course: Communicating Sustainability

As noted above, the course is part of the minor 'Sustainability'⁵¹. Sustainability is a much-debated issue in contemporary society. Many different types of actors communicate about sustainability, for as many reasons and as many goals: information sharing, influencing behavior, countering disinformation, solving sustainability issues, decision making, etc.

We investigate a number of communicative challenges that societal actors that are dealing with sustainability issues might run into. The challenges of lack of understanding and urgency amongst a part of the public, the challenge of motivating people and organizations to change behavior, the challenge of mis- and disinformation about sustainability issues and the challenge of contributing to a fruitful (public) debate on sustainability issues.

In sum, students investigate several communicative challenges around the issue of sustainability and learn about theories that help understand these challenges and provide guidelines for effective communication strategies.

The learning goals are:

By the end of this course, students will be able to:

- Define and apply core concepts and describe current theories in sustainability communication from various perspectives, with a focus on communication science theories.
- Critically analyze sustainability communication campaigns and public debates about sustainability in news- and/or social media.
- Apply theoretical knowledge obtained in the course to debate standpoints and choices concerning communication about sustainability.

Students from all schools, including exchange students (from almost a dozen different nationalities) take the course (and mostly the entire minor program). As mentioned before, some students expressed general interest in the topic of (communicating) sustainability as a motivation to take the course, while others have expressed more specific motivations: going from developing skills to successfully campaign for more sustainable consumption on social media, to the desire to work in the field of sustainability communication, or management.

Why Formative Assessment?

A first reason to include formative assessment is to be able to cater to the diversity of students' interests and motivations to take this minor. One way of doing that is creating space to include their own issues, media experiences, and motivations in their learning process (Study crafting). Therefore, the first formative assignment in the course asks the students to formulate their own personal learning goals, which they may focus on a specific topic, issue, aspect of communication (production, content, or consequences of communication messages) or into a specific field of study (public relations, journalism, science communication, or mobilization, just to name a few).

51 Minor in Sustainability: An Interdisciplinary Approach | Tilburg University

A second reason to include formative assessment is to be able to provide feedback on basic study and analysis skills. In the first two runs of this course, the expected level of these skills turned out not to be met by everyone, but at the same time, the purpose of the course does not allow intensive skills learning due to limited lecture hours and limits in teaching capacity. Hence, social learning activities where students collaborate and provide each other with annotations in the provided literature, with only modest feedback from lecturers, turn out to be a good way to include effective practicing of these disciplinary skills in the course (Nicol, 2011).

A third reason is to stimulate active participation throughout the duration of this semester long course. Many people working in academic teaching may recognize that a significant group of students tends to start studying actively at the last minute. Such observations let to design activities that stimulated students to regularly work on their assignments in exchange for feedback that would help them with future steps in the assignment as well, making the assignments learning experiences, instead of simply **assessments**. We therefore integrated formative assessment in such a way, that it requires frequent and meaningful contributions of students.

Lastly, an important side effect of replacing graded/summative assignments with formative assignments is that the number of hard deadlines and criteria for assignments is reduced, and students' stress level is reduced. The focus switches from **learning outcomes** and assessment to learning process and feedback. This is in line with the university's goals stipulated in the "Smarter Academic Year" plans (De Jonge Akademie, 2021).

Teaching and Learning Activities

The four goals that we want to achieve with formative assessment are translated into three formative assignments that are supported by small planned and unplanned classroom activities. Below is a schematic overview of the three formative assignments.

Block 1	Block 2	Block 3	Block 4
Students formulate two personal goals for the rest of the course Close reading assignment: students annotate articles to better understand literature in communication sciences.	Students connect course reading materials to the guest lectures.	Group campaign assignment based on course literature.	Individual media analysis assignment based on course literature to uncover communication dynamics.
Students bring an example of sustainability communication to class and analyze together. Every lecture starts with three students explaining their example.			
Students keep a log on how the course activities are affecting their personal learning goals.		Final reflection on personal learning goals.	

Figure 3.1.1

Visual Overview of Course Assignments: Summative Assessments With Red Borders. Bottom Assignments Reoccur Every Block

Block 1 focuses on letting students orient themselves within the course, get familiar with the topic of communication in sustainability and how it is presented in literature through the close reading assignment (Hattie & Timperley, 2007; Marissa, 2021), and plan for their learning through personal learning goals (PLG) (Duchi et al., 2025). Throughout the entire course, students are asked to connect their learning goals to the course material and activities (Hattie & Timperley, 2007).

In block 2, students are asked to link the guest lecture on active citizenship to the course material. This is to prevent the guest lecture from becoming a stand-alone act and help students connect the topic of communicating sustainability across disciplines and from theory to practice.

Both block 3 and block 4 contain a summative assessment, which students must complete in groups (campaign assignment) and individually (media analysis). Again, students are asked to link these assignments to their PLG, and to finally reflect on to what extent students reached their PLG.

Reflections and Experiences

Partly these reflections and experiences lay in the future of the current semester, as the majority of these formative assessments is new and still ongoing (at the moment of writing). Below a preview of reflections so far and those on an earlier version of one of the assignments:

Personal Learning Goals

At the moment of writing, students just received feedback on their PLG, but have not submitted any of their reflections yet, so the following is based on just the first of three tasks in the assignment.

Students formulated a wide diversity of PLG, some quite specific, some more general. Their PLG showed several common interests about strategies for effective sustainability communication and the problem of mis- and disinformation, such as these:

“I want to learn how to come up with and test communication strategies that tackle misinformation and open up constructive debate on sustainability.”

“Learning to develop skills for communication that counters misinformation.”

“By the end of this course I want to have a better understanding of how climate misinformation, especially from political figures and online platforms, impacts the public perception of climate change and sustainability.”

Apart from their common interests, they often differ a lot in emphasizing either knowledge and understanding, or specific communication and analytical skills. While the course is mostly focused on understanding and knowledge, these PLG support the idea that it is important to address the impact and practical implications of theories and research findings in reading material.

After students submitted their two PLG via Canvas, they were provided with short written feedback to them (Feedback circle). It was decided not to address formulations, nor critically evaluate their motivations, but to focus the feedback on how students may actually achieve their goals, by attending them to certain lectures, learning activities, or (compulsory or recommended) readings, in order to support their study crafting, like here:

“PLG1: Figure out what actually works online. My first goal is to learn the psychology behind why some sustainability posts on TikTok or Instagram make you want to change, while others are just super boring or feel fake. I want to understand what makes a message actually stick in someone’s brain and make them care.”

I’m online all the time, and I see so much stuff about “saving the planet.” But most of it feels like big companies just trying to sell me something with a green label on it (greenwashing, right?). I want to know the secret behind the posts that actually feel real and make me and my friends stop and think. Basically, I want to learn how to spot the difference between messages that genuinely work and the ones that are just noise.”

Feedback:

“Your goals are really ‘hands on’, which is a good addition to the general course goals, which are maybe more theoretically oriented. Nevertheless, we also pay attention to the effectiveness of communication strategies (i.e., in lectures 9 and 10) and of course we do that in more hands on in practice seminar 1.[...]”

And here:

“PLG2: Develop skills to identify and counter misinformation, disinformation, and greenwashing [...]”

Feedback:

“Misinformation and greenwashing are complex issues, which we will address, for example, in lecture 7 and lecture 10. Be aware that for most lectures recommended readings are available and might help you reach those learning goals as well.”

So far, this formative assignment of requesting students to formulate their own PLG, seems to be useful to encourage engagement with study materials and reflect on their own learning process (albeit their written reflections still lay in the future), but also provide lecturers with valuable insights into students’ interests and perspectives. And reviewing these at the beginning of the course provides ample opportunities for tailoring lectures to students’ interests in general.

Observations of Sustainability Communications

One of the formative assignments in the course entails that students are expected to collect an example of communication about sustainability, which they encountered in their daily lives, through their personal *media diets* (Gerbner, 1998), upload it to Canvas d shortly explain their pick in class (Social learning, Feedback circle). Some students really liked this assignment and submitted multiple examples throughout the semester, while others did not at all, or contributed only the required minimum. This assignment does several valuable things for the course.

First, it stimulates students to contribute to the shared experience in the course and to consciously observe their mediated environment (Reeve & Tseng, 2011). Through their own contributions as well as those of their peers, they develop ‘an eye for sustainability communications’. In other words, they learn to observe relevant occurrences and to reflect on them. As they learn in the course, messages are not just there, but they are produced, distributed, received, and interpreted. And with the weekly discussion of all of these observations, they practice their skills in observing as well as critical reflection, which also means they must connect theory with practice.

Second, it provides a diversity of useful communication examples to the course, which you as a lecturer cannot match on your own, simply because students draw their examples from their own diverse and widespread manifold of experiences. Observations go from stickers on bins providing recycling instructions to newspaper articles on a failed global plastic summit, to a clip of one of the Lord of the Rings movies⁵², or an advertorial for a wine brand that uses paper bottles. The Lord of the Rings observation was brought in during the second week of the course but could be used as an example when we discuss sustainability communication in the cultural domain in week 10, while the paper bottles may illustrate greenwashing mechanisms in week 12. Thus, each contributed observation may become a learning incentive in multiple ways.

Third (and last), the input of these examples brings about interaction and debate about the examples, their message or similar experiences between students on the spot, and some examples brought in by the students serve as valuable illustrations to empirical research or theories that are discussed later on in the course.

52 LOTR The Two Towers – Extended Edition – The Last March of the Ents

Concluding Remarks

Our experience with formative assignments has led to the following observations. First, giving students opportunities to improve their skills, adapt the course to their interests, and share their ideas with other group members encourages them to take a more active role in their learning. It also increases their motivation to participate. These are important, although implicit, learning goals of the course and support previous research showing the positive effects of study crafting (Duchi et al., 2025). Even though students show considerable differences in how they make use of these opportunities, presenting them with these formative assignments has a positive impact on students, the learning climate, and sense of belonging amongst the course population. The increased frequency of formative learning activities, the feedback and interactions that follow from it make for a social learning process (Bandura, 2006). Students are encouraged to frequently participate in various learning activities, with opportunities to incorporate their own interests (i.e., personal learning goals) and exchange their experiences with each other in an informal, or casual manner, that shows to be inclusive.

Although our experiences are positive in general, implementing these formative assessments and finding the right balance between voluntary and required participation in these learning activities remains a challenge. And although these formative assessments have contributed to bridging between interests, skills, and knowledge of the diverse student group on one side and learning goals of the course on the other, finding a good fit between the diversity of the student group and the course design remains a challenge that asks for constant maintenance, reflection, improvement and experimentation with new forms. Hence, it is a story to be continued.

3.2

From Formative Quizzes to iRAT–tRAT: Strengthening Analytical Reasoning in Philosophy of Law

Vincent Kalis and Bart van de Moosdijk

Context and Rationale

Assessment within the bachelor's program in Law has traditionally been predominantly summative in nature: a student's final evaluation for a course is generally based on a single exam. Philosophy of Law B, a core course that occupies a central position in the bachelor's curriculum, aligns with this approach and uses one written exam as the final assessment.

A substantial share of students struggle with this course because it requires a different way of thinking than they are used to in other law courses. Whereas most courses in the bachelor's program focus on understanding, interpreting, and applying statutory law, Philosophy of Law B demands a different cognitive approach: one that is not centered on applying fixed legal rules, but rather on analytical reasoning, such as questioning underlying assumptions and substantiating normative arguments.

The persistently low pass rates prompted the lecturer to consult an assessment specialist. During this conversation, the lecturer emphasized that any improvements should not result in lowering the academic standard: the course needed to maintain its intellectual ambition.

After implementing a number of adjustments, it became clear that the key did not lie in further modifying the exam itself, but in strengthening the learning process. Throughout the course, students began to practice this type of thinking more deliberately through weekly **formative assessment** questions, in which collaboration and the joint analysis, questioning, and refinement of legal-philosophical argumentation were central.

The introduction of these practice moments also paved the way for exploring the Readiness Assurance Test (RAT), an interactive assessment format from Team-Based Learning in which students collectively reason, question assumptions, and refine their arguments. In this approach, students first test their individual understanding of the material in the individual RAT (iRAT). They then discuss their answers with peers and, as a team, arrive at a shared answer during the team RAT (tRAT).

This chapter describes the experiences of the assessment specialist and the educational specialist who supported Hans Lindahl, Professor of Legal Philosophy in the Department of Public Law and Governance until February 2025, throughout the formative learning trajectory. It highlights the key outcomes, conclusions, and insights gained from the process. The chapter concludes with the lecturer's reflection on the trajectory, in the hope that these experiences will inspire colleagues who wish to build on them.

Course Context and Assessment Design

With more than 470 students per cohort each year, Philosophy of Law B has a substantial reach within the Law program. The exam is known for its high level of difficulty, and students generally perform modestly, particularly on the closed section, which initially consisted of 30 multiple-choice questions with four alternatives.

One possible explanation for these disappointing results is that the multiple-choice questions offer students little opportunity to demonstrate partially correct reasoning steps. As a result, students receive no credit for correct analytical intermediate steps, even when those steps indicate partial understanding. Essay questions would align more closely with this type of reasoning, because they allow students to make their argumentation explicit and to show how they arrive at an answer. However, given the size of the cohort and the fact that the lecturer is solely responsible for grading, an exam consisting entirely of essay questions would be practically unfeasible.

For this reason, the exam includes one essay question. This question is graded with partial credit and has a maximum length of 450 words. The final grade is calculated using a weighted formula, with the multiple-choice score corrected for guessing and accounting for 70%, and the essay question accounting for 30%. No separate minimum score is required for either section. Students pass the course if their weighted final grade is at least 5.5.

Exam Analysis and Starting Point for Improvement

To rule out the possibility that the low pass rates were caused by shortcomings in the exam itself, the lecturer had the exam analyzed by an assessment specialist. The specialist concluded that the questions were substantively correct and contained no construction flaws, with the sole exception that many questions included one **distractor** that did not function properly. Distractors that are rarely selected may be considered non-functioning; if few students choose them, they contribute little to a question's ability to distinguish between students who understand the material and those who do not; this may indicate that they are insufficiently plausible (Rodriguez, 2005).

Based on these findings, it was decided to reduce the number of alternatives to three. This change made it feasible for the lecturer to expand the number of multiple-choice questions by 10, bringing the total to 40.

Although the **reliability** of both the revised exam and the resit increased slightly, pass rates remained low in the subsequent administration. This raised the suspicion that the cause did not primarily lie in the exam, as only a few questions were still not functioning optimally according to the test and **item analysis**. In the follow-up discussion, the focus therefore shifted from the test content to the underlying learning process.

At an early stage, administering the tRAT in the lecture hall was considered. Earlier experience, however, had shown that this setting was not ideal for team discussion: the acoustics and room layout made it difficult for students to discuss effectively. Rather than abandoning the approach, the design was adjusted in consultation with the lecturer. E-voting was retained because the lecturer considered it valuable, while the iRAT-tRAT session was moved to the university's on-campus testing facilities, where the conditions for a quiet and well-organized session could be better ensured.

Principle of the Formative Approach

In developing the formative learning trajectory, the central question was how students could be supported more effectively throughout the course in acquiring the analytical and normative-argumentative skills required to meet the course learning goals of Philosophy of Law B:

1. Analyze why the doctrines of direct effect and the supremacy of EU law raise fundamental questions, and justify in what sense the European Union is a Union and in what sense it is European, based on identifiable characteristics.
2. Compare and evaluate the core positions in the debate on the EU's democratic deficit, using the German Federal Constitutional Court's Maastricht Judgment as the analytical point of departure.

3. Distinguish between standard and alternative theories of political representation and apply these theories to the debate on the constitutionalization of the Treaties establishing the EC, now the EU.
4. Explain and demonstrate how European judges engage in legal interpretation, and analyze how judicial law-finding functions as a form of political representation.
5. Analyze the normative tensions that arise between the EU and its Member States as a result of the EU legal order's claim to autonomy.
6. Examine the conceptual and normative implications for both the EU and its Member States in light of competing theories of political representation.

Because these course learning goals require a different type of thinking than students are accustomed to in other courses, an approach was needed that would guide them through this process step by step. The goal was to find a method that not only allows students to practice the content, but also actively encourages them to make their reasoning explicit, critically examine underlying assumptions, and construct arguments carefully and systematically.

The use of weekly formative practice questions was consistent with Dobson's (2008) finding that online formative quizzes can encourage students to prepare for class, provide timely feedback on their understanding, and support performance on summative exams. In Philosophy of Law B, this made the questions a logical starting point for a structured learning trajectory aimed at strengthening analytical and normative-argumentative reasoning.

The formative moments were not introduced as a series of separate interventions. Such an isolated approach would not sufficiently support the type of thinking students need to develop in Philosophy of Law B. For this reason, a structured method was implemented in which students discussed, during class, the two questions they had found most challenging among the five they had completed online. They then answered these questions through e-voting, followed by a plenary discussion led by the lecturer. In this way, a recurring rhythm would emerge: first complete the questions individually online, next answer the two most difficult questions together in the lecture hall, and finally deepen understanding during the plenary review.

Selection and Quality Assurance of Formative Practice Questions

In designing the formative cycle, one guiding principle was central: practice questions can only fulfill their function when they are both psychometrically sound and substantively relevant. For this reason, considerable attention was devoted to the quality of the questions used in the weekly cycle.

Philosophy of Law B had an important advantage in this regard. Because the lecturer developed new questions each year, a large and diverse question bank of 300 questions had been built up over time. This size made it possible to apply a stringent selection process, ensuring that the formative cycle would be both psychometrically sound and substantively relevant.

Based on the item analyses from previous administrations, an educational specialist selected only those questions that met two psychometric criteria:

- a **Rit** value of 0.35 or higher, ensuring that the question effectively distinguishes between students who have mastered the required analytical reasoning skills and students who still need to develop them.
- a **P-value** between the guessing probability and 0.90, ensuring that questions are neither too difficult nor too easy.

This preselection formed the basis for a substantive review by the lecturer, who evaluated each question for relevance and alignment with the core concepts of the course. After this combined selection and review process, a set of 75 high-quality and representative questions remained. Of these questions, 50 were linked to 10 formative practice moments spread throughout the spring 2020 course period, with the lecturer determining which question best aligned with each lecture.

Practical Implementation

To provide students with automatic weekly access to the practice questions, administration through assessment software was chosen. This allowed the questions to be released on the day before each lecture, making the implementation feasible in practice. It also enabled the collection of psychometric data for each administration, which helped identify the questions students had found most difficult.

With all preparations in place, the intended approach could be carefully embedded within the course. During the first lecture, the lecturer explicitly emphasized the importance of this method to the students. By practicing consistently each week, they would gain timely insight into their strengths and weaknesses and gradually build a solid preparation for the exam. The answers to the practice questions were made available in the assessment software after the corresponding lecture. Students were also given the option to provide feedback on individual questions, allowing unclear wording, ambiguities, or perceived errors to be identified and, where necessary, addressed.

Student Feedback on Weekly Practice Questions

The end-of-course student evaluations indicate that the weekly practice questions were experienced as clearly supportive. Students repeatedly emphasized that they appreciated being able to practice each week and that the formative quizzes, as one student aptly put it, “*helped me learn to think in the right way and understand exam questions better.*”

Extending the Formative Trajectory with iRAT–tRAT

From Weekly Formative Quizzes to an iRAT–tRAT Session

The plan to incorporate an iRAT–tRAT session had been in place since 2020, but it could not be implemented due to COVID-19 restrictions. When the pandemic measures were lifted in 2022, the opportunity finally arose to introduce this long-anticipated interactive session at the end of the course. The iRAT–tRAT format offered a natural extension of the course’s focus on analytical reasoning, allowing students to test their individual understanding first and then deepen their thinking through collaborative discussion and team-based justification.

This structure encourages students to make their reasoning explicit, compare interpretations, and refine their arguments. Held two weeks before the exam, the session offered a meaningful opportunity to deepen their understanding and practice the type of analysis that is central to the course.

Initially, administering the tRAT in the lecture hall was considered, earlier experience had shown that this setting was not ideal for team discussion. The acoustics and room layout made it difficult for students to discuss effectively. The lecturer wanted to retain e-voting, as he considered it a valuable part of the approach. The decision was made to administer the iRAT–tRAT session in the university’s on-campus testing facilities, where the conditions for a quiet and well-organized session could be better ensured.

Participation was voluntary and had no impact on students’ exam grades. Students could register through the learning management system, which allowed the assessment organization to prepare the session accurately.

Pedagogical Rationale for the iRAT–tRAT Approach

The choice for the iRAT–tRAT was not a standalone intervention, but a logical extension of the formative learning trajectory. By combining individual and collaborative answering, this format provides deep insight into one’s own understanding as well as that of others. Students discover where their reasoning is already strong, where gaps remain, and which alternative lines of thought their peers apply. The iRAT phase requires students to conduct an analysis independently, while the tRAT phase confronts them with different interpretations and arguments. This combination strengthens metacognitive awareness and actively supports students in sharpening the legal-philosophical reasoning skills that are assessed in the exam. As a result, the practice exam becomes not merely an assessment moment, but a deepening learning experience that closely aligns with the cognitive goals of the course.

Structure of the iRAT-tRAT Session

The iRAT-tRAT session was administered digitally in the assessment software. This made it possible to conduct the session reliably, collect responses efficiently, and analyze the results afterward.

A set of 15 carefully selected multiple-choice questions was compiled for the session. Each question had three answer options and one correct answer, and each correctly answered question earned 1 point. The lecturer would have preferred to include an essay question, but the assessment software allows only multiple-choice questions during an iRAT-tRAT session.

Because students completed the same set of questions twice, once individually and once in teams, increasing the number of questions was not feasible within the 120-minute session. The first 10 minutes were reserved for seating and logging into the assessment software. Figure 3.2.1 shows the subsequent timeline of the session, which resulted in 65 minutes of net testing time.

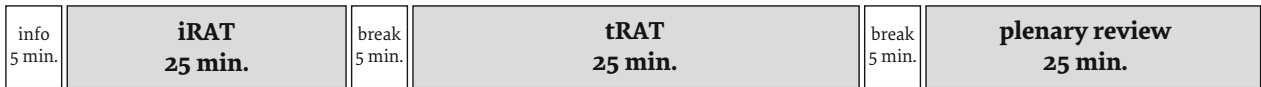


Figure 3.2.1
Timeline of the iRAT-tRAT Session

iRAT: Individual Phase

After a brief introduction, students were given 25 minutes to answer the 15 multiple-choice questions individually in the assessment software. The questions were presented in a fixed order to all students, and during the iRAT students were allowed to navigate freely between questions.

Seating and Transition to the Team Discussion

The iRAT-tRAT session was organized in two testing halls for digital examinations, with tables arranged in groups of four. Figure 3.2.2 shows the seating arrangement used for the iRAT session. The members of each team sat with their backs to one another, allowing them to answer the questions individually while already being seated in their assigned teams. For the tRAT phase, students turned toward each other so that the team discussion could begin immediately.

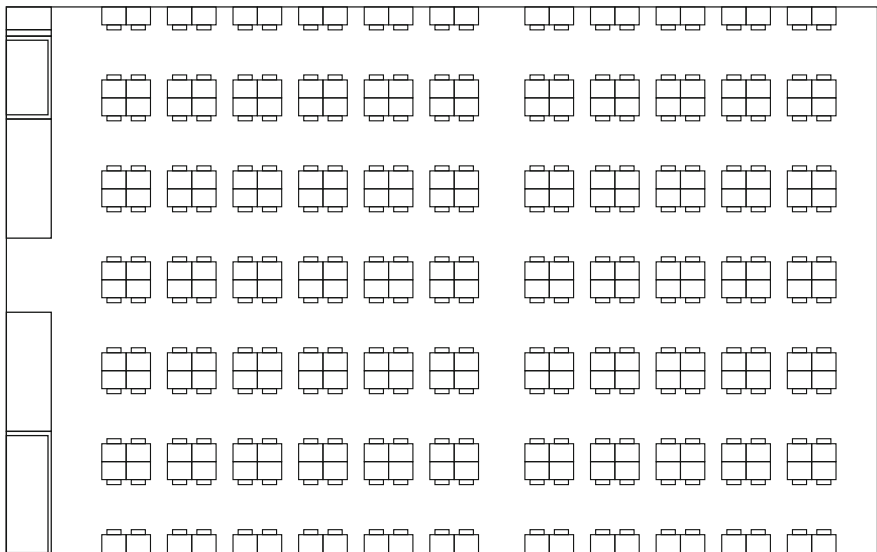


Figure 3.2.2
Seating Arrangement During the iRAT Session

tRAT: Team Phase

In the tRAT phase, the teams worked through the same 15 questions again and submitted one shared set of answers. The 40 minutes available provided ample time to compare lines of reasoning, critically examine assumptions, and weigh arguments collectively. Because the content of the questions was already familiar from the iRAT, the team phase could be used entirely to make reasoning explicit and refine it. The lecturer and support staff observed this process and saw how teams arrived at a collective answer through discussion.

Assessment-related Characteristics of the tRAT in the Assessment Software

The tRAT in the assessment software uses a linear format: a team can proceed to the next question only after correctly answering the current one, and answer options that were chosen incorrectly cannot be selected again. This structure encourages careful discussion before confirming a choice. After the session, students can see their total score, how many questions were answered correctly on the first attempt, and the score assigned to each question based on the number of attempts.

Plenary Review: Collective Deepening

In this phase, the lecturer discussed the questions based on the overall iRAT results, starting with the question that had proved most challenging for the group as a whole. This order was chosen because students were likely to be most interested in the questions they had struggled with, while their attention was expected to decrease as the plenary review progressed. The iRAT results, rather than the tRAT results, were used for this ordering because this allowed the educational specialist sufficient time to review the test and **item analysis** and prepare the slides for the lecturer.

The emphasis was on the reasoning steps that led to the correct answer. The correct answer itself was not the focus; rather, the underlying line of reasoning and the choices within the answer options were highlighted. Common reasoning errors and misinterpretations were explained systematically.

Student Feedback on the iRAT-tRAT Practice Exam

The student feedback shows that the iRAT-tRAT practice exam was experienced as particularly valuable and supportive of learning. Students emphasized that the session not only helped them understand the material more clearly but also gave them a strong sense of preparedness. The collaborative discussion of questions was repeatedly mentioned as effective, as it helped clarify reasoning steps and reduce uncertainty. The additional effort to organize an in-person practice session was also highly appreciated; students noted that this extra opportunity increased their confidence going into the exam. Overall, the responses suggest that the iRAT-tRAT practice exam played a meaningful role in both deepening subject understanding and strengthening exam preparation.

Lessons Learned

The iRAT-tRAT sessions provided several useful insights into the organization, technical setup, and didactic value of this form of Team-Based Learning. The main lessons learned are discussed thematically below.

Preparation and Technical Setup

Thorough preparation is crucial for a successful iRAT-tRAT session. The technical setup largely determines how smoothly the session runs, which makes close collaboration with the assessment organization essential. It is therefore important to involve them from the very beginning, so that scripts, task allocation, and technical coordination can be developed in a timely and consistent manner.

A significant organizational challenge concerns seating students in the assessment software. This is a labor-intensive task in which each student must be assigned a fixed workstation in advance. Errors in this configuration immediately affect both the execution of the tRAT and the subsequent analysis of team results.

It is also necessary to anticipate a substantial percentage of students who register but do not attend. During the first iRAT–tRAT session in 2022, 40% of registered students did not show up; nevertheless, the team allocations could remain intact and the feasibility of the session was not compromised. At the second session in late 2023, this number increased to 50%, possibly due in part to poor weather conditions. As a result, two of the 36 teams consisted of only one student; these students were subsequently reassigned to another team.

iRAT: Individual Timing Differences and Feedback

During the iRAT, students showed considerable variation in how they used their time: some carefully analyzed all answer options, while others finished within 10 minutes. This variation can create restlessness when faster students have nothing left to do. It is therefore advisable to ask students to bring a quiet activity for the moment they finish early.

Students did not see their iRAT scores during the session, since access to these scores would have enabled them to answer all questions correctly during the tRAT phase. To ensure that they could focus fully on the explanation during the plenary session, the complete answer key was posted in the learning management system afterward.

tRAT: Team Size and Participation

During the tRAT, students work together on a single shared screen. In teams of four, this functions well: the screen remains visible to all team members, and participation remains balanced. In larger teams, however, a second ring tends to form around the screen. Students seated further away have a less direct view of the questions and answer options, which makes it more difficult for them to engage actively in the discussion. As a result, in teams of five or more, these students appeared to listen mainly, while those seated closest to the screen took a more active role in the team discussion. The optimal team size therefore seems to be four students, where both screen visibility and participation remain balanced. For future sessions, it would be useful to ask students what they consider the optimal team size.

In addition, it became clear that students at Tilburg Law School, due to their familiarity with individual exams, seem to need some encouragement to discuss actively, share arguments, and defend their own answers. This requires explicit instructions beforehand and a clear statement of the behavioral expectations within the team.

Plenary Review: Structure and Cognitive Load

Because the plenary phase lasted only 30 minutes, the lecturer began with the questions students had found most difficult. He consistently applied the same didactic structure: first explaining why the correct answer is correct and how to arrive at it, and then clarifying how the incorrect answer differs from the correct one.

Cognitive load proved to be an important consideration. Students process the questions consecutively on three levels: individually, in teams, and then in the plenary session. A break of 10 minutes between the tRAT and the plenary review may help reduce the impact of the cognitive load. Finally, it is important that all students have writing materials or a laptop available for taking notes during the explanation.

Lecturer Reflection

After completing his long-standing career as Professor of Legal Philosophy at Tilburg University in 2025, Hans Lindahl reflects on the multi-year trajectory developed in the course Philosophy of Law B:

Philosophy of Law is a difficult subject for law students. A significant part of law courses is dedicated to conveying information, focusing on providing students with basic doctrinal concepts and the general structure of a subdiscipline. By contrast, courses in philosophy of law, certainly Philosophy of Law B, focus on argumentation rather than

information. The course syllabus is thin, and the PowerPoint sheets offer the steps of an argumentative process that a) begins with legal doctrine, then b) lays the groundwork for a philosophical question hidden in doctrinal questions, after which c) philosophical analysis renders the problem in its foundational complexity, concluding by d) returning to the doctrinal issues and showing how philosophy contributes to clarifying the different perspectives that doctrinal development could take in dealing with those issues.

Both the multiple-choice section and the essay question place a premium on testing the students' analytical capacities to draw and apply conceptual distinctions on the basis of the course material. The exams of Philosophy of Law B are surely among the most difficult encountered by students during their bachelor's studies.

Consultation with Vincent Kalis and Bart van de Moosdijk led to a two-tier plan of action aimed at improving exam results without loss of difficulty. The first tier involved having students prepare and test multiple-choice questions themselves on a weekly basis, whereas the second tier consisted of an iRAT-tRAT mock exam. This two-tier plan of action contributed to a better understanding by a wide range of students of the nature of philosophical reasoning and had a positive impact on the exam results of those students who participated regularly in both of its tiers.

A key factor in the formative approach's success was the well-considered technical infrastructure, complemented by the dedicated support from colleagues at Tilburg University. The time and effort invested ultimately resulted in enhanced student understanding, increased engagement with the course content, and appreciative feedback from students. It was both a pleasure and instructive to work with Vincent and Bart. My heartfelt thanks to both! (H. Lindahl, personal communication, June 10, 2026)

Acknowledgments

We would like to express our sincere gratitude to Professor Hans Lindahl for his enthusiasm and openness to innovation within Philosophy of Law B. His commitment to preserving the intellectual ambition of the course while strengthening students' learning process was instrumental in the successful development of this formative trajectory.

We are grateful to the students in the five Philosophy of Law B cohorts from 2019 through 2024 for their participation and valuable feedback, which provided important insights into the educational value of the formative questions, e-voting, and the iRAT-tRAT practice exam.

Special thanks go to Kim Stultiens for her assistance in selecting questions that met the psychometric criteria and in entering them into the assessment software. We also thank Team E-assessment for their close collaboration and contributions to the design, pilot testing, and organization of the on-campus iRAT-tRAT sessions, in particular team leader Thessa van Domburg and assessment software administrator Edwin Yuen.

We would also like to thank lecturers Brandt Weber and Bent Bos for their support of Professor Hans Lindahl, particularly in preparing the questions in the e-voting software and coordinating student registration for the on-campus tRAT sessions.

Finally, we would like to thank Hans Slabbekoorn, Advisor in ICT and Education, for sharing his experiences with the iRAT-tRAT format at Amsterdam UMC, the Netherlands.

3.3

Exams with Mixed Question Types⁵³

Peter de Goeij

Course Overview

Investeren en Beleggen (Investments) is a mandatory course in the second semester of the first year of three BSc programs: Business Administration, Economics and Business Economics, and Tax Economics. It introduces students to two key principles of corporate finance: (i) deciding in which long-term projects a company should invest and (ii) how to raise the capital for these investments. The topics include cash flow analysis, time value of money, capital budgeting, pricing of bonds and stocks, the risk-return trade-off, portfolio theory, and market efficiency.

As the first of two mandatory finance courses for BSc students, it equips students with the knowledge and techniques needed for essential financial decision-making in business practice. As such, it provides the foundation for advanced finance courses such as the third-year minor finance courses and the courses in the Tilburg University MSc Finance program.

The average student enrollment on the course is roughly 400 students. Teaching is provided by two academic staff members specializing in finance, supported by teaching assistants. Instruction consists of lectures, tutorials, and supervised practice sessions.

Learning Goals

The course aims to provide students with a solid foundation in corporate finance and investments by developing both knowledge and skills. Learning goals include interpreting and analyzing financial theories, principles, techniques, and the broader social context of finance. Students also learn about new developments in the field. On the skills side, students practice reading academic literature, conduct basic research by using fundamental research methods, and apply theory to practical problems. Analyzing, structured problem-solving, and adaptability to new financial situations are key **learning outcomes**. By combining disciplinary knowledge with academic and professional skills, the course contributes to multiple Program Learning Outcomes, particularly those related to subject expertise, research literacy, and critical reflection.

The achievement of these learning goals is assessed primarily through exams. The midterm (25%) and final exam (60%) test whether students can interpret accounting statements, apply capital budgeting techniques, and value stocks and bonds. The final exam also covers the more advanced goals (that are taught in the second half of the course), such as applying the arbitrage pricing theory, the law of one price, investment decision-making, risk valuation, and portfolio theory. In this way, the exams provide the main evidence that students have met the intended

53 I thank Louise van Hoek and Ya-Ping (Amy) Hsiao for their inputs that greatly improved this chapter.

outcomes of the course. The group assignment (15%) and **formative assessments** play a supportive role by fostering research skills, teamwork, and continuous practice.

Teaching, Supervision & Learning Activities

The course combines lectures, guest lectures, tutorials, instruction lectures, and self-study. The lectures are in-depth and build theoretical knowledge, while the guest lectures link the finance theory to real-world contexts. During the weekly lectures, students are confronted with online (formative) questions that provide them with feedback on how well they understand the topics of the lectures. The (almost weekly) tutorials and instruction lectures develop problem-solving, methodological skills, and critical reflection. Students are asked to solve exercises in advance, so the lectures can be used to provide them with the necessary feedback on how well they can apply the learned theory to practical problems. Students also must complete an empirical group assignment. This assignment strengthens research abilities, teamwork, and the application of theory to data-driven problems. This is also part of the summative evaluation of the course. Finally, self-study is an essential component of the course: students are advised to dedicate at least as much time to preparation as to scheduled class time, using the practice exercises and video solutions provided via Canvas (the TiU Learning Management System) to test and strengthen their mastery of the material.

Course Assessment Design

There are three summative tests that students take that are used to calculate the final grade of the course. All components are graded on a 100-point scale and after taking the weighted average of these summative tests this is converted to the 10-point final grade scale and if students obtain a grade below 1, it is adjusted to 1, based on the university policy.

The course assessment includes a midterm exam of 25%, a final exam of 60%, and students complete a mandatory group research assignment of 15% using empirical data. If students fail the course at their initial try, there is a resit exam of 85% (covering the whole course equally). Together, these assessments test theoretical knowledge, application of finance concepts, research skills, and teamwork. Students can earn a small extra credit bonus point when they voluntarily participate in four online quizzes that consist of 5 multiple-choice questions (MCQs) each.

Students can practice an old midterm in the TestVision environment but due to digital exam capacity limitations, the midterm exam (that consists of multiple-choice questions (MCQs)) is made on paper on campus. This is also the case for the final (and resit) exam, but then the digital platform ANS is used to correct the exam online, which speeds up exam correction and facilitates feedback to students.

Midterm Exam Design

The midterm exam (25%) provides an early checkpoint, ensuring students have a solid grasp of the foundational concepts from the first part of the course. This course is taught in the second semester of the first year of the BSc programs. Despite being in their second semester, many students are still adapting to independent study and often struggle with consistent preparation. By offering the midterm exam halfway through the semester, the student receives first feedback on how they are doing in the course. If a student performs poorly on the midterm, they still can pass the course by obtaining higher grades for the group assignment and the final exam.

The midterm exam consists of 15 MCQs. Students are given 90 minutes to complete the exam. The reason to offer a midterm exam in MCQ format is to provide students with possible answers for each question. This helps them focus on understanding the material and completing the exam on time. A side benefit is that it is much easier for the lecturer team to correct the exam.

The questions consist of a combination of theory questions and calculation exercises.

An example of an MCQ that combines theory, and calculation can be found below.

Example MC-Question

You have the option to purchase bonds A, B, or C (see the table for the characteristics of the bonds).

Based on the information provided in the table, it is possible to calculate the Theoretical Market Value (TMV). Assume that the market price of each bond is equal to the TMV, which can then be compared to the face value of the bond. The face value is €1,000 for each bond.

Bond	Coupon Rate	Yield-To-Maturity	Time-To-Maturity
A	3%	4%	4 years
B	5%	5%	5 years
C	7%	6%	6 years

Which of the following answers is **correct**?

- A. Bond A is traded 'at par', bond B is traded 'at a premium', and bond C is traded 'below par'.
- B. Bond A is traded 'below par', bond B is traded 'at par', and bond C is traded 'at a premium'.
- C. Bond A is traded 'above par', bond B is traded 'at a premium', and bond C is traded 'below par'.
- D. Bond A is traded 'below par', bond B is traded 'above par', and bond C is traded 'below par'.
- E. Bond A is traded 'at a premium', bond B is traded 'at par', and bond C is traded 'at a discount'.

This question is related to the learning goals of interpreting and analyzing financial theories, principles, and techniques. They also must apply theory to a practical problem. The rationale for including such a question is that it combines theoretical knowledge and understanding with calculation skills. In particular, for this question, they need to understand how to calculate the theoretical market value of a bond. Next, they need to understand the concepts of when a bond is trading at, over or under par (or equivalent terms for the same concepts). Finally, by applying theory knowledge that is covered in the lectures, students should understand/recognize that bond B is trading at par (€1,000) as the coupon rate and yield-to-maturity are the same. This would allow the students to solve this question more quickly, which would give them more time to answer the other questions.

Final exam setup

The final written exam (60%) carries the greatest weight, as it assesses comprehensive understanding and the ability to integrate concepts across the entire course content. The format of the final exam is a mixed exam and consists of three parts for which students can earn up to 100 points.

The first part consists of two constructed-response questions where students can earn up to 40 points. These questions require students to provide detailed written solutions. Partial credit can be awarded if the reasoning is correct, even when the final answer is incorrect. The constructed-response questions are offered to allow students to provide more elaborate answers, which allows the lecturers to test the thinking process of students. Then it is possible to judge the argumentative or thinking step capabilities of students. It is more time-consuming to correct constructed-response questions and therefore it is only limited to 40% of the exam.

An example of a constructed-response question can be found below.

Example Constructed-Response Question

Ben is considering investing in a portfolio of two risky companies A and B (A and B shares) and the risk-free alternative. The prices at which A and B shares are currently traded on the stock exchange are €20 and €15 respectively.

Ben buys 2000 B shares, while he goes short for 500 A shares. Ben also invests €5,000 in the risk-free alternative. Collectively, this portfolio is called the 'total investment portfolio' (TIP).

- A. In the total investment portfolio TIP, Ben has a short position in share A. Explain the purpose of a short sale and how it works in practice.

In addition to the TIP portfolio, it is also possible to make an equally weighted portfolio of the A and B shares. This portfolio is called RP and the beta of this wallet is equal to 1.10. The annual (expected) market risk premium is equal to 6.0% with a standard deviation of 15%, while the return on the risk-free alternative is equal to 2.0% (this is the 1-year risk-free rate). It is also given that the correlation between share A and the market portfolio is equal to 0.90, while the correlation between share B and the market portfolio is equal to 1.05.

Assume that the Capital Asset Pricing Model (CAPM) model is correct, so you can use it in your calculations.

- B. Calculate the expected return of portfolio RP.
 C. If the expected return of share A is equal to 9.0%, calculate the beta of share A. Also, calculate the beta and expected return of share B.
 D. Calculate the beta of portfolio TIP.
 E. Calculate the expected return of portfolio TIP in 2 different ways and show that they are (of course) the same.

This constructed-response question is related to the learning goals of interpreting and analyzing financial theories, principles and techniques. They also must apply theory to a practical problem. The argument to include such a question is that it combines theoretical knowledge and understanding with calculation skills. It combines knowledge of the theory (in this case the CAPM and its implications) and how to use it in a practical setting using calculations. This question also borrows knowledge and skill for the statistics course (calculation of expectations and understanding of variances and covariances). Moreover, in sub question E it is asked to combine their theoretical knowledge of the CAPM and its implications. If they make any calculation mistakes, they will discover them here and can use this knowledge to correct their answer in the previous subsection(s) or, in case of lack of time, acknowledge their mistake. In this latter case, points will be awarded.

The second part consists of ten MCQs that are very similar in setup as the questions of the midterm MC exam, and an example of such a question can be found above. Students can earn up to 30 points for this part. The rationale for including MCQs in the exam is the same as that of the MC midterm exam. To provide students with possible answers for each question which helps them focus on understanding the material and completing the exam on time. It also allows us to cover more specific topics on the exam as it is possible to focus on a small detail of the covered materials. As mentioned before, a side benefit is that it is much easier for the lecturer team to correct the exam lecturer team.

Finally, part three consists of 15 statements where students can earn up to 30 points. These statements consist of a combination of theory questions and calculation exercises but are phrased in the form of a statement that is either true or false. Students can earn 2 points for every statement that they answer correctly, but for every wrong answer, 1 point will be deducted. Students also have the option to not answer the question, but they must deliberately choose this option on the answer form. The statements in part three of the exam are meant to force students to answer true/false questions, but also to confront them with self-reflection if they are certain about their given answer.

Examples of theory and calculation statements can be found below.

Item x2

Statement: If a company goes bankrupt, its remaining assets are sold by the appointed trustee, and all the proceeds are then completely distributed to the shareholders.

- A. The statement is correct
 B. The statement is incorrect
 C. I choose not to answer this statement and receive 0 points

Item x3

As you read in the midterm exam, Dominican baseball player Juan Soto signed a record-breaking contract worth US\$765 million with the New York Mets ahead of the new season, even though he played for their major rival, the New York Yankees, in the previous season. The latter team also made him an offer, but in the end, 31-year-old pitcher Max Fried signed a US\$218 million contract with the New York Yankees.

Max will play for the New York Yankees for 8 years, and the specific financial details are as follows: The first payment of the contract is made at the very start (on March 1, 2025). This payment includes a signing bonus of US\$10 million and a salary of US\$12 million. This same payment will also be made next year, after which there will be six additional annual payments of US\$29 million. Assume all payments take place on March 1 of each year.

The Major League Baseball organization uses a discount rate of 4% per year. What is the Present Value (PV) of Max Fried's contract on February 28, 2025 — one day before the first payment is made?

Statement: If the discount rate r is 4% annually, then the present value of Max Fried's contract on February 28, 2025 is **greater than \$190 million**.

- A. The statement is correct
- B. The statement is incorrect
- C. I choose not to answer this statement and receive 0 points

These statements are related to the learning goals of interpreting and analyzing financial theories, principles and techniques. They also must apply theory to a practical problem. The purpose of including such questions is because it combines theoretical knowledge and understanding with calculation skills. The first statement is a pure knowledge question. The second question is related to the knowledge and skill of how to calculate the present value, but in a real-life example, which is not as straightforward as the textbook examples because it has varying cashflows during the duration of the contract. Phrasing such a question in the form of a statement, instead of an MCQ, does not allow the student to answer the question by 'elimination' as only one potential answer is provided.

The exam is drawn up by one of the lecturers and then solved and corrected by the other members of the faculty lecturer team (i.e., peer review). In addition, we ask one or two of our teaching assistants, who are involved in the teaching of the tutorials and who took the course one or two years before, to also solve the exam to make sure that it is representative and that it would not take too long to solve. If needed adjustments are made.

Resit exam setup

The setup of the resit exam is the same as the final exam. The only difference is that 50% of the questions are related to the materials that are covered in chapters 1-8 of the book and 50% of the questions are related to the materials that is covered in chapters 9-13 of the book ensuring all course learning goals are covered.

Reflections on Challenges and Adaptations

Feedback

Feedback is provided to students on several occasions. During tutorials and instruction lectures, they are provided with feedback on their own solutions. Moreover, via Canvas students are offered a practice midterm exam in TestVision where they must solve 15 MCQs within 90 minutes, thus the same setup as the midterm exam. This practice exam is automated, and students can participate as many times as they want and receive a different exam every time (although if they participate many times, the same question can pop up). Feedback is provided by a standard written-out solution, while for some (more difficult calculation) exercises a short solution video is available.

After each **summative assessment**, there is an inspection planned where students can compare their answers with the solution manual and check and learn from their own mistakes. It is made clear to the students that the inspection is used to learn and not to ask for extra points to obtain a higher grade.

Exam Preparation

In some cases, it is difficult to create new questions as there are only limited ways we can ask those questions. This is especially the case for theory questions. For the calculation questions, this is easier as we can change smaller things to make a different question that still tests the same knowledge and skills to solve it. Changing constructed-response questions is not necessarily difficult, because in many cases these are calculation exercises, however, it takes some time to create these types of questions and make the standard solutions.

Scoring/Grading

Constructed-response questions are scored using a structured calibration process to ensure **fairness**. There are **items**/subjects for which there is only one way to solve or to answer it. If students do not have this correct, they lose all potential points. For other items/subjects, there are multiple ways to derive or come up with the correct answer. The standard solution only contains one solution, so if students solve it in a different way, during the correction of the exam, this is then considered by augmenting the standard solution. In some cases, students make very similar mistakes in the constructed-response questions. If this happens many times, we check the question itself carefully and see if it is clearly formulated. If it is, then the standard solution is not changed. If not, then we revise it accordingly and re-correct the already corrected exams.

Time Management for Grading Using ANS

In courses with many students, grading efficiency is a critical consideration. Constructed-response questions, while valuable for assessing higher-order thinking and problem-solving skills, are time-consuming to evaluate individually. However, advances in digital exam environments, of which TestVision and ANS are examples, have provided us with the tools to streamline the grading process for constructed-response questions.

The final (and resit) exams are programmed into the ANS system. After the exam content has been checked by the lecturer team and the layout and technical things have been checked by the e-assessment support team, the exam is printed on paper. When the exam is completed, all exam documents are scanned and made digitally available in ANS. This allows examiners to grade the answers in a digital environment. The grading team can define assessment criteria digitally, automatically recognizing correct solutions, and flagging alternative valid approaches. For this course, we estimate that it reduces the manual correction time (of the constructed-response questions) by approximately 50% while maintaining consistency and fairness in scoring. The system also supports calibration, allowing examiners to adjust the standard solution if students provide alternative correct answers.

Editorial Feed-Forward for Future Consideration

Please also mention the percentage of course learning goals (CLGs) tested in the midterm (XX out the XX CLGs) so that it is clear why only 15 MCQs are used (because you only test partial CLGs).

If you have any test and **item analysis** report, can you try to use the Spearman-Brown formula to predict how **reliability** changes when you lengthen the exam? Please see this online free calculator. <https://www.cogn-iq.org/statistical-tools/spearman-brown-prophecy/>

This reference can be useful for making shorter exams because of the practical limitations at the organization:

Kingston, N., & Tiemann, G. (2010). Spearman–Brown Prophecy Formula. In Encyclopedia of Research Design (pp. 1403-1404). SAGE Publications, Inc. <https://doi.org/10.4135/9781412961288.n427>

3.4

Introducing Oral Examinations in a Program

Anton ten Klooster

Background

The introduction of oral examinations in the moral theology course was, in theological terms, a *felix culpa*: a happy fault. In 2020, the imposition of restrictions against the spread of Covid-19 led to a change in assessment formats. The final exam of the first-year introduction moved from a written in-person to an oral examination online. After the pandemic, this format was maintained because it was believed to do more justice to students in this stage of their formation. In this chapter, consider the format of the oral examination, its limitations, and possibilities for further implementing it in light of the rise of artificial intelligence (AI).

Oral Examination in the Design of the Moral Theology Track

In the present design of the courses on moral theology, as the catholic form of Christian ethics is often referred to, the first of three courses is concluded with an oral examination. This means that the students who will take the full BA program in theology will have four different types of **assessment**: an oral examination in the first year, an in-person written midterm exam and a term paper in the second year, and a research paper on a single monograph in the third year. The forms of the examinations evolve with the students, with each form assessing further skills in Bloom's taxonomy (see Table 3.4.1).

Table 3.4.1
Forms of Assessment in Relation to Bloom's Taxonomy

Year 1 <i>Ethics and Character</i> , 3 EC	<i>remember, understand, apply</i>	oral examination
Year 2 <i>Catholic Ethics in Practice</i> , 6 EC	<i>understand, apply, analyze, evaluate</i>	midterm exam, term paper
Year 3 <i>New Developments in Ethics</i> , 6 EC	<i>analyze, evaluate, create</i>	research paper, public communication draft

In the larger context of the theology program, this type of examination can be considered as a baptism of fire for presentation skills. Many of the students go on to contexts of ministry, chaplaincy, or education, contexts where they

can be put on the spot to 'say a few words', speak to people in crisis situations, and respond to difficult life questions. While there are no points for composure in the scoring table, it does go a long way for the students themselves in making it through the exam comfortably.

As mentioned in the introduction, the initial use of oral examination as an assessment form was incidental. The continued use of the form was not due to its practicality. In fact, it often requires more of a lecturer: more time, more readily available knowledge of the details of the course, and more justification of the grading. But my view is that this type of examination also gives the lecturer more opportunity to find out what students *do* know and understand, even when they are not immediately able to reproduce a term or definition. Take the following dialogue, for example, between the examiner (E) and the student (S), which is an amalgamation of about a hundred examinations. An example is a question on the parts of prudence in virtue theory, which are docility, memory, and *solertia* (nimble decisiveness).

E: Can you name the three parts of the virtue of prudence, as they are mentioned in the handbook? And can you explain them?

S: ... there was docility, solertia... I can't remember the last one.

E: Would you like a hint?

S: Let me think... Yes, please.

E: You say need to remember it...?

S: Ah, yes: memory!

E: Alright, so the three are...

S: Memory, solertia, and docility.

E: Wonderful! Now can you please explain what they mean?

The objective of the course is twofold: students need to have basic vocabulary to follow the rest of the program and what they remember is tested. But they also need to understand and begin to apply the terms. In this example, there is a small gap in knowledge, but when hinted the student does have an opportunity to further demonstrate their understanding. There are also cases where students receive a hint (i.e., phrasing the question or term) and are still unable to explain, and in this case, they show they neither remember nor understand the course material. In that case, they receive no points, in the example there is a small reduction for the failure to reproduce. And having explained it successfully, the student will probably never forget it again!

Organizing Oral Examinations

The setting of an oral examination is often a one-on-one format, and this can be intimidating to students. This cannot be taken away entirely. What is crucial is to organize the exams in a setting that is as transparent as possible and socially safe. There are several complementary strategies to achieve this. **Transparency** can begin in quite a literal way: making sure that both student and examiner are visible: through street-level windows, or glass doors or such on the inside. A simple way to create a feeling of safety is by not being between the student and their access to the door. In the absence of a second examiner, recording the exam serves as a 'version of record' of the exam. The student is informed of the recording, that it will be stored for a limited time on a university server, and that they can have it listened to by the Examination Board in case of disagreement. This communicates to them that they are not simply at the mercy of the examiner, but that this assessment too can be challenged and, if necessary, corrected. Finally, communicate to the student how long the exam will take, and when they will hear their grade.

What can be done beforehand is to clearly communicate to the students in class what the exam will be like, and how the score will be determined. This can be done at the end of the last class: show them the scoring table that is

available to them on Canvas, and most of all: stress that the very reason we do this kind of exam is not to trick them into saying things that are wrong, but to help them to say all the things they know, and to explain these.

Preparing Students for an Oral Examination

The students take seven classes before they conclude the 3 EC course with an oral examination. Because it is a first-year BA program, the course is designed on the assumption that all the students are new to university education and testing. Part of the first class is used to show on screen where the information on the exam can be found. Throughout the course, the lecturer remarks on what is more and less significant regarding the exam. And at the end of each class, there is a slide with sample questions for the exam. The students hear what kind of answers are required, and they can ask questions. When there is time, there can be a quick quiz in class. In this way, they can already rehearse the format a bit.

At the end of the final class, students are once again shown the online instruction and the **rubric**. It is emphasized that the assessment is designed to help them succeed, not engineered to trick them into failing. There are two key things that are communicated on Canvas: the types of questions and the grading table.

There are four categories of questions, with most focusing on students recalling core concepts and explaining these. It is made clear to them that these constitute the benchmark for passing the exam. Questions on explanation and application, and perhaps a reflection question can help improve the grade even further. The instruction reads:

At the beginning of the exam, the examiner will set a timer for 25 minutes. In this time, you will receive 10 questions, all from the slides. You will need to answer at least 6 questions correctly for a passing grade, the correct answers are those you can deduct from the book used in the course.

Category1 – Question about definitions: ‘name’, ‘what is...’ etc.

For example, what is *epikeia*?

Category2 – Questions about lists.

For example, what are the conditions for the principle of double effect?

Category 3 – Questions about explaining and applying certain principles.

For example, what is the role of prudence in scenario X?

Category 4 – Reflection question. Here you can show you mastered the material.

For example, do you think that virtue ethics is indeed universally applicable?

The four categories of questions assess knowledge (1, 2) and explaining and application (3, 4). To the students, this is communicated in terms of grading: you need to answer type 1 and 2 questions correctly to pass, type 3 and 4 will help you improve your grade. The assessment description explains this:

You build up your grade by answering questions, moving from topic to topic in the order of the book and the classes. The basis for your grade are questions from categories 1 and 2: you will need to know the terms of the book and be able to explain them. The more you can explain and show you understand the material (category 3, and eventually 4), the higher your grade will be. Each time, we will start at category 1 and move to the next one when you have shown sufficient knowledge. If you are unable to answer the category 1 questions, the exam is terminated and you do not receive a passing grade. If you miss certain parts, you will score a bit lower but will still be able to get credit for a partial answer. For example, you cannot name the parts of prudence, but after hearing them from the examiner you can explain what they are. You will get credit for the explanation.

The scoring table follows a similar narrative format. This does justice to the fact that oral examinations, certainly in ethics, are not an exact science.

Grades are given in half-decimals, and within the following ranges:

- 1: There were no correct answers at all to category 1 questions.
- 2-3: You did not answer category 1 questions correctly and were unable to explain basic principles.
- 4-5: You had some correct answers, but there are considerable gaps in your basic knowledge and understanding.
- 6-7.5: You have a broad grasp of the material and can explain things, but there's enough room for improvement. But you're good to go for year 2!
- 8-9: You have broad knowledge and show a deeper understanding.
- 9.5: You really know how to explain what is happening in the book. Okay, I got you on one or two things, but I had to try hard.
- 10: You did not only answer the exam questions perfectly, you understand the structure of the book and its relation to the field of ethics.

As is usual in the Dutch system, the full marks (10) are not the starting point but an exception. In five years, it has been given once. It is reserved for the student who not only meets expectations but exceeds what may be expected at this level. Of course, in relation to the material, the student needs to know the basics of the course, and giving a brilliant explanation of quantum physics in relation to ethics unfortunately will not suffice.

Taking and Assessing the Examination

With the student prepared, the examiner can assess their knowledge and insight. When they hear the exam annually, at some point they will know the answers by heart: the definition of conscience, the three (or four) conditions of the principle of double effect, and what the Latin term *habitus* describes. There is often a precise element in these questions, with follow-up questions to ascertain this. For example, the first question on conscience, with a well-prepared student.

E: *Can you give me the definition of conscience, please?*

S: *Conscience is a judgment of reason of the moral quality of actions that someone has done, will do, or plans to do.*

E: *Wonderful, thank you. And who is doing the 'judging' in this case?*

S: *The person themselves.*

It does not matter if it refers to 'someone', an 'agent' or 'the person'. The follow-up question makes clear that the student understands what conscience is, according to this definition. In the follow-up, the student can explain this a bit more.

E: *So, what does that mean, to judge with 'reason'?*

S: *Kind of like we discussed in class: it's not simply a feeling or an intuition, but it can be argued rationally.*

This is more than enough for a good answer. No need to draw the discussion out further. On paper, this all looks simple. But there is always a percentage of students that show up ill-prepared, thinking they can bluff themselves through the exam. Typically, they will freely be associated based on words. In response to the question of conscience, they might reproduce popular images of 'knowing what is right and wrong', but it lacks reference to reason, or actions of the person themselves.

The order of the questions follows the order of the classes, and the students know this beforehand. If they struggle with the first topic, the lecturer can reassure them and say there are more chapters to discuss. Each part begins with

the category 1 and 2 questions, and if they do well in these, more detailed questions are added. For example, if they have explained the principle of double effect, which has three terms to assess the moral acceptability of an action, they are asked to give an example. This is a higher skill, and if they do consistently well on these, this earns them an 8 or 8,5 depending on the clarity. If this is too much to ask, their grades are in the 7 or 7,5 if they at least understand and explain the concepts well, and a 6 and 6,5 if there are some gaps in their knowledge but they get the course material overall. Students who excel get a small number of more difficult questions, still staying close to the material. If a student does that flawlessly and can reflect on the course entirely, by answering a meta-question well, they do get full marks.

Because the exam is designed to bring out what students do know, the final question -after the reflection- is always the same: “You have answered a selection of the questions you studied for. Is there something you would have liked to answer, something you prepared so well and now weren’t asked anything about it?” Usually, there is something, and it can help assess a student’s understanding if the examiner is still in doubt about a half in the score. After this, the grading is explained to the student, with the emphasis on what went well. To give a few examples, in ascending order.

I saw you struggled with some of the basic concepts. For example: you weren’t able to explain habitus, or name the cardinal virtues. Since these are the basic type questions, I will have to fail you. There will be another opportunity, and if you have questions before that, you can reach out to me, of course.

All the information is there, and you can reproduce it. But there are gaps. You can explain it, but there are still bits and pieces to be added. But given that you’ve shown it’s all there, I give you a 6.0.

It’s all there, you explained it well. Some of the details could have been more precise, and that tricky list of conditions for a just war got to you. But you know what’s in the course material and can explain it fairly well, so that is a 7.0.

This was a good exam: you knew the terms, explained them well. If I push a bit further, there are things to better understand, but that’s why we’re only in year one, this reflects an 8.0.

This was almost perfect. You struggled a bit with the concept of natural law, which is understandable at this stage. Since that was the only thing between you and a perfect exam, you receive a 9.5.

After this explanation, the student is asked “Do you think that’s a fair assessment?” They usually do, or it is discussed further; in a rare case, there is one extra question when a student insists was not able to demonstrate their knowledge. At this point, the recording is ended, and the student leaves the room.

Lessons Learned: Personal Reflection

In the age of GenAI, many lecturers are thinking of using oral exams to solve the problem of authorship issues. In fact, oral examinations do not solve that many problems. They are time-intensive, they take some preparation during the course since many students are new to the format, and the examiner needs to stay alert while hearing different students explain the same basic concepts for hours on end. The main ‘problem’ for which it is a solution is the lack of variety in forms of assessment. But when done well, it is a transparent, reliable, and valid way of testing a student’s knowledge. In the school of theology, it also helps to diversify assessment forms. For a long time, there was an emphasis on written assignments, and more specifically term papers. This is great for students who do well in this domain. But some students are simply better at expressing themselves verbally, because they are dyslexic, for example, or just because writing is not their strong suit. Having oral examinations in the mix does justice to the

diversity of the student population in terms of innate skill and preference. As mentioned, most of the students end up in positions where they will be doing more speaking than writing, so knowing how to respond, and to be in dialogue about what students know, is a wonderful skill.

The first condition for oral examinations is time. In this case, there is a thirty-minute time slot for each student. With them having to travel to the university, exams can usually begin at 9 or 9.30 am and begin the last exam at 5 pm. With a lunch break for the humanity of the examiner this means one lecturer can roughly take ten exams per day. Realistically, this means that one examiner can do this type of examination for no more than 30 students, or the workload would become excessive. Alternatively, the exams could be slightly shorter, but this may disadvantage students who need more time to think and formulate. Another option is to cluster students, as was sometimes done in the school at the beginning of this century. This only works if the examiner can make sure they both answer a similar number of questions. To some, it actually may help with the nerves, and in dialogue they may know more – which will need to be corrected for in terms of grading. But some people do better alone, or may just not get along with the person they're taking the exam with. So, in principle, it is preferable to assess students individually.

Oral examinations are good for quickly assessing lower-level learning skills. The more complicated ideas get, the more time is needed to hear their explanation and probe the depth of the student's knowledge of them. In this case, an oral examination may work with a very limited number of questions. In the age of AI generated texts, it might be a helpful assessment in addition to written work. In that case, the student submits a written work, and the oral examination allows the lecturer to push the student on certain things and ascertain that they understand the ideas they have presented in written form and can expand on these.

But most of all, if the conditions are there to take them in a socially safe setting and with sufficient time, oral examinations are great for lecturers who want to help their students learn to express their knowledge and to demonstrate what they do know. With the proper preparation through the course, by weaving references to the assessment into the classes as we discussed above, it is a form that students can learn to appreciate, and it prepares them for similar situations in their professional future.

To conclude, in my view, oral exams are perhaps the most subjective form of assessment, with numerous construct-irrelevant factors influencing both student performance and examiners' judgments. Without raising examiners' awareness of these sources of subjectivity and adequately preparing students, the **validity**, **reliability**, and **transparency** of oral assessments are compromised. Consequently, there is no assurance that the intended **learning outcomes** are being accurately assessed. If oral exams are adopted primarily as a response to AI-generated work, this objective is unlikely to be achieved unless these fundamental quality issues are addressed.

3.5

A More Neuro-Inclusive Assessment of Presentation Skills

Yvette Drissen, Hanne Jacobs, Sinsemilla Blokvoort, Frea van Dooremaal, Minne Bakker, and Ya-Ping (Amy) Hsiao

Introduction

In this chapter, we present an educational innovation developed to improve presentation skill training and **assessment** in the Bachelor program of Philosophy. The innovation was implemented in the first-year course called ‘Contemporary Social and Political Philosophy’⁵⁴, which already contained a presentation skill component. In the original setup, students received information on the expectations and the grading **rubric**, after which they were ‘thrown in at the deep end’. Throughout the course, each student had to give a 30-minute presentation on a primary philosophical text followed by a 15-minute Question and Answer session (Q&A). In order to pass the course, the students needed to pass the presentation assignment meaning that they had to get at least a grade of 5,5 for this component (30% of total grade).

Through feedback from students and the study advisor, we learned that several students found giving a high-stake, one-off presentation daunting and that certain aspects of our grading rubric and the format of the assignment might disadvantage neurodivergent students. More specifically, the rubrics and setup mostly seemed to benefit students who already felt comfortable giving presentations in the traditional neurotypical way (e.g., making eye-contact, speaking with an animated voice, demonstrating self-confidence, responding on the spot to questions). The grading rubric and assignment setup hence appeared to erect barriers in the educational environment for neurodivergent students to develop presentation skills in a way that feels authentic to them. The student feedback supplemented with insights from the academic literature (see Blokvoort et al., 2025) urged us to rethink what a ‘good presentation’ even means and made it questionable whether the standard way of training and assessing presentation skills is flexible and inclusive enough in that it allows for a variety of successful presentation styles. In response, we completely revised the way we train, teach, and assess presentation skills. These changes were implemented during the academic year of 2024-2025. While our innovation continues to work in progress, we invite other lecturers to reuse the materials we have developed and adapt them to their particular context.

54 <https://uvt.osiris-student.nl/onderwijscatalogus/extern/examenprogramma/16300/7A200-2025?taal=nl>

The Innovation

The course learning goals focusing on presentation skills remained the same as before, namely:

- The student can analyze a complex philosophical text and reconstruct the most important argumentation structure.
- The student can give a clear presentation in academic style that relates to, but also goes beyond, the content of the lectures.

However, the path that students need to follow to achieve these goals has changed substantially. Importantly, we first adjusted the rubrics/evaluation criteria. Engaging presentations come in different forms and formats. We wanted to encourage students to explore and develop their own authentic presentation style, which allows for flexibility in how they deliver a clear and effective presentation, while maintaining high standards. So, instead of grading students on the extent to which they make eye-contact, they are now evaluated on how their “non-verbal communication strategies support their message”. This can include walking around (rather than standing in one place, if this is what is helpful for the student to deliver their message effectively), holding a pen in one’s hand (in case a student does not know what to do with their empty hands), or looking at the wall behind the audience (instead of looking into people’s eyes). The rubric now also includes an interaction element, urging students to think about how they deliver their message and ensure that it is inclusive and suitable for a diverse audience. See Table 3.5.1 for the whole rubric of the final presentation.

The instruction and training of presentation skills throughout the semester follow a modular structure that aligns with the content of the course. The first four lectures are about justice (module 1), the following three lectures are about democracy (module 2), the three lectures after that are about power (module 3), and the final three lectures are about freedom (module 4). The *presentation skill component* of this course aligns with these content modules. That is, the texts that the students use in developing their presentation skills are also the required literature for the course, which gives them an extra opportunity to study the material. Moreover, the tutorial in the module on ‘power’ is used to reflect with the students on how power dynamics play a role in a presentation (who is given a stage to begin with and what does a stereotypically ‘good’ presenter look like?).

To facilitate the students’ development towards the final presentation, the students now work on their presentation skills on a gradual, step-by-step basis through a set of assignments that lead up to the final presentation.

Module 1 is mostly focused on the *form* of the presentation. In terms of assignments, students select a video of a presenter they recognize themselves in and look up to. They are asked to analyze the presentation style, including what they can learn from the presenter, to formulate their own personalized learning goals and action points, and to give peer feedback on the same assignment submitted by a fellow student.

The deliverables for module 1 are as follows:

- Video analysis of the presentation form: individual assignment, graded by the lecturer.
- Self-reflection report, learning goals, and action points: individual assignment, pass-or-fail, but the lecturer provides qualitative feedback.
- Qualitative peer feedback on self-reflection, learning goals, and action points of one other student. This is an individual assignment that needs to be completed within one week after handing in the other two assignments. This element is pass-or-fail. The lecturer merely checks whether the student fulfilled the assignment and gives qualitative feedback where needed.

Table 3.5.1
Grading Rubric Final Presentation

Criteria	Excellent (4)	Good (3)	Sufficient (2)	Insufficient (1)
CONTENT				
Focus	Key elements are highlighted exhaustively and with precision. Attention is perfectly distributed according to the importance of the elements.	Key elements are clearly highlighted. Attention is largely in line with the importance of the elements.	Key elements have been highlighted reasonably well. Attention is partly in line with the importance of the elements.	Key elements are poorly or incompletely highlighted. Attention is not commensurate with/attuned to the importance of the elements.
Argumentation	The author's arguments are fully understood and excellently explained. Transitions between arguments are very clear and coherent.	The author's arguments are well understood and well explained. Transitions between arguments are clear and reasonably coherent.	The author's arguments are partly understood and explained. Transitions between arguments are somewhat clear and coherent.	The author's arguments are poorly understood and explained. Transitions between arguments are unclear and incoherent.
Conceptual clarity	Relevant terms are used correctly and appropriately. Crucial terms are defined in an exemplary manner.	Relevant terms are used correctly. Crucial terms are well defined.	Relevant terms are sometimes used correctly. Definitions of crucial terms are reasonably clear.	Relevant terms are used incorrectly or unclearly. Definitions of crucial terms are unclear or absent.
Text analysis	The selection of text passages is excellent and they are clearly explained.	Crucial text passages are selected and well explained.	Crucial text passages are reasonably well chosen and explained.	Crucial text passages are poorly chosen or unclear in their explanation.
Critical reflection	Critical considerations are highly relevant and argued for in a compelling way.	Critical considerations are relevant and well-argued.	Critical considerations are partly relevant and reasonably argued.	Critical considerations are irrelevant or poorly argued.
FORM				
Conveying the message	Highly accessible, appropriate language register, and nonverbal communication facilitate the speaker's message.	Accessible, mostly appropriate language register and nonverbal communication helps get the speaker's message across to some extent.	Mostly accessible, sometimes appropriate language register and nonverbal communication facilitates the speaker's message to a limited extent.	Inaccessible, inappropriate language register and nonverbal communication hardly facilitate the speaker's message.
Structure	The objective, context, and structure are clearly stated. The conclusion is well defined and provides a good overview and feedback. The structure is logical and easy to follow.	The objective, context, and structure are clearly stated. The conclusion is well defined and provides an overview and feedback. The structure is reasonably logical and easy to follow.	The objective, context, and structure are reasonably well defined. The conclusion is somewhat limited and offers a restricted overview and feedback. The structure is somewhat logical and easy to follow.	The objective, context, and structure are unclear. The conclusion is poorly defined and does not provide an overview or feedback. The structure is not logical and difficult to follow.

Visual support	Slides are well-designed and support the message, contain an appropriate amount of information, and have very clear structural indicators.	Slides are well-designed, usually contain an appropriate amount of information, and have clear structural indicators.	Slides are reasonably well-presented, sometimes contain too much or too little information, and have somewhat clear structural indicators.	Slides are poorly designed, contain too much or too little information, and have unclear structural indicators.
Time management	The allocated time has been strictly adhered to and the time spent on each component is perfectly balanced.	The allocated time was well respected and the time spent on each component was well balanced.	The allocated time has been reasonably respected and the time spent on each component is somewhat balanced.	Allotted time is poorly respected and time spent per component is not balanced.
Interaction	The student captures the interest of a diverse audience and takes into account different ways in which people engage information. Answers during the Q&A are very clear and well-founded. The student asks enough relevant questions during the Q&A from fellow students.	The student generates a certain degree of interest among a diverse audience. Answers during the Q&A are clear and well-founded. The student asks a reasonable number of relevant questions during the Q&A session with fellow students.	The student generates limited interest from the audience and pays little attention to the different ways in which people engage information. Answers during the Q&A are reasonably clear and well-founded. The student asks somewhat relevant questions during the Q&A session with fellow students.	The student barely arouses the interest of a diverse audience and does not take into account the different ways in which people engage information. Answers during the Q&A are unclear and poorly substantiated. The student asks no questions or irrelevant questions during the Q&A with fellow students.

The instructions of the video analysis as well as the assessment criteria used by the lecturer can be found in the online supplement folder of this chapter (i.e., Instructions and Rubric Video-Analysis Module 1). The instructions for the reflection report, learning goals, action points, and the peer-feedback rubric can be found in the online folder (i.e., Instructions Self-Reflection, Learning Goals, Action Points, and Rubric Peer-Feedback Module).

In Module 1, the lecturer creates small groups consisting of 4 students. Students can indicate in an online form with whom they would like to collaborate. The lecturer makes sure that each of them is in a group with (at least) 1 classmate of choice. The remainder of the assignments in the upcoming modules are executed within these groups.

Module 2 is about the *content* of the presentation. The lecturer selects a few videos that align with the course material and asks students to write a group analysis of one of the videos. This analysis includes a summary after which the students answer questions such as, “how well did the speaker succeed in comprehensively explaining complex ideas?”, “did the speaker provide context?”, “was it easy to follow the argumentative steps?” and “(how) did the speaker use quotes?” Here, the students are graded *as a group*. The students can also adjust their individual learning goals in light of their content analysis and need to write a reflection on how they are going to incorporate the insights from the content analysis in their own presentations.

The deliverables are:

- Video-analysis about the presentation content: this is a group assignment, and all group members receive the same grade.
- Self-reflection report and (adjusted) learning goals: this is an individual, pass-or-fail assignment (as always, the lecturer provides qualitative feedback where needed).
- Peer-feedback on self-reflection and learning goals of 1 other student. This is an individual, pass-or-fail assignment and needs to be completed within one week after handing in the other two assignments.

The instructions and rubrics of the video-analysis can be found in the supplement materials (i.e., Instructions and Rubric Video-Analysis Module 2). The instructions for the self-reflection, (adjusted) learning goals and the peer-feedback rubric can be found in the online folder (i.e., Instructions Self-Reflection, Learning Goals and Peer-Feedback Rubric Module 2).

In module 3, students get their first opportunity to practice their presentation skills in the form of a panel discussion. Each group chooses a text from the required readings and divides it into sections, after which each group member presents one section. Every presentation takes about 5 minutes. Then, a group discussion takes place in which the students are expected to ask each other clarificatory questions and engage in a discussion. Afterwards, the students write an individual report in which they reflect on their learning goals.

The deliverables for module 3 are:

- Recording panel discussion: one group member hands in a video file on behalf of the group. This video file contains a 5-minute presentation of each group member and a 15-minute Q&A. Each student is graded individually for this assignment.
- Slides of the assignment: that same group member also uploads a document that contains a compilation of the slides of the whole group.
- Self-assessment: after uploading the video and slides, the student has one week to complete a self-assessment. This self-assessment is only qualitative in nature (so the student does not provide a grade for their presentation) and the handing in of the self-assessment is graded pass-or-fail.
- Peer-feedback: after uploading the video and slides, the student has one week to give peer-feedback to 1 other student. The feedback is only qualitative in nature (so the student does not provide a grade on their peer's presentation) and the handing in of the peer-feedback is graded pass-or-fail.

The document, Instructions Panel Discussion and Rubrics for Students and Teacher Module 3 (see the supplement materials) show the instructions students received and the rubrics that the students and the lecturer used.

Finally, in module 4, students give a final presentation in their own group. Each group member selects a unique and new text from the required readings of their choice. The individual presentation should last for about 15 minutes and is followed by a 15-minute Q&A. Afterwards, students write an individual reflection report in which they look back on their progress.

In short, the student delivers the following materials:

Final assignment: Recording of their presentation and the Q&A (which the lecturer grades individually).

A final reflection report (an individual, pass-or-fail assignment on which the lecturer provides qualitative feedback).

The grading rubric the lecturer uses to assess the final presentation can be found in Table 3.5.1. The instructions for the final reflection report are included in Box 3.5.1. Students do not need to provide peer feedback.

Box 3.5.1

Final Reflection Report Module 4

You have now almost completed the presentation skills component of Contemporary Social and Political Philosophy. All that remains is a reflection report of no more than one page in which you look back on your final presentation and your development throughout the course. Reflecting on what you have learned helps you to retain this knowledge better.

Answer the following questions in your report:

1. Refer to the assessment form for the final presentation (see Table 3.5.1). To what extent do you think you have met the various criteria? What went well and where is there room for further development? You only need to highlight 3 to 5 criteria that are most relevant to you.
2. Review your own course learning goals. To what extent have you achieved them?

3. How do you look back on the process from February to now? What was it like to work on your action points? For all three questions, be specific. Link your reflections to actual situations, behaviors, thoughts, and/or feelings in recent months.

You do not need to provide peer feedback on a classmate's report.

Upload supporting documents:

- Your reflection report

All assignments, reflection reports, and peer feedback together form a portfolio. The reflection reports and peer feedback are all pass-or-fail. Students are required to receive a pass on all the reflection reports and peer feedback. For each assignment, the students receive a grade. Every module contains one such assignment, meaning that the student receives four grades in total. The final grade for the presentation skill component consists of a weighted average of those four assignments. The individual analysis of an inspirational speaker (module 1) counts for 10% towards the final grade. The group analysis of a presentation (module 2) also counts for 10% towards the final grade. The small presentation (module 3) counts for 20% and the final presentation (module 4) for 60%. The reason that each component has an increasing weight is because we wanted to give students the room to practice in a lower stakes environment first, slowly working towards their final higher stakes presentation.

Just like in the original setup, the students need to get a passing grade of a 5.5 or higher on the portfolio in order to pass for the course. Yet, with the introduction of the different sub-components, the stakes feel much lower compared to before. By the time the students reach the final presentation, they already had plenty of time to practice. For an overview of the course setup per module, including the deliverables, timeline, and the weight of each component, see Table 3.5.2.

The students must hand in all their work in FeedbackFruits, which they can access through links in the Canvas environment. The platform allows students to hand in videos of their presentations, written documents, as well as (PowerPoint) slides. It also offers various peer-feedback options. Throughout the course, students were required to provide individual peer-feedback on other people's learning goals and reflection reports (three times in total, for modules 1-3). To be clear, the lecturer did not attend any presentation 'live'. They assessed all the material through the platform.

The contact hours between the lecturer and the students all took place during tutorials scheduled over the course of one semester. In total, there were five tutorials, two during the first module and one at the beginning of each consecutive module. The goal of these tutorials was to explain the new assignments that were coming up and let students know what they expected. It was also an opportunity to practice different elements of the presentation. The very first tutorial, for example, is dedicated to the correct formulation of learning goals and action points. It also provided the students with tips on how to formulate proper (peer) feedback. Other tutorials were dedicated to text analysis and the analysis of power dynamics at play when giving a presentation. And in yet another tutorial, we did all kinds of small exercises. For instance, students were given a short text in Esperanto, which they were asked to read out loud in front of others. Even though no one spoke Esperanto (or rather, *because* no one spoke Esperanto), we could focus on one's posture, gestures, and intonation, and even experiment a little (putting your hands in your pocket, holding them in front of your chest and making little movements as you speak, etc.). The students worked on their portfolio assignments outside of the 'tutorial time', but we did keep a room available at a fixed moment each week, in case students wanted to make use of it (and we heard that some groups did make use of it).

The course was designed with Universal Design for Learning (UDL) principles in mind. The goal of UDL is to encourage learner agency, understood as "the capacity to actively participate in making choices in service of learning goals" (CAST, 2026). In practice, this meant that students were given various choices to shape their own learning process: they could pick their own inspirational speaker (module 1), could choose between different videos (module 2), or select their preferred text which they then had to present (modules 3 and 4). The lecturer also gave students the

option to indicate *one* other classmate with whom they'd like to be in a group, so that there would be at least one familiar face. Moreover, students were asked to formulate their own personalized learning goals (within the bounds of the more general assessment criteria) and work on those things that were important or meaningful to them. While one person may want to practice speaking at a slower pace, another person may benefit more by developing strategies for how to handle nerves. They were also invited to hand in their work in a variety of creative formats, beyond the 'typical' written format. One student made use of this and handed in a video-essay instead of a written report for module 1. The course provides multiple ways to engage with the material: by *listening* to the lectures, by *reading* the texts, and *presenting* (some of them) in turn. The benefit of UDL is that the educational design is flexible enough so that every student can adapt it to their own needs and preferences (regardless of whether you are autistic, in a wheelchair, a non-native Dutch speaker, shy, someone who is already very experienced or not experienced at all).

Finally, oral assessments are more prone to bias, due to the impossibility of anonymizing the presentations (Gerlt et al., 2023). We tried to mitigate bias in a few ways, for instance, by complementing presentations with anonymized

Table 3.5.2
Overview Course Setup

Module	Presentation skills component	Tutorial number	Date	Content of the tutorial	Deliverable portfolio (FeedbackFruits)	Deadline portfolio assignment	Weight
1	Form	1	31/1	Explanation of portfolio Explanation of personal learning goals and peer feedback	Self-reflection, learning objectives, and action points (INDIVIDUAL) Video analysis form (INDIVIDUAL) Peer feedback (INDIVIDUAL)	28/3 (21:00)	10%
		2	14/2	Analysis of text module 1	Nothing, but the skill that is being practiced is crucial for subsequent modules	x	x
2	Content and structure	3	28/2	Explanation and practice with content analysis presentation	Video analysis content (GROUP) Learning objectives and reflection report content (INDIVIDUAL) Peer feedback (INDIVIDUAL)	28/3 (21:00)	10%
3	Communication	4	17/4	Power dynamics based on text module 3	Recording of panel discussion (SUBMITTED AS A GROUP, BUT ASSESSED INDIVIDUALLY) Slides (SUBMITTED AS A GROUP, BUT ASSESSED INDIVIDUALLY) Self-reflection (INDIVIDUAL) Peer feedback (INDIVIDUAL; WITHIN GROUP)	2/5 (21:00)	20%
4	Everything together	5	2/5	Final presentation of a text from module 3 + 4	Recording of your presentation and Q&A (INDIVIDUAL) Self-reflection (INDIVIDUAL)	30/5 (21:00)	60%
							100%

written assignments. Besides, students needed to hand in recordings of their presentation(s), thereby allowing the lecturer to only grade a limited number of them at the time to avoid fatigue. Moreover, following Gerlt et al., (2023), self-reflection and actively questioning stereotypes may help as well, which is even an explicit component of the course.

Looking Back and Looking Ahead

Looking back, students who followed the course during the academic year of 2024-2025 did not receive significantly different grades than the ones who followed the course in the prior year. As such, students from the 2024-2025 cohort as well as the students who came before achieved the course learning goals. The biggest difference lies in how these goals are achieved. By listening to the students and incorporating literature on UDL, authentic assessment methods, and internal and peer feedback, the training and assessment of presentation skills now became a much more fulfilling process *for the students*, as evidenced by informal feedback from the students themselves. For example, they indicated that the course encouraged them to step outside of their comfort zone. They also shared that they gained insight into what is needed for an academic presentation and that they feel more confident in front of groups now. They liked that, by practicing their presentation skills, they also had to study the course material at the same time, thereby killing two birds with one stone. This feedback suggests that achieving the learning goals in this renewed way is experienced as much more fruitful than through a scary one-off presentation without having the room to experiment and grow.

At the same time, the lecturer and the students also identify room for improvement. For example, due to the various subcomponents that needed to be uploaded in FeedbackFruits (two assignments per module + a separate deadline for peer-feedback in each module), one loses oversight quite easily. This requires students and lecturers to keep a more extensive admin than usual.

Additionally, in the pilot version, it was simply assumed that students know how to collaborate. While the collaboration in most groups went well, this was certainly not the case for all. In the future, extra attention could be paid to collaboration. For instance, it would be good to let the students discuss the role each of them wants to fulfil in the group and let them reflect on that in their reflection report.

Something else that will be incorporated next year, is the opportunity for students to present in front of a bigger group, with the lecturer present. Some students indicated that they did not really get the chance to challenge themselves, because of the small-scale setup. So, even though they still achieved the learning goals, they personally wanted to take on this extra challenge, which will be provided in the upcoming edition(s).

Then, regarding the evaluation criterion of 'interaction' (see Table 3.5.1), it turned out that it was rather difficult for the lecturer to judge whether the student indeed considered a diverse audience in their presentation. For a future version of the course, the lecturer should either operationalize this criterion more or remove it altogether and include it as a requirement in the self-reflection.

Finally, FeedbackFruits has a lot of functionalities and works very intuitively. However, there are still some limitations. For example, if you create a 'peer-feedback module', the lecturer cannot give direct feedback on the student's work, only on the peer feedback *other* students have provided. And if you create a 'lecturer feedback module', then the students cannot give peer feedback as well. It would be inconvenient to ask the students to hand in the same work twice, once in every type of module, especially when uploading large video files. This led the lecturer to formulate their feedback and grades for module 3 in a separate Word-document and send it to every individual student by mail. This is something to consider when setting up the course in the future.

All in all, our current way of assessing presentation skills is an improvement compared to before but remains work in progress as well. We therefore mostly want to invite presentation skill lecturers in other programs to draw from our innovation what they need and adjust it to their own context.

3.6

Becoming Critical Writers Through Dialogic Feedback and Co-Teaching

Sam Passeport, Constanze Hopfner, and Jos Swanenberg

Introduction

How can lecturers support Master students with differing levels of background knowledge and skills in becoming critical writers in the field of Culture Studies? How can students engage with complex concepts and texts and begin to find their voices as writers? How can this process be aided through formative feedback and assessed with a final writing assignment? These are the challenges the lecturer team faced when designing the course Critical Culture Studies, a course all Master students of Culture Studies take in the first semester. We will describe the course and assessment design in this chapter, to exemplify good practices for the design of a writing assignment within the contexts of relational pedagogies (Gravett et al., 2024).

Course Design

Three-quarters of students joining the master's program in Culture Studies have completed their prior education in disciplines other than Culture Studies. Students bring with them personal and academic experiences and histories that shape their writing styles and aspirations as writers. Consequently, incoming students possess varying levels of conceptual and theoretical knowledge of Culture Studies, as well as differing levels of academic research and writing skills. To support an equitable start for all master's students, the first 10 days of the academic year focus on an intensive 6 ECTS(European Credit Transfer and Accumulation System)-course in "Critical Culture Studies". The master's program is set up in a hybrid format, with a small number of students who can follow the synchronous sessions online via Zoom, while the rest of the students are present on campus, together with the lecturers.

Course Learning Outcomes

By the end of Critical Culture Studies, students develop advanced knowledge of key concepts in Culture Studies, such as power, diversity, multimodality, orientalism, feminism, and affect. Through lectures and in-class close readings of foundational thinkers, they learn to interpret academic texts and connect key theories to contemporary societal issues. At the same time, students strengthen their skills in applying these concepts to real-world contexts. In workshops and group projects, they collaborate on presentations, podcasts, or clips that analyze present-day problems through a critical lens. These collective assignments prepare the ground for the final individual writing assignment (a paper), where each student expands his/her/their analysis with independent research and argumentation.

In sum, the course's overarching goals are to develop a strong foundation, a shared language, and a common reference frame among all students on which subsequent courses in the master's program build. Moreover, students are encouraged to draw from their lived experiences and histories as writers, and to strengthen their own voice and writing style. Finally, the course nurtures academic character building by fostering intellectual independence and critical thinking, in alignment with the character elements of the Tilburg Educational Profile (TEP) (Tilburg University, 2017b).

Course Schedule

The course Critical Culture Studies serves as the official kick-off into the master's program. In the first two weeks, students exclusively take this course before they start other courses. The course is organized into three parts:

- Part 1: In week one, students are introduced to the key theoretical concepts of Culture Studies through five key lectures and subsequent guided close reading sessions. Each day is centered around the crucial work of a leading scholar in the field and is led by a team of two lecturers. Lecturers follow a similar design for each session but also apply their own creativity and practical wisdom to engage students in discussing texts while maintaining the balance between standardization and lecturers' autonomy. Most lecturers use guided questions to prompt students to engage with texts.
- Part 2: At the end of the first week, students pick the theoretical concept they are most interested in and form groups of five to six students to work together on that concept. Groups are supervised by one of the lecturers who taught them about the concept in week one. Two workshops and one to two sessions with the group supervisor help students prepare for the group presentation. At the end of week 2, student groups present their presentations to each other and the lecturers.
- Part 3: After presenting, students choose a topic for the individual writing assignment (the paper) related to their group's concept. Through weekly synchronous writing circle sessions, the supervisors guide students in their writing process. In the sessions, the students and their supervisor discuss questions, each other's work, and peer and lecturer feedback. Outside of these sessions, students work on their paper independently.

Situating the Course Within Its Assessment Culture

The disciplinary assessment culture of the Cultural Studies program is uniquely positioned toward relationality and dialogues, in contrast to performative and measurement-oriented assessment practices that tend to position students as “objects of **assessment**” (Nieminen et al., 2025). The nature of discipline itself lends weight to such an approach. Cultural Studies invite sustained engagement with questions of positionality, identity, culture and environment, where objects of study are not treated as detached bodies of knowledge but as matters that call upon students' lived, embodied and affective experiences. This situates both students and lecturers within a co-shared learning space, orientating participants toward attentiveness, responsibility, and responsiveness to one another and to the world.

While grades and formal evaluation moments remain present, they do not singularly define how learning is recognized or valued. Instead, assessment is largely enacted through ongoing feedback talk, characterized as “contingent, episodic and dialogic interaction[s] between students and lecturers in the classroom” (Heron, Baker, et al., 2023, p. 172). These encounters are not always explicitly framed as assessment but are woven into the everyday pedagogical relations that shape students' engagement with the course (Gravett & Lygo-Baker, 2025).

Small class sizes and frequent co-teaching practices further cultivate this shared pedagogical space, in which dialogue is prioritized over transmission (Freire, 2020). Lecturers regularly invite embodied exercises, pause and listen, and legitimize affect and lived experience as ways of knowing. These micro-moments of connection accumulate into a pedagogy of care (Gannon et al., 2019; Noddings, 1984; Walker & Gleaves, 2016) that extends across physical, online and hybrid spaces, aligning with Biesta's (2021) conception of education as world-centered and supporting an authentically **formative assessment** culture.

Formal Assessment Moments

Students must complete two assessments for the course: first, a presentation they designed together as a group, which is graded with a collective pass-or-fail; secondly, an individual writing assignment which accounts for 100% of their final course grade.

Group Presentations

The assignment instructions for the group presentation are stated in Box 3.6.1.

Box 3.6.1

Instructions for the Groupwork

The group assignment consists of a presentation of the groupwork during the second week. The presentations take place on [date]. You will present your team's findings to your peers and lecturers. The presentation should address the topic and aim, the progress, and the outcome of the team assignment. The presentation is supported by visuals, for example, a PowerPoint presentation.

We will have approx. 15 minutes per presentation + questions, so try to limit the presentation to 10 minutes maximum.

The group assignments address a specific place: students find their chosen topic in their surroundings (on campus, online, in the city, in a museum, etc.), connect it to one of the theoretical concepts from class, and describe and analyze how they see the concept can be applied but also needs adapting to present times.

The presentation addresses:

- the 'problem': situate your topic and state the aim of your group project
- the concept and analysis: demonstrate how you applied the theoretical concept of choice in the context of 'here and now'
- the findings: discuss and conclude, by giving arguments based on your analysis

Using the Frayer Frame (Figure 3.6.1) might help order your thoughts.

Assessment consists of feedback to the presentation; this group assignment will not be graded but is assessed as pass-or-fail. This implies that it does not count for the final grade, but to pass this course, a pass for the presentation is required.

The coordinator does the assessments of the group assignment and receives input from other lecturers during the presentations. The group assignments address a specific place: students find their chosen topic in their surroundings (on campus, online, in the city, in a museum, etc.), connect it to one of the theoretical concepts from class, and describe and analyze how they see the concept can be applied but also needs adapting to present times.

The presentation addresses:

- the 'problem': situate your topic and state the aim of your group project
- the concept and analysis: demonstrate how you applied the theoretical concept of choice in the context of 'here and now'
- the findings: discuss and conclude, by giving arguments based on your analysis.

The assessment criteria are based on these topics: a clear problem statement, related to a theoretical concept from the literature (as presented in one of the lectures) situated in a present-day context, followed by an analysis of how the concept can be applied to understand this problem. All students in the group are present during the presentations; delegates present before class, and all students join the discussion with lecturers and other students after the presentations.

The procedure is as follows: teamwork is on a topic selected by yourself, which is connected to one of the lectures in the first week. Students are divided into teams by signing up on Canvas. You decide collectively what topic to work on, and how to apply the chosen concept to data found in the field (on campus, in the city, in a museum etc.).

Frayer Frame

To support students in creating the presentations, they are introduced to Frayer frames (see Figure 3.6.1). Frayer frames are templates students can fill out when learning about a new concept through reading, viewing, or listening to information. Each frame consists of four fields, one field each to describe the concept's definition, its characteristics, examples/models, and finally non-examples (see Note below Figure 3.6.1). Using Frayer Frames is a suggestion rather than an imposition. Several students report it as a helpful tool for clarifying concepts.

Definition	Characteristics
Examples/models	Non-examples

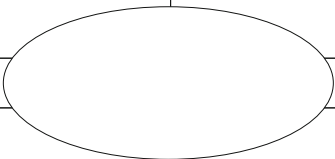


Figure 3.6.1

Frayer Frame

Note: Non-examples are illustrations of instances in which a concept (or definition) does not fit. Non-examples provide clarification of the concept's boundaries and can support avoiding misinterpretations.

Individual paper

The individual writing assignment is explained to students by the lecturer in a dedicated session and is conceived holistically to prompt students' self-discovery and engagement with concepts, rather than standardized products (Corbin et al., 2026). Additionally, they receive the following written assignment instructions (Box 3.6.1):

Box 3.6.1

Instructions for the Individual Writing Assignment

Write an individual paper on a topic discussed during the Critical Culture Studies course.

Apply one or more of the theoretical concepts addressed in the first five lectures. Alternatively, you can choose to elaborate further on the topic that you addressed in the team assignment for the group presentation on September 5. Show you master the subject of your choice by using the input from the lectures and the literature discussed during the course.

Develop a well-defined problem statement and show you that you can discuss this subject at an academic level. Make sure to develop a research question and a conceptualization (describe and argument how your topic relates to one of the theoretical concepts from class).

The length of the paper should be max. 3000 words, including references.

You need to apply the APA Style and Grammar Guidelines ([Links to a guide](#))

Level of permitted AI usage: AI-assisted editing (Level 2)

You are allowed to use GenAI tools only to review completed drafts for editing, including spelling, punctuation, grammar, syntax, and referencing. Submit the technology statement in which you state what GenAI tools are used and how they are used.

To clarify expectations for the writing assignment to students, to stimulate their self-assessment skills (Boud, 1995) and finally to streamline grading for lecturers, the supervisor and the instructional designer tailored a SOLO **rubric** (Table 3.6.1). SOLO is an acronym for Structure of Observed Learning Outcomes and was originally developed by Biggs and Collins (1982). During week one, one of the lecturers discussed rubric with students in depth. Students were given excerpts of papers and together discussed with SOLO level applied, to ensure they understood the quality criteria and standards.

Table 3.6.1

*SOLO – Structure of Observed Learning Outcomes, Adapted From Biggs and Collis (1982)**

Criteria	Pre-structural (1) Limited understanding	Unistructural (2) Isolated understanding	Multi-structural (3) Surface understanding of several parts	Relational (4) Integrated and coherent understanding	Extended Abstract (5) Insightful and transferable understanding
Points	0–2 points	3–4 points	5–6 points	7–8 points	9–10 points
1. Problem Statement & Research Question (25%)	No clear problem statement or research question; topic is unfocused or unrelated to course themes.	A basic problem statement or research question is present but vague, narrow, or disconnected from literature.	Problem statement and research question are clear, relevant, and linked to course themes; some connection to literature.	Problem statement is well-developed; research question is precise, situated in literature, and guides the analysis effectively.	Problem statement and research question demonstrate originality, critical framing, and potential for broader academic or societal insight.
2. Use of Theory & Literature (30%)	No or very limited use of course concepts or readings; concepts misapplied or undefined.	One theoretical concept or source is identified but not fully explained or applied.	Multiple concepts are correctly defined and supported with course readings and at least some external sources.	Concepts are well-defined, critically applied, and integrated with both course readings and independent literature.	Theoretical framework shows deep understanding, critical engagement, and extends or applies concepts in novel contexts.
3. Argumentation & Academic Writing (30%)	Writing is fragmented or descriptive only; lacks structure, coherence, or academic style.	Some structure and argument are present, but ideas remain isolated, underdeveloped, or weakly connected.	Clear structure (problem-evidence-analysis) with logical progression; arguments are supported but remain surface-level.	Argument is coherent, well-developed, and supported by evidence; writing demonstrates academic style and clear organization.	Argument is nuanced and original, with sophisticated connections and critical insights; writing is precise, cohesive, and engaging at an advanced academic level.
4. Referencing & Formal Presentation (15%)	Referencing absent, incorrect, or inconsistent; APA and formatting guidelines ignored.	Some attempt at APA referencing, but inconsistent and incomplete; limited adherence to format.	APA style generally followed with minor errors; structure and formatting mostly respected.	APA referencing consistently accurate; formatting is fully compliant with assignment guidelines.	Referencing is consistent and enhances the argument; formatting is precise.

* The rubric was explained to students in week one.

SOLO lends itself well to the context of the course in describing different levels of quality (or complexity) of integration of theoretical knowledge and writing skills. Whereas Bloom's taxonomy tend to focus on verbs that describe learning goals, the five levels of SOLO allow students to develop inner criteria for self-assessment and have a shared language to discuss with lecturers and peers about their work. As posited by Corbin et al. (2026): "through engagement with exemplars, students can encounter a range of valid essay forms and begin to discern what makes them effective. [...] Over time, this cultivates the capacity to write evaluatively [...] fram[ing] assessment as a space for reciprocal meaning-making" (p. 207). Employing the SOLO rubric in this way is also a form of dialogic feedback, engaging students and their lecturer in co-regulated learning (Allal, 2020; Heritage, 2016) and supporting them in building inner criteria for self-assessment.

The weighing of the criteria ensures that both content (developing depth by linking theory to a societal problem) and skills (argumentation and academic writing) are addressed. Together, these components prepare students for later writing tasks in the master's program, including the thesis. The performance standards are based on the SOLO – Structure of Observed Learning Outcomes (Biggs & Collis, 1982).

Preparing students for the individual writing assignment through group reading

To break the feeling of isolation in reading and writing, students are guided to discuss text segments together in plenary and small groups during week 1. First, the lecturer reads a text passage aloud. Then, using guiding questions, the lecturer and the students dissect the text together. Examples of guiding questions used are:

- What are some words that are in the way of understanding the text?
- What are the key ideas/concepts of this passage?
- What is difficult in this passage?

Students then implement this process of reading and reflecting with guided questions and discussing them together in smaller groups with another text passage. This process allowed them not only to build experiences with these academic skills but also realize the power of the community in generating knowledge. Students mention that reading and reflecting together allows them to go deeper into the theories than they could have on their own. After week 2, students continue diving into the theories within their writing circles (for 4 weeks), which positions them more confidently when starting their individual writing assignment. These circles allowed students to experience a process like thesis supervision (which will take place later in the year).

The red thread that connects all course elements: dialogic feedback and co-teaching

A team of cooperating academic lecturers delivers lectures in pairs, guiding student teams through collaborative assignments, and supervising writing circles when students begin working on their individual writing projects.

All these activities are dialogic, integrating formative assessment culture. In week 1, students engaged in various plenary and group dialogues, answering and asking questions, responding to the texts and concepts of the course and engaging with various invitations and informal feedback talk (Heron, Medland, et al., 2023) with their lecturers. In week 2, feedback was embedded throughout the process, with lecturers guiding the teamwork, peers contributing during discussions, and material (such as various rubrics and writing exemplars) that allowed students to develop inner feedback through comparisons (Nicol, 2021; Nicol & McCallum, 2021). Therefore, feedback was continuous and multi-layered: formative during lectures, workshops, and groupwork, and summative in the final essay.

Reflection on Challenges and Potential Adaptations for Future Course Iterations

To gather reflections and further improve the course, lecturers and students were asked to fill in (separate) digital course evaluation surveys. In total, nine out of eleven lecturers, as well as seven of the sixty students completed the surveys. In the following, we summarize the main points of feedback.

Several positive reflections were submitted. Lecturers noted that students appeared eager to learn, engaged, and motivated, both in online and in-class settings. This was especially true when students could connect theory with their personal lives or current topics. Lecturers appreciated meeting students before the other classes begin, having students from all tracks together, and for students to become familiar with each other and the lecturers, suggesting a relational culture was established (Gravett, 2024). Students appreciated this as well. In addition, students shared that the group reading sessions, guided by the lecturers, were helpful for understanding the theories. The sessions also successfully stimulated students to deeply engage with the reading materials. Each lecturer supervised a group of 5-6 students when working on the group presentation and on the individual writing assignment. Supervising was conducted both online and in person and this was evaluated positively. When observing the presentations, the lecturers were pleased with their groups' performances. Generally, students demonstrated interest, were proactive, well-prepared, or worked independently.

Lecturers and students also shared challenges they experienced and gave input into what can be improved in the course design. Several lecturers and students felt that the number of texts to read, both preparatory and in class, might have been too much.

After week one, it was somewhat unclear how to organize supervision for both lecturers and students, particularly in finding the right balance – allowing exploration and mistakes or giving clear instructions – especially given the limited time. Students felt that the decrease in supervision and guidance from lecturers throughout the course was too abrupt. Students would have appreciated more structured guidance throughout the whole course, not only in the first two weeks.

Furthermore, the interaction with online students could be improved. For example, online students would have preferred more breakout rooms in the first week, or other opportunities to connect with in-class students. In addition, it was sometimes forgotten to repeat questions and contributions from on-campus students into the microphone, meaning online students could not hear these. Also, in contrast to their fellow on-campus students, online students felt like they were not given the chance to go first when the group presentations took place. These are examples of how seemingly small moments can cause (unintentional) disconnections for online students.

Upon reflection, the teaching team also decided to take additional steps to further enhance assessments and the grading process. Before the next course iteration, it will be checked if the SOLO rubric is working as intended. To do this, the entire course team will discuss the meanings of the criteria and standards of the rubric. The lecturers will use the rubric to independently re-grade a sample of papers (from the first course iteration). The outcomes will be discussed among the team to ensure that lecturers are calibrated in terms of grading. During the next course iteration, lecturers will also consult the course coordinator to apply the four-eyes principle before determining the final grade of the paper in case a paper was assigned with a grade around the pass-or-fail **cut-off** point or if the paper received a grade above 8.

To improve the assessment of the group presentation, specific instructions on how to avoid free-riding will be added to the assignment description. Also, a specific policy on how to report issues occurring during group assignment will be stated. Furthermore, it will be ensured that all students take part in the group presentation (as opposed to allowing individual students to represent their group).

3.7

WIKI-Assignment in Psychopharmacology Medical Psychology

Paula Mommersteeg, Ya-Ping (Amy) Hsiao, and Leanne Prevc-Hekman

Course information

Psychopharmacology Medical Psychology is a core course in the master's program in Medical Psychology (6 ECTS, European Credit Transfer and Accumulation System). It integrates both theoretical content and skills training and is delivered by a team of nine academic staff, including a coordinator, lecturers, and assignment supervisors, with contributions from guest experts.

The course addresses the role of psychopharmacotherapy in managing psychiatric conditions as well as mental health complaints that often accompany somatic illnesses. Students study the fundamental principles of psychotropic drugs and their application in conditions such as depression, anxiety, sleep disorders, schizophrenia, dementia, pain, placebo, and substance abuse. Throughout the course, students develop their ability to read, interpret, and summarize scientific knowledge.

Several textbooks are available for the Psychopharmacology course. However, there are no books specifically focused on psychopharmacology within medical psychology. Nevertheless, medical psychologists encounter people with both a somatic illness and a mental health condition in their work. Because there is no suitable literature focused solely on medical psychology, students will work in groups, to combine information about mental health/wellbeing in patients with a somatic illness, connected to medication use.

Course learning goals and assessment design

This course has five learning goals and contributes to eight program **learning outcomes**.

Building on foundational knowledge of psychotropic drugs, this second-semester course deepens students' understanding of psychological (dys)functioning and the mechanisms underlying pharmacological treatments. It complements the pathology course on somatic conditions, the Clinical Neuropsychology course, which focuses on neurological disorders, and the Psychosomatic Models of Explanation course, which examines how psychological factors influence the development and worsening of somatic conditions.

Students further develop their ability to translate scientific knowledge for broader audiences through an in-depth online ‘WIKI’-assignment⁵⁵. A WIKI, in this case, is a website on a specific topic developed and peer-reviewed by a group of students. The finished product is an accessible website. Working in small groups, they research the combination of mental health complaints within a specific somatic condition, related to medication use and present their findings in an online accessible website format.

The Psychopharmacology course runs for 17 weeks in a hybrid format, combining 12 in-person lectures on pharmacodynamics, pharmacokinetics, and psychopharmacological treatments with 10 weeks of guided self-study devoted to the WIKI project. Students engage in literature review, synthesis, peer feedback, and revision, supported by instructions, online guidance, and at least one coaching meeting with the assignment supervisor. The WIKI assignment fosters skills in research, integration, and science communication.

Table 3.7.1 shows how each learning goal maps onto the program learning outcomes and how these are assessed through the two graded components: the final exam (multiple-choice questions/MCQs and constructed-response/CR questions) and the WIKI assignment.

Table 3.7.1
Course Assessment Overview

Upon successful completion of the course, students are able to:	Program Learning Outcome (PLO's)*	Exam with mixed questions (MCQs & CR)	Wiki assignment
1. describe and explain neurobiological, physiological, and psychological processes which form the basis for (psycho)pharmacological functioning (e.g., antidepressants, antipsychotics, sleep medication).	1.1 and 1.2	X	
2. describe neurobiological and physiological aspects of psychological or psychiatric conditions (e.g., depression, anxiety, ADHD) in the realm of (psycho)pharmacology.	1.1 and 1.2	X	
3. describe and explain how (psycho)pharmacological substances can affect psychological or neurobiological symptoms and conditions and their side effects	1.1 and 1.2	X	
4. create an online WIKI based upon current (scientific) knowledge, treatment options (e.g., Farmacotherapeutisch Kompas) which provides insight into the association between a somatic condition and psychological functioning, with a focus on relevant medication use. Making scientific information readable for a wide audience, including patients and their relatives, and healthcare providers.	2.3, 3.1, 4.1, 4.2, 5.2		X
5. create an online WIKI with a group, give peer-review feedback on other WIKI(s), and process peer-review feedback.	3.1, 4.2, 5.1, 5.2		X

*See Table 3.7.2 for Program Learning Outcomes

The course *Psychopharmacology* supports the development of key program learning outcomes (PLO's) by integrating biological, psychological, and clinical perspectives central to medical psychology (see Table 3.7.2). It provides students with foundational knowledge of neurobiological and physiological mechanisms underlying psychopharmacological treatments, linking these processes to psychological functioning and clinical practice. Through lectures on depression, anxiety, pain, dementia, and addiction, students learn to understand and evaluate pharmacological interventions within the broader context of chronic somatic and psychological comorbidity.

55 Note that only students can share, modify and add information to the WIKI posts. Visitors can only view the posts.

The course promotes interdisciplinary awareness by involving lecturers from medical psychology, psychiatry, and pharmacology, encouraging students to assess the contributions and limitations of related disciplines. Engagement with English-language scientific literature fosters familiarity with the terminology and culture of the medical field, enhancing students' ability to communicate effectively across professional boundaries.

A central element of the course is the group-based WIKI assignment, in which students synthesize and translate current scientific evidence into an accessible online reference on psychopharmacology and its application in clinical contexts. This assignment develops skills in critical appraisal, synthesis of complex information, and communication to both specialist and non-specialist audiences.

Finally, the course strongly emphasizes self-directed learning and continuous knowledge development. By combining independent study, peer collaboration, and evidence-based reasoning, students learn to integrate biological and psychological perspectives and to apply them critically within the evolving domain of medical psychology.

Table 3.7.2.

The Program Learning Outcomes (PLO's) Covered in This Course

1.1	Knowledge of medical psychology (including biological processes) related to theories, methods, and techniques used in scientific research and clinical practice.
1.2	Knowledge of epidemiology, biological processes, pathophysiology, and psychopharmacology of common—particularly chronic—physical conditions such as heart failure, stroke, COPD, cancer, diabetes, Alzheimer's disease, brain injury, and neurodegenerative disorders.
2.3	Ability to make an original contribution to the knowledge base within at least one subfield of medical psychology.
3.1	Awareness of other disciplines and/or stakeholders and the ability to assess their strengths and limitations.
4.1	Understanding of, and ability to navigate, the language (e.g., English) and culture of the medical profession in order to function effectively within the medical field, both with specialists and non-specialists.
4.2	Ability to communicate conclusions in the field of medical psychology—and the scientific knowledge, motivations, and considerations underlying them—clearly and unambiguously to audiences of specialists and non-specialists.
5.1	Ability to study independently in order to review and expand one's knowledge and to stay responsive to new developments in the field.
5.2	Ability to independently synthesize complex information or viewpoints from different perspectives to arrive at a well-reasoned, evidence-based opinion.

WIKI-assignment

The WIKI assignment was chosen because it actively supports the application and integration of knowledge central to psychopharmacology. While the written exam assesses individual understanding of core mechanisms and pharmacological principles, the WIKI requires students to connect these concepts to real-world contexts by researching and explaining how somatic and psychological conditions are related with attention for the use of (psycho)pharmaca.

The WIKI assignment requires students to collaboratively create an online reference resource for use by people with a medical condition and their peers, medical psychologists, researchers, and healthcare providers. Students work in groups with shared contribution, apply and integrate knowledge from psychology, pharmacology, and somatic disorders, and present scientifically grounded information in a format accessible to both professionals and patients. The assignment also involves adding value through personal insights supported by literature, providing peer-review feedback, and using peer feedback to improve the final WIKI.

By translating complex scientific information into clear, accessible language, students practice perspective-taking, deepen their understanding, and develop essential communication skills in medical psychology. Collaboration and

peer review foster critical thinking, reflection, and synthesis across different viewpoints. This active learning process reinforces the foundational knowledge and reasoning skills assessed in the written exam, ensuring that the WIKI both measures higher-order learning and supports strong exam preparation.

For this assignment, students work in small groups, each describing one somatic illness combined with a common (mental) health condition. Attention will be paid to medications prescribed for these conditions. To this end, students use the knowledge gained during lectures and from the required reading, supplemented with information from scientific publications, the internet, and books. Their findings will be published on our WIKI website⁵⁶. This WIKI will be updated throughout the course, and students receive (online) feedback from the lecturer and peer review from colleagues. Table 3.7.3 displays the weekly activities, **assessments**, and feedback.

One goal of this assignment is for students to ultimately consult WIKIs when they graduate as medical psychologists in practice. However, when these WIKIs are published online at the end of the course, they will be accessible to everyone: medical psychologists, physicians, and other healthcare professionals who work with people with somatic illnesses and mental health conditions, as well as patients and their friends and relatives. The WIKIs serve not only as learning materials but also as online references.

Table 3.7.3

Weekly WIKI Activities, Assessment and Feedback

Week	WIKI activity	Assessment & Feedback
1	Create groups of 4 and think about topics	
2	Read about the assignment and discuss potential topics	Submit group (Canvas)
3	Work on the assignment in groups (3 hours)	Submit topic (Canvas)
4-7	Work on the assignment in groups (3 hours)	
8	Work on the assignment in groups (3 hours) Coaching meeting with assignment supervisor	Formative feedback from the course coordinator
9-10	Work on the assignment in groups (3 hours)	
11	First complete version of WIKI	Submit first draft WIKI for peer review in Canvas
12	Individual peer-review of another WIKI	(formative) peer-feedback in Canvas
13	Process peer feedback and finalize WIKI	Process peer feedback
14	Work on the assignment in groups (3 hours)	Submit final WIKI assignment in Canvas
15-16	Supervisor WIKI assessment (summative), Canvas	
17	WIKI and response lecture: Launch of WIKI online	WIKI launch lecture

*Overall 30 hours per person are estimated for the WIKI assignment. About 3 hours per person per week during 10 weeks. This semester includes many 'empty' weeks due to National holidays.

Grading

The assessment workload is well aligned with the 6 ECTS (168 hours) study load. The exam (60%) assesses individual mastery of pharmacological concepts, with preparation time distributed across lectures and self-study. The WIKI assignment (40%) extends over roughly ten weeks, allowing manageable weekly progress (about 3 hours) supported by the group, lecturer feedback, and peer review. This gradual structure prevents overload and promotes consistent engagement. Together, both assessments balance theoretical knowledge with applied and collaborative learning, ensuring feasibility within the expected workload.

Since the exam and WIKI assignment assess different learning goals, they cannot compensate each other. Each assessment (exam and WIKI assignment) requires 6.0 to pass the course (therefore, for example, if students get a 9

⁵⁶ <https://sites.google.com/tilburguniversity.edu/tiupsychofarmacologie>

for the WIKI assignment, but a 5.0 for the exam, then students still need to resit the exam to pass the course. Plagiarism check and anti-free-riding measures are in place. Groups are expected to address unequal contributions internally first. If issues persist, they must be documented and reported to the course coordinator, who may request evidence of individual contributions (e.g., peer review) and take appropriate measures, such as grade adjustments or additional individual assignments, in line with institutional academic conduct policies (see Criterion 6, Appendix 3.7).

AI Index and risk mitigation:

Since 2024 students have been documenting their use of ChatGPT, while keeping in mind the following considerations:

As part of this online group assignment, you could choose to utilize an AI language model, e.g., ChatGPT or similar (subsequently referred to as 'ChatGPT' though more options are available), for assistance in generating content for your WIKI page. While ChatGPT can provide valuable insights and suggestions, it is important to keep the following considerations in mind:

1. **Supplement, Don't Substitute:** ChatGPT is a tool to supplement your research and creativity, not a substitute for your own analysis and understanding. Ensure that the content generated aligns with your group's collective knowledge and meets the assignment requirements.
2. **Critical Evaluation:** Carefully review and critically evaluate the output from ChatGPT. Verify information independently and assess the relevance and accuracy of the content before incorporating it into your WIKI page.
3. **Originality and Collaboration:** Maintain the originality of your group's work and ensure that all group members actively contribute to the creation of the WIKI page. Collaboration is key, and ChatGPT should be used to enhance your collective efforts.
4. **Cite Sources Appropriately:** If ChatGPT suggests information or ideas, make sure to appropriately cite sources. Accurately attribute content generated by ChatGPT to maintain academic integrity. When acknowledging the use of AI, provide a brief description of the tool and its role in the content creation process.
5. **Acknowledging AI:** In your WIKI page, acknowledge the use of ChatGPT by including a brief statement in the acknowledgments section. For example: "We acknowledge the use of ChatGPT, an AI language model, to assist in generating content for this wiki page".
6. **Describing AI Contribution:** Briefly describe the role of ChatGPT in the content creation process. Use a textbox when uploading the link to your WIKI assignment to describe if and how you used the AI language model. For instance: "ChatGPT was utilized to provide additional insights and suggestions, enhancing our research and analysis efforts".
7. **Lecturer's Guidance:** If you have any doubts or questions about the use of ChatGPT or the content generated, seek clarification from your lecturer. Your lecturer's guidance takes precedence in this assignment.

Remember that responsible and ethical use of AI tools is essential. By incorporating content generated by ChatGPT, you are responsible for ensuring the integrity of your work and adhering to academic standards.

Assessment criteria and grading processes

The WIKI-assignment is graded by the following criteria:

Appendix 3.7

Wiki Assignment Rubric

1	Description of the disorder Is the disorder described sufficiently clear? Do we understand what the disorder entails, how it is caused, what the prevalence/incidence is, and any other relevant characteristics? Is the pharmacological treatment, if applicable, clearly described?	2 points
2	Description of interaction with psychological disorder Is the psychological complaint clearly described? Has relevant and recent literature been used, largely based on PubMed publications? One point if the description is basic, only mentions standard relationships, is based on just one or two articles, or uses outdated literature. Medication is described only in general terms, without being tailored to the disorder. Two points if the description transcends textbook knowledge; this is knowledge not easily found on an average website. The content goes beyond what has been taught in lectures (so not just the monoamine hypothesis of depression). Includes a brief introduction to medication and its application to the disorder, preferably including overlaps or noted side effects. The text reads easily and clearly.	2 points
3	‘Open section’ in which students independently explore an angle This section is open to interpretation by the students. 0 points if barely any attention is given to this section, if it is unclear, or if it adds no value beyond the previous sections. 1 point if it aligns with what was discussed in lectures, is clearly based on (PubMed) literature, but is mostly a repetition of parts of the literature without an original contribution. 3 points if it is more advanced. Students clearly add their own contribution to the relationship between the disorder and the psychological component—something that did not previously exist. The text reads easily and clearly.	3 points
4	Layout, readability, structure, and formatting 0 points: unclear, poorly done, unclear who the WIKI is intended for, messy overall. 1 point: acceptable—font size and text blocks are fine. However, readability is reduced; many concepts are unexplained or difficult to read. Some clear issues with layout. 2 points: excellent—good font size (at least 10), clear headings, introduction, concluding section, supplemented where possible with tables/figures. Clear information appropriate for the target audience, written clearly and pleasantly. Language suitable for a broad audience, including relatives and people with the disorder.	2 points
5	Processing peer-review feedback Peer review: Requirements met for submitting the peer-review assignment, unless insufficient—see point 6. Grade adjustment if the authors do not address feedback comments that appear frequently and that the lecturer considers relevant.	1 point
6	Collaboration and (individually provided) peer review Grade adjustment if the lecturer has been clearly informed that collaboration is not going well, e.g., if one student is contributing significantly less than the rest of the group. Also possible grade adjustment if a student provided NO peer-review feedback or feedback that is clearly insufficient (barely any effort).	Grade adjustment if needed

The course team, consisting of the course coordinator and four co-lecturers, applies the same assessment criteria to ensure consistency and **fairness** in assessment. Each assignment is graded by one examiner because of the limited capacities. Previously, in case of uncertainty, a second examiner has occasionally evaluated another WIKI, but grades were in a similar range. Lecturers have experience with grading the WIKI's. In cases of uncertainty, co-lecturers can consult with the course coordinator, particularly when a student's grade is close to the pass-or-fail threshold or when assigning very high grades (9 or 10).

Reflections

Student interviews conducted in 2023 provided valuable insights into the learning experience. Their feedback informed the 2024/25 course redesign, which introduced additional supportive materials, structured WIKI supervision, and updated content (e.g., antipsychotics, shared decision-making). Both student comments and supervisor reflections continue to guide iterative improvements each year, ensuring alignment with learning needs and program goals. Some students invest a significant amount of time in the WIKI assignment. Over time, the course team has adjusted its weighting, increasing the WIKI assignment from 30% to 40% of the final grade. We believe this change better reflects the time and effort students dedicate to the task and results in a more balanced overall assessment.

Additionally, the course team emphasizes close monitoring of the group process to prevent free-riding. A separate passing requirement is set for the wiki assignment to ensure that all students demonstrate a minimum level of individual contribution and engagement. When concerns about unequal participation arise, they are addressed promptly in accordance with the course's free-riding policy, thereby safeguarding fairness and academic integrity in group-based assessment.

Several intentional compromises were made to balance learning outcomes with workload. Reducing the frequency of supervision increased student autonomy and encouraged ownership of the WIKI project. Replacing some traditional individual assignments with the WIKI shifted the focus toward collaborative, applied learning while maintaining rigorous engagement with psychopharmacology content. These adjustments improved the authenticity of the learning experience while keeping demands manageable for both students and staff.

For medical psychology students, even though they do not need to prescribe medications, it is essential to understand a patient's problems, the medications prescribed during treatment, and how to integrate this information into a coherent treatment plan. Working on the WIKI assignments can help students develop these skills by encouraging them to research clinical cases, synthesize complex information, and translate it into clear, structured guidance for practice.

3.8

Groupwork in Personnel Psychology

Djurre Holtrop and Lonneke Schellekens

Course Overview

In the course Personnel Psychology (6 ECTS, European Credit Transfer and Accumulation System), students learn about a wide variety of topics that concern people at work. Personnel psychology is concerned with the fit between the individual and the work environment. Having 'the right person in the right place' is thought to result in positive work results such as increased performance, retention and employee well-being. Accordingly, topics that will be addressed in the course are, among others, personnel recruitment and selection, performance evaluation and compensation, turnover, and talent development. In this course, students are exposed to theoretical/academic content, and they will need to apply their learnings in practice.

Course Learning Goals (CLGs)

The course directly supports the program learning outcomes by linking core knowledge of Work and Organizational Psychology to broader skills required to work in the field. First, students gain knowledge and understanding of central concepts, theories, and models (course learning goal 1 (CLG1)) and learn to critically engage with the scientific literature in the field (CLG2), thereby contributing to the program's emphasis on disciplinary knowledge. Through the application of the learned concepts in the analysis of organizational problems (CLG3) and the generation of evidence-based recommendations (CLG4), students develop the ability to apply theoretical paradigms to practical contexts and to critically assess the relevance of empirical findings. CLG4 further contributes to the program's focus on judgement, as students evaluate the usability of theoretical models in practice and reflect on their applicability in practice. Additionally, students are trained in communication skills through written and oral assignments (CLG5). Finally, all course components encourage an attitude of continuous learning and reflection on the connection between science and practice, aligning with the overarching program's learning skills goals.

Teaching, Supervision & Learning Activities

The course, Personnel Psychology, covers basic findings and new developments in the field of Personnel Psychology. The course consists of two parts: Theory and Challenge. These two parts are closely *connected to the* **assessments** used.

Theory

The theory part includes several class meetings. For these meetings, students are expected to study several papers carefully. The class meetings have a range of activities: class discussions, in-class assignments that students will work on with partners or small groups, guest lectures, and coaching for the practical.

The theory part will conclude with an individual blogpost assignment in the form of a “community of practice blog” and an exam about all readings. To prepare for the blog post, one class meeting is dedicated to ‘writing for a professional audience’. In this session, students practice translating knowledge convincingly in accessible writing. Many other sessions allow students to practice contextualizing knowledge, often discussing ‘how the readings are (ir)relevant for practice. The exam is a mix of multiple-choice and constructed-response questions, testing how well students can recall the content from the readings and the extent to which they can apply this knowledge to hypothetical scenarios.

Challenge

A substantial part of the course includes a practical assignment that involves taking on a consultant role in an organization. Students work in teams to solve a real-life challenge for an organization. The course coordinator selects several interesting challenges students can choose from. The organizations are both recruited through an organization called Masterchallenge (<https://masterchallenge.me/>) and through the course coordinator’s network. Masterchallenge collects a matching fee from the organizations that they have recruited for the course. This collaboration is vital for recruiting sufficient challenges on a yearly basis.

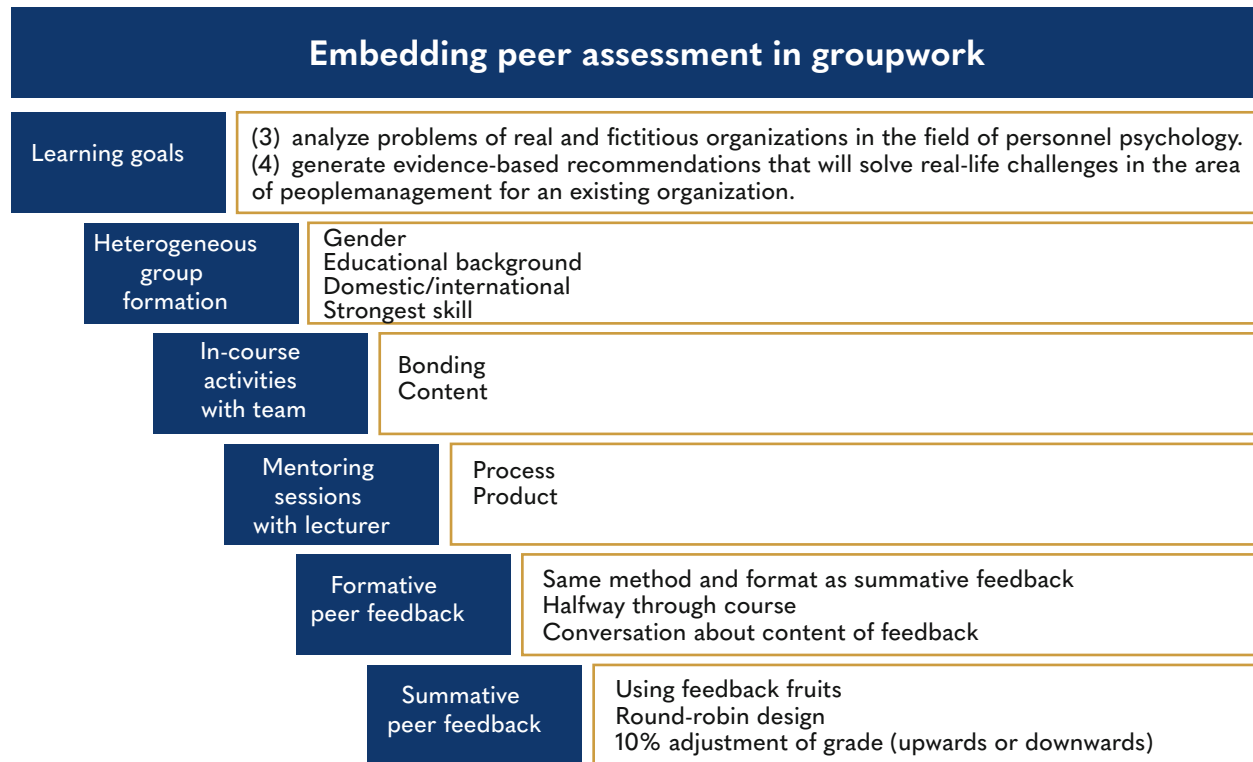
The task

The student team will come up with solutions, combining contextual and academic knowledge (gathered during the theory part of the course, and also with their own research effort). The student team will be conducting research/solving the challenge in close collaboration/consultation with the organization within a few weeks, which makes solving the challenge demanding. Organizations commit to regularly meeting with students when signing up.

The matching process

Heterogeneous teams are formed by the lecturer based on a brief survey. The groups are created to have a mix of gender, origin (international/domestic), educational backgrounds (psychology/non-psychology), and strongest skills (e.g., writing, presenting, coordinating group efforts). Team heterogeneity is particularly important to minimize exclusion of minority group members, because it ensures that most students have unique backgrounds (i.e., there is not really a majority). For example, most students will natively speak Dutch—when more than one student in the group is non-Dutch it is easier to keep communications in English and minimize exclusion. After team formation, the student teams select their top four challenges based on their collective preferences and submit a short-written motivation for each challenge. Next, the organizations rank the student teams based on the written motivation. Finally, the course coordinator matches the student teams and organizations based on their preferences.

Throughout the course, there will be work group meetings with the lecturer to discuss students’ progress and/or to help solve the problems students encountered. In these mentoring sessions, the course coordinator pays attention to both process (how collaboration within the team and with the organization is going) and content (the progress and quality of the work). Halfway through the course, students will give each other feedback and discuss it with one another, guided by the lecturer. This formative feedback moment serves three purposes: 1) students learn how their behavior will be assessed by their peers in the assessment phase, 2) students can ask questions about the scoring method to avoid ambiguity in the assessment phase, and 3) students have time to adjust their behavior to improve their final evaluations in the assessment phase. The same method and format are used at the end of the course, after handing in the report, where students summatively assess each other’s contribution. At the end of the course, the student team hands in a group final report and presents their solution to the company in the final session. See Figure 3.8.1 for an overview of the peer assessment procedures.

**Figure 3.8.1***An Overview of Peer Assessment Procedures*

Assessment Components

Assessment components:

- Exam (individual) – 35%
- Blogpost assignment (individual) – 25%
- Attendance (individual) – pass-or-fail
- Report on practical assignment (group with individual component) – 40%

The assessment weights are determined by the approximate amount of time students spend preparing the assessments. For the exam, this includes studying the course materials. For the blogpost, this includes reading a limited selection of course materials, additional papers, and writing the blog post in an appropriate manner. For the report, this includes working with the organization, group meetings, mentoring sessions, giving a presentation, and writing the report. The weights have been slightly adjusted over time and now appear to roughly capture the relative time investment required for each assessment.

Report and Presentation on Practical Assignment

At the end of the course, students hand in a group final report and present their solution to the company in the final session. The report will be graded by the lecturer with the grading **rubric** available on Canvas. In the final lecture, students will present their work to the contact person(s) of the challenge company. The lecturer, the other student teams and companies will be present. The presentation is not graded but serves an important role. It allows student teams to collect invaluable feedback on their consultancy report. Moreover, the presentation allows students to practice invaluable skills for any scientist practitioner: teamwork, leadership, and communication. To maximize the usefulness of the presentation, students are encouraged to schedule a debriefing session with the participating organization.

Report Individual Adjustment Grade

The report grades will be adjusted to students' individual contribution to the report. Effectively, after the report has been submitted, students will be asked to rate the contribution of themselves and of their fellow students. These evaluations will result in a maximum 10% change in the students' grades, depending on how the group members rated each other's contribution to the report.

The adjustment of individual grades is determined through the following procedure. First, both the group average and the individual student averages are calculated. Next, each student's deviation from the group average is computed using the formula: $(\text{individual average} + 10) / \text{team average}$. Finally, this deviation is capped at a maximum of 10 percent to ensure **fairness** and consistency in grade allocation.

Guidance and Feedback

Throughout the course, students receive intensive support from the lecturer, ensuring they never feel left on their own. From the very beginning, expectations are set out clearly so that everyone knows what is required. Active involvement is encouraged at all stages, with communication playing a central role in the learning process.

The lecturer maintains regular check-ins with each team, both through formal meetings and more informal conversations, to monitor progress and address emerging needs. Clear rules are established for giving and receiving peer feedback, helping students approach it constructively. To help students to develop feedback skills, halfway through the project a (formative) feedback session is scheduled. During this session, groups are guided by the lecturer as they give, receive, and collectively discuss feedback. The purpose of this exercise is to encourage individual reflection on areas of skill that require further improvement. In addition, students provide feedback on the same criteria that are later applied in the peer assessment at the conclusion of the project. This alignment contributes to the quality criterion of **transparency**, as it ensures that students are well-prepared for the final evaluation and understand the standards against which their performance will be assessed.

In addition, students can obtain feedback on their project in several ways:

- Students are encouraged to ask questions during the regular seminars and listen and discuss the issues of other teams.
- Students receive three comments or suggestions based on their presentation for the partner organization in the final week of this course.
- Students are encouraged to have regular conversations with their partner organization.

Assessment Criteria

Report on Practical Assignment

For this assignment, it is important to realize that the students work on very different projects. Some are conducting a literature review about employee retention strategies, others are conducting interviews and finding sources on how a worker's council can be best implemented, and others are studying how more women can be attracted for computer engineering positions. This means that the rubric for assessing the group reports and the instructions for writing the report need to be flexible but also encourage a somewhat comparable structure. As such, the course offers guidelines for content that can be included in a consultancy report, encouraging students to not simply describe the results of their work, but also describe the process of getting to the results and reflecting on the results. Students present their preliminary conclusions to the course lecturer and the organization towards the end of the course. After the presentation, they receive feedback from the course coordinator and the organization about the quality of their work with suggestions for improvement. Like the blog post, students are required to write the report in accessible language that is informative to the organization. The organization is not included in grading the report, because they do not (necessarily) have the correct qualifications and these are not appointed examiners by the Examination Board.

Students may include the following components in the report:

- An executive summary
- Introduction to the goal of the project (and how these fit in the bigger picture)
- What your team did to understand the challenge
- What you found out and how this matches best practices in Personnel Psychology
- How the presentation and other feedback from the organization informed your thinking
- What you advise the organization based on your findings and knowledge
- How the organization can achieve this proposed course of action
- Implications of the proposed course of action for the organization
- How the effectiveness of the proposed changes can be evaluated
- Limitations of your work / proposed course of action
- A figure/diagram that describes your diagnosis and/or advice in the context of the organization and the findings/recommendations of the project
- References (predominantly scientific; excluded from word limit)
- Table or matrix (1 page) which gives a clear breakdown of the role of each team member as well as their tasks and contributions to the overall project (excluded from word limit)

The rubric that is used to grade the groupwork paper:

<p>1. Introduction (10%) A general introduction in which the key contributions of the report are already clarified, this can be in the shape of...</p> <p>a. An executive summary (much like the abstract of a scientific paper) of the proposed course of action and how this was developed, and/or</p> <p>b. An introduction to the goal of the project (and how this fits in the bigger picture)</p>
<p>2. Diagnosis (20%) The quality of the diagnosis of the challenge:</p> <p>a. A description of the scope of the challenge</p> <p>b. What your team did to understand the scope of the challenge</p> <p>c. What you found out and how this matches best practices for personnel psychology</p>
<p>3. Advice or actions (20%) The quality of the advice or actions that followed the diagnosis:</p> <p>a. What you advise the organization based on your findings and knowledge</p> <p>b. How the organization can achieve this proposed course of action</p>
<p>4. Impact and limitations (10%) The quality of the considered consequences and limitations/risks are of the advice or actions:</p> <p>a. Implications of the advice or actions for the organization</p> <p>b. How the effectiveness of proposed changes can be evaluated</p> <p>c. Limitations of your work / proposed course of action</p>
<p>5. Main figure (10%) A figure/diagram that describes your diagnosis and / or advice in the context of the organization with the advice/actions of the project</p>
<p>6. Team member contributions (5%) Table or matrix (1 page) which gives a clear breakdown of the role of each team member as well as their tasks and contributions to the overall project (excluded from word limit)</p>
<p>7. Visual appeal (10%) The report is visually appealing:</p> <p>a. It includes an attractive cover page with a logo/image that also includes details on the author team</p> <p>b. It uses charts and graphics to visually illustrate your most important points (for example, consider including an infographic).</p> <p>c. It uses clear headers and lay-out features.</p> <p>d. Formatting is consistent and report feels as a whole.</p>
<p>8. Writing (15%)⁵⁷ The report is written with style:</p> <p>a. The style of writing is suitable for a consultancy report, it is professional and accessible</p> <p>b. Writing is clear and concise, and the report is well structured</p> <p>c. The report does not contain repetitive or redundant information</p> <p>d. The text does not include spelling, grammar, or typographical mistakes</p> <p>e. APA 7 guidelines were followed for references</p> <p>f. A reference list at the end (predominantly scientific; excluded from word limit)</p>

⁵⁷ This criterion may be updated based on the School's most recent AI policies in assessment.

Each criterion is graded based on a scale ranging from 4 to 10 points, using generic level descriptors.

10	9	8	7	6	5	4
Excellent	Very Good	Good	More than sufficient	Sufficient	Insufficient	Fail
Meets all descriptors at an extremely high level; extremely clear and well developed	Meets all descriptors at a high level; very clear and well developed	Meets most descriptors at a high level; clear and well developed	Meets all descriptors at a sufficient level; some lapses in clarity and development	Meets most descriptors at a sufficient level; sometimes unclear or under-developed	Does not meet descriptors at a sufficient level; often unclear or undeveloped	Incomplete; can be compensated

Criteria for Peer Feedback and Peer Assessment

Students rate their own and each other's contribution with seven items using a 5-point scale. Scale ranges from '1 = never' to '5 = as often as needed'. These **items** are based on peer-reviewed work by Oakley et al. (2004) and a practical guide by Hundey et al. (2020) but slightly adjusted to the Dutch and course context. Collectively, these criteria are meant to capture if students participated sufficiently in the project. To be clear, students do not rate the quality of each other's work, that is the job of the course coordinator. This system promotes active participation and punishes social loafing; It is intended to promote equal participation. In practice, most groups rate each other '5' on all criteria and no individual grade adjustment takes place.

The seven items are:

- 1 – Attendance: How often did the student attend team meetings?
- 2 – Preparation: How often did the student make a serious effort at assigned work before the team meetings?
- 3 – Responsibilities: How often did the student make a serious effort to fulfill his/her team role responsibilities on assignments?
- 4 – Communication: How often did the student notify a teammate if he/she was not able to attend a meeting or fulfill assigned tasks?
- 5 – Meeting activity: How often did the student attempt to make contributions in group meetings?
- 6 – Listening: How often did the student listen to his/her teammates' ideas and opinions respectfully and give them careful consideration?
- 7 – Equal effort: How often did the student cooperate with the group effort?

Reflections on Challenges and Adaptations

Using peer assessments to evaluate students' contributions to group projects has proven both valuable and demanding. One unexpected challenge was the amount of time required to manage the **formative assessments** in a class of around forty students. Scheduling and facilitating mid-course feedback moments took more effort than anticipated, particularly because we were designing a new system. Students generally adapted well to the technology used to facilitate the assessments, such as FeedbackFruits, and many appreciated the opportunity to provide structured feedback to one another. Nevertheless, the initial wording of the peer assessment questions caused confusion, as students interpreted them in different ways. Through iteration and feedback from both students and colleagues, these questions were refined to improve clarity and ensure the exercise was smoothly integrated into the course design.

Introducing peer assessment also required conscious trade-offs. Embedding additional assessment moments added to the teaching workload and reduced time available for other tasks. However, this design choice functioned as a

preventive intervention that substantially improved group dynamics, helping to reduce conflicts and the need for reactive “firefighting” later in the course. Furthermore, shifting the evaluation of contribution from an implicit character judgment to an explicit learning goal marked an important step in aligning assessment with desired outcomes. In this sense, peer assessments not only measured but also reinforced the behaviors necessary for effective teamwork.

The approach also raised questions of fairness, but in practice, the system enhanced perceptions of equity. Because the assessments are based on the principle that students must contribute sufficiently to receive recognition, contributions are rewarded in a transparent and systematic way. Peer feedback provided insights that would otherwise remain invisible to the lecturer, as students directly observe each other’s behavior during the project. These perspectives helped capture dimensions of collaboration that external lecturers could not easily assess. Moreover, the assessment criteria established clear norms for group behavior, functioning as concrete behavioral targets for students to strive toward. This clarity helped to set expectations, guide student conduct, and ensure that teamwork was not only evaluated but also actively scaffolded as part of the learning process.

3.9

Team Assignments in Strategic Consultancy

Koen van den Oever and Aswin van Oijen

Course Overview

Strategic Consultancy is a Master-level course offered within the Strategic Management program at Tilburg University. The course is taught in English and carries 6 ECTS, European Credit Transfer and Accumulation System (168 hours) of study load. It is positioned in the third study unit (January–March) and builds on core preparatory courses, such as Business-Level Strategy, Corporate-Level Strategy, International Strategy, Strategy Analytics, and Research Skills.

Course Learning Goals

The Strategic Consultancy course develops essential consulting skills while preparing students for real-world strategic management practice. The formal course learning goals, aligned with Bloom's taxonomy, and the explicit mapping of the course learning goals to the specific program **learning outcomes** are shown in the course assessment overview below. This overview is the school wide format for accounting for assessment **validity**.

These course learning goals are explicitly connected to the broader program learning outcomes of the MSc in Strategic Management. In this way, the Strategic Consultancy course does not stand alone but serves as a bridge between theoretical learning and professional practice, ensuring that students meet the knowledge, skills, and character requirements of the entire master's program.

Course code: 325087-M-6 Course name: Strategic Consultancy Course year: 2026-2027 Course coordinator: dr. A.A.C.J. van Oijen Teaching staff: dr. A.A.C.J. van Oijen, dr. K.F. van den Oever Coursework assessment types and percentages: <ul style="list-style-type: none"> • Midterm (35%) • Team assignments (65%) 				
Learning goals <i>On successful completion of this course, students will be able to:</i>	Intended learning outcomes	Midterm (35%) ☒ Pass required?	Team assignments (65%) ☒ Pass required?	Formative assessment (no grade, Feedback only)
1. Make an assessment of the key strategic issue an organization faces [evaluation] and devise a proposal for the approach to solve this issue [synthesis].	3, 4	X	X	Feedback is provided per team in each tutorial
2. Evaluate existing tools to deal with strategic issues and choose the ones that are appropriate for a particular strategic issue [evaluation].	4, 8	X	X	
3. Gather and analyze data from practice using strategic tools [analysis].	4, 6, 8		X	
4. Design and present a solution for an organization's key strategic issue [synthesis].	4, 7, 8		X	
5. Make an assessment of what the work of a consultant in practice involves, and reflect on own professional identity and development within a consulting context. [evaluation].	3, 5		X	

Teaching, Supervision & Learning Activities

The course is designed as an intensive seven-week cycle, often described as a “pressure cooker.” For example, already in the first course week students have to set up an intake meeting with their client as well as start preparing for an individual, written, digital exam, which is administered at the end of the second week. Other important milestones are submission of a proposal on how to deal with the client's issue, pitches of preliminary advice, and presentation and submission of the final advice. In general, students have a tutorial each week in which there is a deliverable that is formatively assessed. This phased approach mirrors professional consulting projects, where early scoping, iterative feedback, and final delivery are critical to success.

Formative feedback is embedded throughout the course. Coaching sessions, tutorial discussions, and trial runs provide low-stakes opportunities for improvement. Peer feedback during pitches allows students to learn from each other's approaches. Summative assessments consist of the midterm exam and the advice slide deck, but these are scaffolded by multiple preparatory submissions and trial presentations. Feedback is thus multi-layered: students receive insights from lecturers, peers, clients, and through reflection on their summative assessments, ensuring that learning is both continuous and contextual.

Assessment Components

The course employs both formative and summative assessment. Formative moments include trial runs and coaching feedback. Summative components are the written exam (35%) and the advice slide deck (65%). The heavy weighting of the advice slide deck is justified by its practical relevance and workload, as students must solve a real strategic problem. The exam ensures that all students have the necessary theoretical foundation.

Each component (exam and assignment) allows one resit opportunity, under the same format and criteria. For the written exam, a resit is scheduled after classes end and after the final advice slide deck submission. For the assignment, students may improve their original work but cannot re-engage the client. A resubmitted assignment is capped at a maximum grade of 6.0 (enough to pass). If students failed previously and re-enroll in another year, they are considered “repeaters” and must complete all assessments again under the new year’s structure.

Assignments are central to the Strategic Consultancy course because they bridge theoretical understanding with authentic practice. Unlike conventional written exams, these assignments replicate the structure of professional consultancy work, requiring students to engage with real clients, manage projects under time pressure, and deliver actionable solutions.

The first (ungraded) deliverable is a client proposal, generally prepared in the format of a PowerPoint presentation. This proposal includes the background of the client’s strategic issue, the specific research question to be addressed, the initial perspectives, the project plan with milestones, and the team’s credentials. The proposal directly addresses learning goals related to issue identification and the formulation of strategic approaches. It requires students to demonstrate evaluative and synthetic skills, while also practicing how to scope and communicate with professional stakeholders.

The core (graded) deliverable is the final advice slide deck, supported by detailed speaker notes. This document functions as the equivalent of a professional consultancy report, laying out the analysis, arguments, and recommended actions. This task integrates virtually all learning goals. It requires data analysis, tool application, structured problem solving, synthesis of findings, and the design of actionable recommendations. Moreover, the use of speaker notes ensures clarity and **transparency**: anyone reviewing the deck should be able to understand the argument even without attending the presentation.

While AI tools might be used to brainstorm or refine slide wording, the assignment demands client-specific data analysis, issue tree development, and hypothesis testing – activities that AI cannot perform without student input. Speaker notes provide an additional check, as they reveal the reasoning process and depth of reflection. To further safeguard authenticity, lecturers and clients evaluate the deck not only for content quality but also for contextual fit and professionalism.

Across all tasks, the course employs a multi-layered strategy to mitigate AI-related risks through (1) personalization, (2) oral verification, (3) reflective elements, and (4) artifact verification. First, personalization ensures each deliverable is tied to a unique client engagement, making generic or recycled AI outputs inadequate. Second, interim pitches and final presentations allow lecturers and clients to assess whether students can defend their work in real time. Third, reflective elements, including self- and peer assessments, require students to articulate their thought processes, which cannot be outsourced. Fourth, deliverables must reference specific data, interactions, and project milestones, serving as evidence of genuine consulting work. Together, these four approaches to mitigating AI-related risks resemble the Swiss cheese model of risk management: weaknesses or gaps in one layer of defense are offset by other layers, preventing a single point of failure from allowing the risk to materialize.

Assessment Criteria

The client proposal is not graded. Not grading reduces the stakes students are facing, which are already intense, as students have told us they have been hardly, if at all, exposed to working with a client on a strategic issue. For the assignment, we have included a list of criteria and guiding questions for grading in Table 3.9.1. Relative weights are assigned to different criteria based on how close they connect to the learning goals of the course. For instance, we deem the introduction to be important and give it a weight of 15%, given that it is mirrored in the first learning goal (assessment of the key strategic issue or need an organization faces). We also include a **rubric** in Appendix 3.9. Since a detailed rubric is used, the summative feedback is the grade per criterion and brief feedback per criterion.

Table 3.9.1
Criteria for Grading Strategic Consultancy Assignments

Topic	Guiding Questions
Introduction (15%)	<p>Are the situation, complication, and question clear? Are they logically connected?</p> <p>Do the situation and complication correctly reflect the client's strategic challenges and needs?</p> <p>Is the question useful for the client?</p> <p>Is answering the question feasible within the scope of the project?</p> <p>Has the intake with the client produced the right information for a successful identification of the strategic challenge and successful completion of the project in general?</p>
Advice (15%)	<p>Is the advice compelling? In other words, is it clear, to-the-point, convincing, sound, appealing, and inventive? Is the advice detailed enough or is it rather too abstract?</p> <p>Are important and complex concepts briefly explained?</p> <p>Does it answer the question that was presented in the Introduction? Is the advice useful for the client?</p> <p>Is the advice based on an evaluation of alternatives which have been properly identified?</p> <p>Is it based on a (possibly iterative) hypothesis?</p>
Argumentation (20%)	<p>Is the issue tree built up like a pyramid?</p> <p>Do the arguments at the consecutive levels answer the "why" question?</p> <p>Are there arguments in the boxes of the issue tree? Or just single words?</p> <p>Does it make strategic sense?</p> <p>Is it MECE (mutually exclusive and collectively exhaustive)?</p> <p>Is it supported by data and figures at the lowest level?</p>
Data and tools (15%)	<p>Have valuable data sources been used? Have the data been processed correctly?</p> <p>Have the appropriate strategic tools been selected? Have they been implemented properly?</p>
Recommendations (10%)	<p>Have sufficient recommendations been added on how to implement the advice?</p> <p>Are the recommendations ready to be implemented by the client or are the recommendations vague?</p> <p>Has a timeline been provided?</p>
Backup slides (7.5%)	<p>Have high-quality backup slides and appendices to address follow-up questions from the client been added?</p> <p>Do the backup slides and appendices indeed contain information that is too detailed or not vital enough to be included in the main slides and text?</p> <p>Is there an appropriate number of back-up slides (not too few, not too many)?</p>
Structure and form (10%)	<p>Are all the main components of the advice present?</p> <p>Have they been put in the correct order?</p> <p>Does the structure make it easy to quickly find specific parts of the advice in the presentation?</p> <p>Do the different slides logically follow one another?</p> <p>Are the main elements balanced properly?</p> <p>Have active titles been used? Do they entail the main message of the slide?</p> <p>Have speaker's notes that would fully allow a third party to do the presentation been added to all slides?</p> <p>Is the information presented in an attractive way?</p> <p>Is the language correct (e.g., can an audience lacking prior understanding of strategic management concepts understand the presentation)?</p> <p>Have tables and figures been used to further support the presentation/report?</p> <p>Have all the data, tools, and (if and when useful) theory been referred to correctly?</p>
Professionalism (7.5%)	<p>Has the team fulfilled the minimum requirement for the assignment?</p> <p>Has the team acted in a professional manner throughout the process, both towards the client and the lecturers?</p> <p>How satisfied is the client with the advice?</p> <p>What is the client's view of the team's degree of professionalism?</p> <p>Would the client again hire a similar student consulting team, when faced with another strategic challenge?</p> <p>Have all team members filled out the self and peer assessment? What adjustments of the individual grades are warranted?</p>

Reflections on Challenges and Adaptations

When we designed the course, we started with two team assignments without any individual assessment. Moving from two to one team assignment had important benefits and drawbacks. With two assignments, students had opportunities to learn from one assignment and put into practice the experiences and knowledge they learnt in the first assignment. However, given that the course lasts 7 weeks, students had merely 2 to 3 weeks to work on the assignment, which was much too challenging for them. However, working with one assignment does imply that students do not receive summative assessment halfway through the course, which for a portion of students is significantly more useful in changing their behavior (mostly increasing their effort), which now, for these students, only occurs after getting the grade for their assignment, forcing them to do a resit instead. From the lecturer's side, having one assignment reduces the amount of coordination substantially, as the number of assignments that need to be 'found' is halved.

We also adapted the course based on what we learnt from the clients, an additional benefit of working with authentic assignments. This ranges from slide design to how to conduct an intake with a client. In general, we have received a lot of support from students who took the course as well as the clients they helped. Clients frequently decide to participate year after year with different assignments. They use the input given by students to write white papers, make strategic decisions and/or to scout new talent. Students frequently come back after the course as they seek to give back by bringing in client assignments from the employers they started to work for after graduation.

The rubric is highly effective for grading assignments that present markedly different challenges, promoting greater objectivity and transparency across assessments. It is also particularly useful in coaching sessions, where each session may focus on specific criteria (e.g., the introduction). The weighting of criteria further enhances the grading process; however, there are instances where the overall grade should reflect a failing performance. In such cases, the scoring per criterion may need adjustment, as the current rubric can be too lenient on lower-end performances.

Using authentic assignments supports our multi-layered AI-related risk reduction strategy. Because students must prepare advice and present it orally to a client, examiners can directly observe their understanding, particularly during the follow-up Q&A. While the Q&A is not intended to assess verbal representation of the content, it reveals the depth of students' engagement beyond their final recommendations. It also provides an opportunity for students to demonstrate additional considerations they weighed but ultimately chose not to implement, maintaining focus on their final advice.

Appendix 3-9
Grading Rubric for the Team Assignment

5 or lower – fail	6 – Sufficient	7 – Good	8 – Very Good	9-10 – Excellent
Introduction (15%)				
<ul style="list-style-type: none"> The Situation, Complication, Question (SCQ) is not a critical reflection of the client's challenge, and/or students do not go beyond the initial briefing of the client No potential to actually resolve an issue the client is facing. Situation, complication, and/or question are unclear and/or not clearly distinguishable Situation, complication, and/or question are overlapping largely. Situation, complication, and/or question do not follow logically from each other Question is too simple, too broad, or nonsensical given the background information. 	<ul style="list-style-type: none"> The SCQ shows barely any critical reflection of the client's challenge, and/or students barely go beyond the initial briefing of the client Does not fully clarify how the SCQ contributes to addressing a real-life problem relevant to the client. Situation, complication, and/or question are sufficiently described and clearly distinguishable Some overlap remains between the distinct elements of the SCQ The connection between the elements of the SCQ could be significantly improved. A minimum level of ambition is set 	<ul style="list-style-type: none"> The SCQ shows an adequate critical reflection of the client's challenge, and/or students go beyond the initial briefing of the client Clarifies how the SCQ contributes to addressing a real-life problem relevant to the client. Situation, complication, and/or question are adequately described and clearly distinguishable Near zero overlap remains between the distinct elements of the SCQ The connection between the elements of the SCQ is adequate A level of ambition is set adequate for the MSc level 	<ul style="list-style-type: none"> The SCQ is a good critical reflection of the client's challenge, and/or students beyond the initial briefing of the client Crystal-clear how the SCQ contributes to addressing a real-life problem relevant to the client. Situation, complication, and/or question are well described and clearly distinguishable No overlap between the distinct elements of the SCQ The connection between the elements of the SCQ is excellent. The story is coherent and logical. A level of ambition is set far above the MSc level, given the practical constraints of the assignment The description of the SCQ is of a top-tier consultancy level. 	<ul style="list-style-type: none"> The SCQ is an outstanding critical reflection of the client's challenge, and/or students well beyond the initial briefing of the client Crystal-clear how the SCQ contributes to addressing a real-life problem relevant to the client. Situation, complication, and/or question are very well described and clearly distinguishable No overlap between the distinct elements of the SCQ The connection between the elements of the SCQ is excellent. The story is coherent and logical. A level of ambition is set far above the MSc level, given the practical constraints of the assignment The description of the SCQ is of a top-tier consultancy level.
Advice (15%)				
<ul style="list-style-type: none"> The advice is not an answer to the question. Major issues mentioned in the SCQ remain and/or the advice contradicts information given in the SCQ The advice is (too) straightforward in the sense that the client probably already knows/researched this. The advice is unclear and not specific 	<ul style="list-style-type: none"> Advice that gives an answer to the question. Some significant issues mentioned in the SCQ remain and/or the advice partly aligns with information given in the SCQ The advice is (too) straightforward in the sense that the client probably already knows/researched this. The advice is unclear or not specific 	<ul style="list-style-type: none"> Advice that gives an answer to the question. No significant issues mentioned in the SCQ remain and/or the advice aligns with information given in the SCQ The advice has some original components/insights new to the client The advice is to a large extent clear and specific 	<ul style="list-style-type: none"> Advice that gives an answer to the question. No issues mentioned in the SCQ remain and/or the advice aligns very well with information given in the SCQ The advice has some original components/insights new to the client The advice is clear and specific 	<ul style="list-style-type: none"> Advice that gives an answer to the question. No issues mentioned in the SCQ remain and/or the advice aligns very well with information given in the SCQ The advice is very clearly original and new to the client The description of the advice is of a top-tier consultancy level. The advice is clear and specific

5 or lower – fail	6 – Sufficient	7 – Good	8 – Very Good	9-10 – Excellent
Argumentation (20%)				
<ul style="list-style-type: none"> The argumentation is not mutually exclusive nor collectively exhaustive or contains major mistakes in both of the elements. The argumentation structure does not adhere to the pyramid principle. Lower level arguments do not support the preceding level. No or hardly any relevant (academic/practitioners) sources used. Reveals significant lack of understanding of sources. Argumentation is vague and hard to follow. The argumentation is not consistent throughout the project. The argumentation is not applied to the client. 	<ul style="list-style-type: none"> The argumentation contains still major issues in either mutual exclusivity or collective exhaustiveness. Subsequent levels in the issue-tree are not always an answer to the 'why' question. Some relevant (academic/practitioners) sources used. Limited understanding of sources. Argumentation is not fully clear and some elements are hard to follow. The argumentation is somewhat consistent throughout the project. The argumentation is applied to the client. 	<ul style="list-style-type: none"> The argumentation contains still minor issues in mutual exclusivity and collective exhaustiveness. Subsequent levels in the issue-tree are mostly an answer to the 'why' question. Relevant (academic/practitioners) sources used. Proper understanding of sources. Argumentation is clear and relatively easy to follow. The argumentation is consistent throughout the project. The argumentation is applied to the client. 	<ul style="list-style-type: none"> The argumentation only contains minor mistakes in either mutual exclusivity or collective exhaustiveness. The argumentation fully adheres to the pyramid principle; the lower level arguments provide reasoning for the preceding level. Very relevant (academic/practitioners) sources used. Very good understanding of sources. No sources included that provide no added value. Argumentation is fully clear and easy to follow. The argumentation is consistent throughout the project. The argumentation is applied to the client. 	<ul style="list-style-type: none"> The argumentation is mutually exclusive and collectively exhaustive. No aspects are missing to support the advice and the arguments do not overlap. The argumentation fully adheres to the pyramid principle; the lower level arguments provide reasoning for the preceding level. Very relevant (academic/practitioners) sources used. Excellent understanding of sources. No sources included that provide no added value. Argumentation is fully clear and easy to follow. The argumentation is consistent throughout the project. The argumentation is applied to the client.
Data & Tools (15%)				
<ul style="list-style-type: none"> No data is used to support the argumentation. No tools or used, or mainly inappropriate tools are selected to address the question. Tools contain evident logical errors or omissions that prevent reliable conclusions. 	<ul style="list-style-type: none"> Some arguments are supported by data. Used tools are partly suited to address the question or support the advice Tools contain some logical errors or omissions that prevent reliable conclusions. 	<ul style="list-style-type: none"> Most arguments are supported by data. Used tools are appropriate to address the question or support the advice. Tools contain few logical errors or omissions that prevent reliable conclusions. 	<ul style="list-style-type: none"> Most arguments are supported by high quality data. Used tools are well suited to address the question or support the advice. Students display some ability to recombine elements of different tools and/or adjust existing tools to more appropriate address the clients' question. Tools contain no logical errors or omissions that prevent reliable conclusions. 	<ul style="list-style-type: none"> All arguments are supported by high quality data. Used tools are very well suited to address the question or support the advice. Students display an ability to recombine elements of different tools and/or adjust existing tools to more appropriate address the clients' question. Tools contain no logical errors or omissions that prevent reliable conclusions.

5 or lower – fail	6 – Sufficient	7 – Good	8 – Very Good	9-10 – Excellent
Recommendations (10%)				
<ul style="list-style-type: none"> Not, or insufficiently, connected to the advice. Poor recommendations that are unlikely to be implementable by the client. Provides no clear guidance what the client should do next. 	<ul style="list-style-type: none"> Somewhat connected to the advice. Standardized and/or mechanical presentation of recommendations. Provides little guidance what the client should do next. 	<ul style="list-style-type: none"> Recommendations clearly tie in with the rest of the advice. Recommendations are clear and actionable. Shows awareness of key issues for the client and presents a clear roadmap for next steps. 	<ul style="list-style-type: none"> The reader immediately comprehends how the recommendations and advice are related. Recommendations are self-explanatory. They will clearly aid the client in implementing the advice. Shows full understanding of key issues for the client and the surrounding environment. 	<ul style="list-style-type: none"> The reader immediately comprehends how the recommendations and advice are related. Recommendations are self-explanatory, non-obvious, and provide clear value-added by themselves. They will clearly aid the client in implementing the advice. Shows full understanding of key issues for the client and the surrounding environment.
Back-up (7.5%)				
<ul style="list-style-type: none"> No back-up slides are included Back-up slides are of poor quality The back-up slides are insufficiently supported. The back-up slides are not or barely relevant for the question, client and section 	<ul style="list-style-type: none"> Back-up slides are included Back-up slides are of sufficient quality The back-up slides are sufficiently supported The back-up slides are somewhat relevant for the question, client and section 	<ul style="list-style-type: none"> Back-up slides are included Back-up slides are of adequate quality The back-up slides are adequately supported. The back-up slides are relevant for the question, client and section 	<ul style="list-style-type: none"> Back-up slides are included Back-up slides are of a good quality The back-up slides are well supported The back-up slides are relevant for the question, client and section. 	<ul style="list-style-type: none"> Back-up slides are included Back-up slides are of an excellent quality The back-up slides are very well supported. The back-up slides are highly relevant for the question, client and section.

5 or lower – fail	6 – Sufficient	7 – Good	8 – Very Good	9-10 – Excellent
Structure and Form (10%)				
<ul style="list-style-type: none"> • Speaker notes are not included, or are of insufficient quality. • The prescribed structure during the course has not or barely been followed. • It is difficult or not possible to distinguish between different elements of the report. • There are a plethora of spelling/grammar mistakes which give an unprofessional impression of the report. • No care has been exercised in visually editing the report up to professional standards (e.g., layout of tables and figures, referencing). Would be unacceptable in a professional consultancy setting or for presentation to a client. • Title does not represent the content of the assignment. • The language used is inappropriate for the audience of the report. • No active titling has been used, or major flaws in the majority of the slide titles. 	<ul style="list-style-type: none"> • Most slides are accompanied by speaker notes of adequate quality • Prescribed structure has generally been followed. • Some different elements of the report are overlapping. • There are several spelling/grammar mistakes which are detrimental to the professional appearance of the report. • Some care has been exercised in visually editing the report up to professional standards (e.g., layout of tables and figures, referencing). Would not be to the author's credit in a professional setting, but might be acceptable for internal use in an organization. • Title suits the content of the assignment • The language used is suited for the audience of the report. • Active titling is used for the majority of the slides, but they are no integral part of an overall coherent story, and/or the titling does not cover all content of individual slides. 	<ul style="list-style-type: none"> • Speaker notes are of good quality and carefully explain the information contained on the slide. • Prescribed structure has been correctly implemented. • Barely any different elements of the report are overlapping. • There are barely any spelling/grammar mistakes. • Adequate care has been exercised in visually editing the report up to professional standards (e.g., layout of tables and figures, referencing). Would be suited in a professional setting, although not a distinctive feature. • Title suits the content of the assignment • The language used is well suited for the audience of the report. • Active titling is used for the majority of the slides, there is an overall coherent story, and/or the titling matched most of the content of individual slides. 	<ul style="list-style-type: none"> • Speaker notes are of very good quality and allow a novice external reader to fully understand the content of slides. • Prescribed structure has been correctly implemented. • No elements of the report are overlapping. • There are no spelling/grammar mistakes. • Much care has been exercised in visually editing the report up to professional standards (e.g., layout of tables and figures, referencing). Would be well suited in a professional setting, and provides a distinctive feature over other reports. • Title suits the content of the assignment • The language used is very well suited for the audience of the report. • Each slide contains very good active titling that represents the individual slide but is also part of a larger coherent story, albeit some minor gaps in the story remain. 	<ul style="list-style-type: none"> • Speaker notes are of the highest quality and fully illuminate information contained on the slides, without introducing any non-essential information. • Prescribed structure has been correctly implemented. • No elements of the report are overlapping. • There are no spelling/grammar mistakes. The report has been eloquently written. • Extensive care has been exercised in visually editing the report up to professional standards (e.g., layout of tables and figures, referencing). Would be well suited in a professional setting, and provides a distinctive feature over other reports. Would need little to no editing to represent a top-tier consultancy firm. • Title suits the content of the assignment • The language used is very well suited for the audience of the report. • Each slide contains excellent active titling that represents the individual slide but is also part of a larger coherent story.

5 or lower – fail	6 – Sufficient	7 – Good	8 – Very Good	9-10 – Excellent
Professionalism (7.5%)				
<ul style="list-style-type: none"> Student team has been unable or unwilling to take meaningful initiatives or to advance own ideas, yet has largely ignored suggestions for improvement. Students fail to balance demands from the university and the client. Student team solely blames external causes for issues, without any critical self-reflection. Student team comes to meetings with lecturers unprepared, with unclear questions, or with a completely changed setup right before a meeting. The team did not fulfilled all formal requirements. The client was dissatisfied with the team's performance in terms of content and process. 	<ul style="list-style-type: none"> Has required extensive coaching regarding all aspects of the assignment, but has shown a general willingness to work, to accept guidance and suggestions, and to learn. Student team lacks a critical attitude towards the client and provided information. Student team comes to meetings with lecturers prepared. The team has fulfilled all formal requirements (e.g., deadlines.) The client was somewhat dissatisfied with the team's performance in terms of content or process. 	<ul style="list-style-type: none"> Has required a normal level of coaching, but student team has also displayed own initiatives. Student team has been willing to accept advice and suggestions, but has during discussions also been willing and able to defend own choices. Student is open to information and suggestions from the client, but also maintains a professional and critical consultancy attitude. The team has fulfilled all formal requirements (e.g., deadlines.) The client was somewhat satisfied with the team's performance in terms of content and process. 	<ul style="list-style-type: none"> Student team has conducted the bulk of the work independently. A large proportion of the assignment reflects the student team's own thinking and initiatives. Has kept the lecturers well informed of plans and progress during the meetings. Has taken the initiative to raise specific questions for discussion. The team has fulfilled all formal requirements (e.g., deadlines.) The client was satisfied with the team's performance in terms of content and process. 	<ul style="list-style-type: none"> Student team can fairly take all credit for an original and high-quality assignment. Student team comes to meetings with lecturers very well prepared with high quality questions. The team has fulfilled all formal requirements (e.g., deadlines.) The client was highly satisfied with the team's performance in terms of content and process.

3.10

Simulation Design Case: Humanitarian Response Simulation Exercise

Antonio Massella and Kerry Forrester

Course Overview

The Advanced Project Management (APM) course forms a central component of the International Bachelor in Global Management of Social Issues (Department of Organization Studies, TSB) at Tilburg University. It is a core offering that integrates both content and immersive, skill-based learning. The course builds on knowledge and insights from earlier courses, such as Introduction to Organization Studies, International Organizations, and Financial and Project Management. This simulation design case focuses on one component of the APM course: the Humanitarian Response Simulation Exercise.

The simulation exercise is a fundamental component of the course where students apply the theoretical knowledge gained in the APM course through a dynamic, immersive simulation exercise (Simex). The team involved in the simex consists of a Lead Facilitator with professional experience in international disaster management and humanitarian response as well as academic staff, lecturers, and PhD students who are involved in the one-and-a-half-day exercise. As a pedagogical tool, the Simex deepens theoretical understanding while reinforcing practical competencies in project and financial management, stakeholder engagement, and intra/inter-organizational coordination. This exercise is designed to expose students to the real-world complexities of humanitarian coordination under crisis conditions, based on training and assessment methodologies used by institutions such as the United Nations, Red Cross, other international agencies, as well as multi-stakeholder humanitarian country teams.

Simulation Overview

Simex is designed to replicate the complex, dynamic environment of an international humanitarian response to a large-scale disaster. The scenario focuses on a rapid-onset humanitarian crisis in Timor-Leste, triggered by a cyclone that causes severe flooding across the country. The disaster resulted in widespread impact in the capital city, Dili, along with flash floods and a major landslide in the Liquica district.

The entire simulation unfolds within a single day, but it represents a two-week period of real-world crisis response. Participants are assigned different roles reflecting the structure of a typical international humanitarian operation, such as national government officials, United Nations representatives, Non-Governmental Organizations (NGO) staff, emergency responders, and cluster (or sector) coordinators. Each participant is required to react to events and make decisions according to their assigned role, engaging in distinct activities and responsibilities.

Throughout the exercise, participants receive “injects”, timely pieces of information delivered via email or other communication channels, that simulate real-world updates, reports, and directives from the field. These injects provide evolving details about the situation, present new challenges, and require participants to make informed decisions, coordinate actions, and communicate effectively under pressure.

Set in a facilitated, high-stress environment characterized by uncertainty and rapid change, the simulation mirrors the realities of a wicked problem context, where essential decisions and coordinated actions must be taken quickly in response to an evolving humanitarian emergency.

The simulation methodology is modelled on approaches used in the training and preparedness of professional disaster managers and humanitarian staff from organizations such as the United Nations, international and national NGOs, and emergency response teams.

Course Learning Goals and Program Alignment

The general aim of the course is to deepen students’ knowledge and understanding of project management especially applied to organizing for wicked problems in which multiple stakeholders are involved and need to be managed. After completing this course, students can explain what multi-organizational projects are and who the key stakeholders are. They apply project management knowledge in assignments and simulations. They also analyze challenging, real-world wicked problems.

The Simex aims to consolidate and assess the students’ ability to navigate the project management cycle in a fast-paced, uncertain, and high-stakes environment. By role-playing key actors such as heads of UN agencies, NGO leads, and government officials, students are challenged to apply their knowledge in a simulated emergency context. In their role-playing capacity, individual students are required to join one or more temporary networks, thematic or cluster/sector groups, or stakeholder consortia (that exist in reality) as part of the international humanitarian architecture manifested at the country level. They are required to manage coordination meetings, respond to cascading injections of information, and produce outputs or deliverables such as stakeholder maps, situation reports, budgeted cluster implementation plans and appeal documents. Various individuals and different groups are required to address a variety of needs or predicaments, at different time; therefore there is a need for clear and regular communication, as well as coordination, between networks to ensure maximal efficiency, effectiveness in resource use, and to avoid duplication of effort and wastage as far as possible, as is the case in real-life largescale humanitarian response operations.

An example illustrates this point: while all students respond to one evolving crisis scenario, they do not work on the same problem: for example, three to four nominated students join a rapid multi-cluster needs-**assessment** “helicopter” mission to conduct key-informant interviews and focus-group discussions with roleplayed affected community members, capture photos, and bring back tabulated data and maps from an affected community that they visit.

At the same time, different teams work on their own tasks while staying closely coordinated. The Logistics Cluster arranges air transport and obtains government approvals. The Health Cluster prepares the first situation report and implementation plan. The Information Management Unit turns field data into stakeholder maps. Meanwhile, the Humanitarian Country Team brings together key stakeholders to review the latest needs assessment, agree with government officials on priority actions, and ensure that organizations do not duplicate each other’s work. Information is shared quickly through email, WhatsApp, and short meetings. This allows decisions in one area, such as the results of a rapid needs assessment, to immediately guide decisions in other areas, including sector priorities and budget planning. This highlights the importance of continuous communication and coordination to improve efficiency and avoid duplication.

This exercise tests all the course learning goals (CLG), meanwhile several Program **Learning Outcomes** are covered by strengthening students’ understanding of social science methodologies, real-world applications, and multidisci-

plinary approaches to complex global issues. It enhances their ability to analyze practical challenges, make informed decisions under pressure, and critically reflect on ethical, political, and social considerations. In addition, it builds key professional skills, including project management theory and models (covered earlier in the course), communication, and teamwork, preparing students for graduate study and careers in international organizations or civil society.

Exercise Delivery and Pedagogical Structure

The simulation component is introduced in Week 6 of the APM course. Students first receive an orientation and role-specific briefing that outlines expectations, their own individual character profiles, and foundational rules. The full-day simulation takes place in person and is followed by a structured debrief the next morning, facilitated by the Lead Facilitator and Facilitation Team members, including PhD students. During the simulation, students work in real-time to interpret evolving information, coordinate response strategies, and engage with injects delivered via email. These injects simulate the volatility of humanitarian crises and require students to adapt quickly, think clearly under pressure, self-regulate in high-stress environments, delegate tasks, and negotiate with peers occupying other humanitarian actor roles.

Learning activities include information mapping, resource management, production of situation reports, and formulation of response budgets. Communication is facilitated through physical and digital channels, primarily through e-mail messages, WhatsApp, selected social media platforms, as well as Google Workspace, supporting collaboration and the management of information. Students are also encouraged to maintain reflective journals or notes and to contribute to facilitated dialogues during the debrief session.

The facilitation of the exercise requires a large team with 7-9 staff frequently involved acting as facilitators, role-played characters and observers. Additionally, a social worker is included, ensuring the exercise is an emotionally and psychologically safe and engaging experience for all students.

Feedback is provided through multiple mechanisms. During the simulation, assigned facilitators observe student engagement and offer real-time prompts when critical gaps arise. Group feedback is delivered in the debriefing workshop, while personalized commentary accompanies each student's self-reflective written assignment.

Furthermore, feedback is not limited to the reflection report. While brief prompts may be offered during the simulation, the primary feedback moment is the structured group debrief, where facilitators synthesize observations and give targeted feedback on (i) planning and prioritization, (ii) coordination and "meeting craft," (iii) information-management and sharing (accuracy, timeliness, traceability), (iv) leadership and management (role clarity, delegation, psychological safety), and (v) decision-making under uncertainty (use of data, assumptions, and ethical trade-offs). The facilitation team also provide explicit feedback on how student "characters" engaged external stakeholders, especially a facilitation-team role-played international media/TV crew, including media management, strategic messaging for resource mobilization, advocacy and lobbying, press-briefing etiquette, message discipline, and handling pressure. Individual, personalized commentary then accompanies each reflective essay to consolidate learning and capture personal growth points.

Course Assessment Design – Simex

The APM course uses three **summative assessments** to determine the course grade of which the Simex is one of the assessment components. The Simex component of the course is a written self-reflection exercise answering six questions which collectively count for 40% of the overall course grade.

The Simex is run in a standard large lecture hall (no Virtual Reality is required), provided the room has fully moveable furniture and ample wall space for posters, rolling updates, and maps. After a brief orientation clarifies participants' roles and mechanics, the exercise formally starts: the facilitation team becomes effectively "invisible," shifting to light-

touch observation, and the role-played UN/Government leadership launch the official declaration of the emergency situation through an initial briefing that is immediately followed by a choreographed surge of timed injects—primarily via email/on-screen—to reproduce the first-hours of information overload of a rapid-onset emergency.

Periodic audio-visual news flashes (e.g., BBC World Service-style) punctuate the day; facilitation team members role-play affected community leaders for direct engagement; and a continuous cyclone soundscape evolves with the scenario. At the outset, students are encouraged to re-arrange the room space (e.g., cluster tables, set up an ops corner and press area) as they see fit to optimize coordination, information-management, and decision-making. Realism is sustained by the cadence of injects and minimal facilitator intervention during play, with synthesis and feedback reserved for the structured group debrief that follows the following day.

Individual student performance is not graded during live play. The simulation period prioritizes collaboration within designated stakeholder groups, whether institutional representations (e.g., UN agencies, national authorities, NGOs), thematic clusters, or broader peer-based teams, so that students can practice coordinated planning, clear communication, and joint decision-making. Summative assessment occurs after the exercise: students are required to complete a structured self-reflection exercise designed to deepen critical thinking and reinforce applied learning. This reflective assignment explores six key dimensions: firstly, students analyze the use of the multi-actor and inter-organizational network approach to humanitarian response, drawing on their observations from the simulation. Secondly, they examine how simulation processes evolved throughout the day, with particular attention to decision-making, coordination, cooperation, information sharing, and communication, selecting and reflecting on two of these processes in depth. Thirdly, students explore the tension between top-down and bottom-up leadership styles in complex emergency scenarios. Fourthly, they identify which specific individual skills, such as interpersonal communication, conflict resolution, or negotiation, they most employed and developed. Fifthly, students are asked to reflect on what they learned about the roles and responsibilities of international organizations and the types of decisions required in high-pressure environments. Finally, the sixth dimension of the self-reflection exercise invites students to reflect on collaborative or network strategies as discussed in the first APM lecture. Specifically, students are asked to consider network phenomena such as *structural holes*, *degree centrality* (hubs), and *betweenness centrality*—concepts that were introduced as part of the theoretical foundation of collaborative networks. During the simulation, students may have recognized that certain roles functioned as key connectors or hubs, or that information and influence were unevenly distributed across the simulated response network. By analyzing how such phenomena influenced communication flows, coordination dynamics, or decision-making bottlenecks during the exercise, students develop a more nuanced understanding of networked collaboration in emergency response. This reflection dimension aims to encourage deeper awareness of how network structures can impact efficiency, cohesion, and the effectiveness of multi-actor humanitarian efforts.

This self-reflection assignment is a summative assessment and is weighted at 40% of the overall course grade.

Assessment Rationale and Strategies

The assessment structure reflects a deliberate alignment between course goals and evaluative methods. The simulation participation component emphasizes direct engagement and collaborative performance, acknowledging the experiential value of the exercise. The individual reflective questions, which are weighted most heavily among the other summative components of the APM course, serve as a capstone intellectual output in which students critically interrogate their decisions, group dynamics, and project management strategies.

All responses are linked to the main overarching scenario, constant release of additional pieces of information and developments, stakeholder negotiations, and deliverables that occurred during the Simex. The use of oral debriefs, facilitator observations, and individual accountability maintains participant engagement throughout the exercise.

Evaluation Criteria and Academic Standards

Assessment criteria are transparently aligned with the learning outcomes and emphasize analytical rigor, theoretical application, professional engagement, and clarity of communication. For the self-reflective written exercise, the highest marks are awarded to those demonstrating an insightful and well-structured narrative that meaningfully integrates personal experience with relevant academic concepts from the APM, and related courses. Mid-tier grades reflect competent engagement with room for deeper analysis, while lower scores suggest superficial responses or limited clarity. The simulation deliverables are evaluated based on the appropriateness and coherence of the produced outputs, the quality of coordination, and the effective use of planning tools such as stakeholder maps and Gantt charts as well as strategic planning and resource mobilization templates such as Humanitarian Response Plans and emergency response grant application forms.

Reflections, Feedback, and Adaptive Exercise Design

Students are required to explicitly anchor their reflections in the theories introduced at the beginning of the chapter and in class. In their papers, they apply collaborative-network theory (e.g., the multi-actor/inter-organizational network approach) and draw on frameworks used in the debrief (such as stakeholder power-interest mapping, process views and prioritization tools) to interpret information flows, coordination dynamics, and decision-making trade-offs observed during play. Students also compare the mainstream, centralized project management cycle-based model with a decentralized coordination approach currently being piloted in a selected number of humanitarian contexts (“Flagship Initiative”) as a theory-to-practice lens. The grading **rubric** reinforces this expectation by awarding marks for correct use and definition of concepts, ensuring analysis is not merely descriptive but theory-informed.

A critical element of the course is its adaptive nature. Student feedback from the 2025 iteration highlighted initial confusion regarding simulation roles and rules. Many reported ambiguity in their assignments and a degree of cognitive overload due to the volume and intensity of information. While this replicated real-world humanitarian dynamics, facilitators acknowledged the need for clearer initial instructions and structured transitions.

In response, several adaptations were introduced. Character profiles were clarified, facilitator interventions were used sparingly but more strategically, and inject delivery was streamlined through digital means. Printed materials were reduced in an attempt to reduce excessive paper usage and to reflect the reality of current-day humanitarian operations. Furthermore, digital tools ensured continuity and access to resources. These adjustments, based on formal debriefing sessions and individual student evaluations, enhanced the clarity, inclusivity, and pedagogical integrity of the course.

Ultimately, the simulation succeeded in providing an intellectually and emotionally engaging and invariably exhausting experience for all players. It equipped students with critical insight into the challenges of international humanitarian response and collaborative disaster management, while sharpening their strategic thinking and adaptive leadership capabilities. The APM simulation exercise stands as a compelling case of how simulation can be leveraged in higher education to develop professional competencies within an academic setting.

3.11

Simulation-Based Assessment in Collaboration in the Supply Chain

Chong Zhang

Course Overview

Collaboration in the Supply Chain (CiSC) is a 3-ECTS (European Credit Transfer and Accumulation System) core course in the MSc Supply Chain Management program at Tilburg University. Built around The Fresh Connection (TFC), it combines conceptual supply chain content with decision-making, collaboration, and reflection skills. In the simulation, students manage a virtual fruit juice company in teams of four, taking the roles of Vice President (VP) Supply Chain, Purchasing, Operations, and Sales. Across several game rounds, each team designs and executes an integrated strategy to improve Return on investment (ROI).

The course is taught by two lecturers and typically hosts 80–120 students, organized into 20–30 teams. Teaching takes place through plenary lectures, synchronous Q&A sessions, and asynchronous teamwork around the simulation.

Course Learning Goals (CLG)

After completing CiSC, students are able to:

- (CLG1) Develop a coherent company strategy that aligns purchasing, supply chain, operations, and sales decisions with overarching business goals in the simulation.
- (CLG2) Understand cause-and-effect relationships in the simulated supply chain and interpret game results to improve future decision-making.
- (CLG3) Analyze supply and demand dynamics, forecasting, and sustainability issues (including CO₂ foot printing) in a global supply chain context.
- (CLG4) Reflect on intra-team and inter-team collaboration (including communication, negotiation, and conflict handling) and evaluate how team processes affect performance.

Link to Program Learning Outcomes

CiSC contributes to several program **learning outcomes** in the MSc Supply Chain Management program in the following ways:

- Strategic and integrated supply chain thinking: students design and justify integrated strategies across functions.
- Analytical problem solving in supply chains: students analyze data and dynamic supply chain behavior and derive justified decisions.

- Application of theory to practice: students use academic SCM literature to interpret and improve real-life or simulated business situations.
- Collaboration and professional skills: students collaborate in diverse teams, manage cross-functional tensions, and reflect on their own role.
- Responsible and sustainable decision-making: students recognize the sustainability and CO₂ foot printing implications of supply chain decisions.

Teaching Activities

Teaching activities, simulation, and **assessment** are woven into a coherent learning cycle. An opening lecture introduces the course, the simulation, and the assessment structure. Students then work in teams on a sequence of simulation rounds, supported by asynchronous tutorials (four team discussions and four buddy team interactions), and synchronous tutorials (three Q&A sessions). A closing lecture synthesizes learning and connects simulation experiences back to theory. Figure 3.11.1 illustrates an overview of teaching activities.

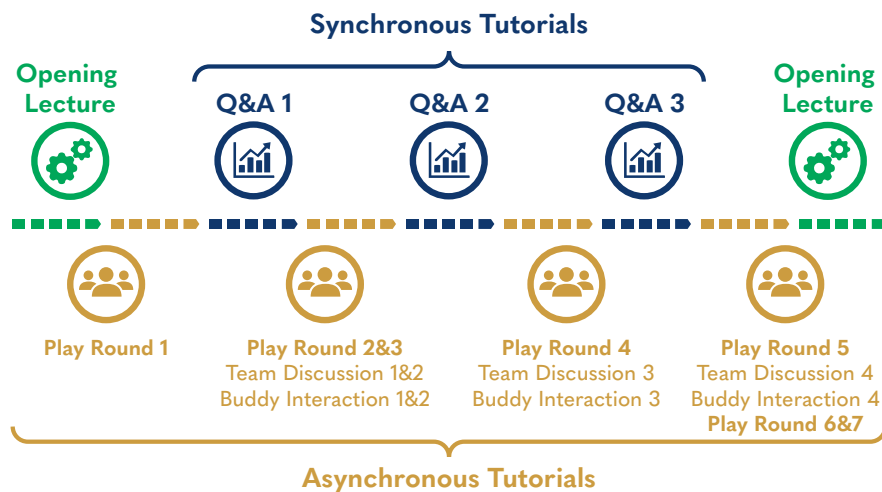


Figure 3.11.1
Teaching Activities Overview

- Opening lecture: introduces course goals, assessment, simulation context, team roles, and the learning cycle. It clarifies how team reports and assignments will build on gameplay.
- Asynchronous team discussions: within each team, students analyze previous round results, discuss relevant literature, design decisions for the next round, and record reflections in a team report.
- Asynchronous buddy team interactions: pairs of teams exchange analyses and provide feedback to each other. They are explicitly asked not to copy decisions but to use each other as critical friends.
- Synchronous Q&A sessions: each team has short time slots with a lecturer to debrief game rounds, clarify questions, and discuss strategy and collaboration issues.
- Closing lecture: brings together insights from game data, survey results, and assignments. Top-performing teams, or teams with particularly interesting learning trajectories, present their strategies and lessons learned.

Learning Cycle

Tutorials are the core component of the CiSC course. They are designed around five key building blocks that guide the learning process:

(i) Search for Concepts – before & after tutorials

Students are encouraged to use academic literature to justify their decisions in the game and their responses to assignments.

(ii) Apply Concepts to the Game – before, during & after tutorials

Students are encouraged to apply the concepts found in the literature and reflect on how these concepts have improved their decision-making.

(iii) Reflect Within Your Team – during tutorials

Students are encouraged to collaborate with their team members to review and discuss previous decisions and strategize for the upcoming round.

(iv) Reflect Across Teams – during tutorials

Each team will be paired with a buddy team. The goal is to exchange experiences and learn from the decision-making processes of other teams.

(v) Q&A With Lecturers – during tutorials

Each team is allocated a timeslot to ask the lecturer specific questions related to the game, allowing for clarification and a deeper understanding of strategic decisions.

The above five building blocks form a learning cycle and are interconnected through three learning stages:

1. **Learning from Theory:** This stage involves searching for literature to support your decision-making processes.
2. **Learning from Collaboration:** This stage centers on gaining insights through interactions with team members, buddy teams, and lecturers.
3. **Learning by Doing:** This stage, focused on gameplay, involves analyzing both successes and failures to inform and guide (corrective) actions.

Figure 3.11.2 illustrates the building blocks and learning cycle of the CiSC course.

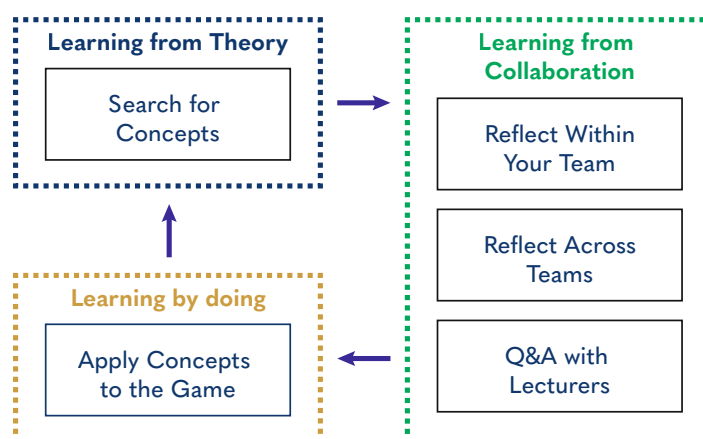


Figure 3.11.2
Building Blocks and Learning Cycle

Teaching Activities Connected to The Learning Cycle

Each teaching activity in CiSC is deliberately linked to the learning cycle, as shown in Figure 3.11.3, which summarizes how instruction release, tutorials, gameplay, and document submission are sequenced and connected.

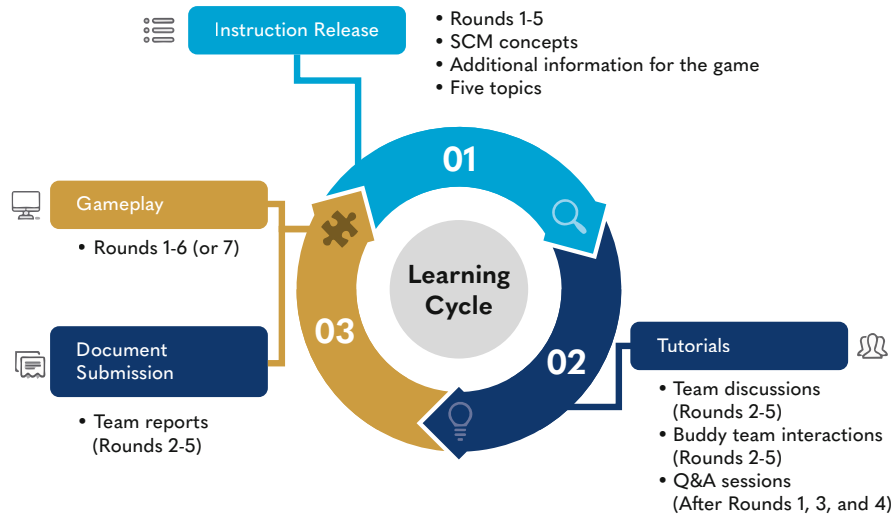


Figure 3.11.3
Teaching Activities and Learning Cycle

Assessment Design

Table 3.11.1 shows how each learning goal contributes to the categories of program learning outcomes of MSc Supply Chain Management and how each is assessed by multiple assessment types.

Table 3.11.1
Course Learning Goals, Program Learning Outcomes Categories, and Assessment Type(s)

Course learning goals (CLG's)	Program learning outcomes categories	Assessment type(s)
CLG1	Applying knowledge, Problem solving	Team assignment, Individual assignment
GLG2	Applying knowledge, Problem solving	Team assignment, Individual assignment
CLG3	Applying knowledge, Problem solving, Information skills	Team assignment, Individual assignment
CLG4	Learning and development, Communication skills, Responsibility and independence	Simulation game score, Individual assignment, Team assignment

The assessment is designed so that the simulation drives learning, while written assignments validate individual and team learning outcomes. There is no invigilated exam. Each assessment component is displayed in Table 3.11.2.

- The team ROI score (10%) rewards serious engagement with the simulation but has a limited impact on the final grade. Only the best round counts, which encourages experimentation and learning over time.
- The individual assignment (60%) is the main vehicle to assess whether each student has personally achieved the learning goals. It is role-specific and asks students to connect their own decisions, game data, and academic literature.
- The team assignment (30%) evaluates firm-level strategy, cross-functional alignment, and collaboration with buddy teams. It builds on all game rounds and team reports, capturing collective, longitudinal learning.

Table 3.11.2
Assessment Components

Component	Weight	Type	Main focus
Team ROI score (best round)	10%	Summative	Serious engagement with the simulation, high-level performance incentive
Individual assignment	60%	Summative	Role-specific reflection, theory application, individual learning outcomes
Team assignment	30%	Summative	Firm-level analysis, cross-functional alignment, collaboration reflection
Team reports after rounds	0%	Formative	Ongoing reflection, note-taking, input for final assignments

Grading Policies

The minimum acceptable grade for each assignment is 55 out of 100. To pass the course, students must achieve a final grade of at least 5.5. Late submissions are not accepted and receive a grade of zero.

Students who fail to meet the minimum criteria of 55 for an assignment or a final grade of at least 5.5 (rounded) must take a resit. Each student has one resit opportunity for both the individual and team assignments. After grades are released, a perusal session is conducted by two examiners. The resit is scheduled within the same semester, following the perusal. For the resit, students can improve their assignments based on our feedback and insights from the perusal session. The resit grade calculation mirrors the initial process.

Should a student still receive an insufficient grade on the resit, they become a “repeater,” requiring them to retake the course, including participating in tutorials, completing all reports, and redoing the final individual and team assignments. Previous semester grades are not carried over for repeaters.

Assignment Descriptions

Individual assignment:

- Role-specific reflection (e.g., Vice President of Purchasing, VP Purchasing) of 2,500–3,000 words.
- Typical tasks:
 - Analyze role-specific decisions and game outcomes.
 - Explain how SCM theories informed these decisions.
 - Reflect on mistakes and corrective actions across rounds.
 - Discuss learning about decision-making.
- AI risk mitigation:
 - Questions require student-specific game data, screenshots, and team experiences, which generative artificial intelligence (GenAI) cannot fabricate convincingly.
 - Students must include a Statement of Authorship and AI Usage with prompts and tools used, following an AI Index Level 4 policy that allows AI as a support tool but not as a generator of final text.

Team assignment:

- Team reflection of 1,500–2,500 words.
- Typical tasks
 - Written from a firm-level perspective across all game rounds and topics (strategy and alignment, supply and demand management, supply chain sustainability, supply chain collaboration, etc.).
 - Includes analysis of the company’s overall strategy, how this strategy was adapted to internal and external factors, and a reflection on the team’s collaborative skills.
- AI risk mitigation:
 - The task requires joint reflection on unique gameplay, team discussions, and buddy team interactions.

- The assignment includes questions on how teams interacted with their buddy teams and how they responded to specific feedback, which again ties work to authentic interactions.
- Students must include a Statement of Authorship and AI Usage.

Assessment Criteria

The Individual Assignment is graded on:

- Reflective capacity (how did you reflect on the game?)
- Seriousness (although you play a game, did you capture the intended and unintended course learning goals?)
- Applying theory to practice (sensible use of references, at least 10, among which at least 5 should come from a list of recommended journals)
- Plausibility of the answers (given the theory and the data provided in the text)
- Level of consistency and logic in the argumentation
- Structure of the assignment

Team Assignment is graded on:

- Completeness (did you cover all questions?)
- Analysis (using models and techniques discussed in class, as well as those found by yourself in the literature)
- Originality (out-of-the-box thinking)
- Reporting quality (readability, good use of figures and tables)
- Applying theory to practice (sensible use of references, at least 10, among which at least 5 should come from a list of recommended journals)
- Plausibility of the answers (given the theory and the data provided in the text)
- Level of consistency and logic in the argumentation
- Structure of the assignment

The detailed **rubrics** are displayed in Appendix 3.11 at the end of this chapter.

Reflections on Challenges and Adaptations

Game Complexity

The simulation is complex, with many parameters per VP and increasing difficulty over the rounds. Some students initially felt overwhelmed by the number of decisions and the pace of the game.

The following adaptations are planned:

- Round-specific instruction slides that introduce only the new concepts for that round, with concrete tips and references.
- Clearly communicated time windows for each asynchronous tutorial, including recommended durations.
- A structured format for Q&A sessions that first addresses general issues and then focuses on team-specific questions.

Communication

Many moving parts (team formation, game registration, round deadlines, team and buddy meetings, Q&A slots, assignments) made the course logistics demanding for students.

The following adaptations are planned:

- A well-structured Canvas homepage with a clear timeline for each activity and document release.
- Regular announcements before each deadline.
- Standardized templates for submissions to make expectations transparent.

Motivation and Fairness of The ROI Component

Using ROI as an assessment component requires balancing motivation and **fairness**.

The following adaptations are planned:

- Including ROI as a small, graded component to stimulate serious engagement with the game.
- Keeping its weight low and concentrating most of the grade on reflective assignments, so that the course primarily rewards learning rather than performance alone.

Inclusion and Accessibility

The simulation requires a paid license fee, and students differ in their prior experience and confidence with gaming and analytics.

The following adaptations are planned:

- Designing an alternative path for students who cannot or prefer not to pay the game fee, based on static screenshots and equivalent analytical tasks.
- Securing university funding to cover the license fees for three years.
- Explicitly explaining that, although some students may have used the game in their bachelor's program, this course operates at a deeper level by emphasizing the application of Supply Chain Management theories and reflection on collaboration, and the overall course and assessment design supports this deeper learning.

Assessment Workload

The assessment workload is manageable for students, as we give them 4 weeks after the final lecture to complete the assignments.

However, it's not very manageable for lecturers to grade, especially when the number of students exceeds 100. We are still thinking about how to improve this part but have not come up with a solution yet.

Editorial Suggestions for Future Improvements

For future editions, it may be valuable to provide a clearer rationale for the **cut-off score** (e.g., 55 out of 100). In particular, **transparency** could be improved by explaining how such cut-off scores relate to student performance standards and learning outcomes. Where possible, passing grades should be grounded in a formal **standard-setting** approach (for example, through consultation with subject matter experts to determine the expected performance of a borderline student), rather than being presented as a fixed convention.

Although the current threshold is determined by institution-wide policy and therefore lies beyond the control of individual authors, a brief discussion of the underlying principles of grading and standard setting could help readers better understand the rationale and limitations of the system, and highlight directions for future policy development.

To improve transparency, using AI for **formative assessment** and feedback as a starting point for peer and self-assessment may help students better understand the assessment criteria.

Appendix 3.11 CiSC Assignment Grading Rubrics

Individual Assignment Grading Rubric	
Criteria	Points
Part A. Question 1 <ul style="list-style-type: none"> Coverage of All Three Phases: Check whether all three phases (Briefing, Gameplay, and Debriefing) are explicitly mentioned and clearly defined in the student's own words. Clarity of Interrelatedness: Evaluate how clearly the student explains the influence of one phase on another. A strong answer should reflect on how one phase shaped others (e.g., how briefing influenced gameplay decisions and debriefing reflections). (Without discussing this interrelation, the maximum score for this question is 12.) Use and Application of Literature: At least three academic references are expected. The literature should be used to support arguments and reflections, not just to define terms or cite concepts. Look for evidence that the student connects theoretical insights to their gameplay experience in a meaningful way. 	__/20 pts
Part A. Question 2 <ul style="list-style-type: none"> Adherence to Instructions: Students must choose three SCM subjects from the provided list. Do not invent or combine subjects outside this list (e.g., avoid using general categories like "service level" or "contract index" unless clearly tied to one of the specified subjects). Explanation of Subjects: Each subject/concept must be clearly defined. Students should also explain how the concept relates to the decisions made or situations encountered during gameplay. Clarity in Extent of Engagement: Students must clearly indicate the extent to which each subject was encountered using "low," "mid," or "high". Avoid vague terms such as "often" or "most times". Justifications should explain why each subject was encountered at that level, ideally with specific examples from gameplay. (Without this clarity, the maximum score for this question is 12.) Use and Application of Literature: At least three academic references are expected. Literature should be used to support reflections and arguments, not simply to define a term. Look for integration of theory into gameplay experiences. 	__/20 pts
Part B. Question 2 <ul style="list-style-type: none"> Inclusion of Cause-and-Effect Diagram: A cause-and-effect diagram is required for this question. It should be included clearly in the main body of the text, not hidden in the appendix. Missing or unclear diagrams will lead to a deduction. Correct Focus: The central effect in the diagram must be the key KPI identified in Question 1, not ROI or another unrelated metric. The diagram should show how that specific KPI is influenced. Clarity of Causal Relationships: The diagram should clearly distinguish between decisions made by the student's own VP, decisions made by other VPs, and any relevant external factors. External Factors External factors (e.g., CO₂, market volatility, lead time disruptions) must be included as potential causes. A complete diagram considers both internal and external influences on the KPI. 	__/10 pts
Part B. Question 2 <ul style="list-style-type: none"> Inclusion of Cause-and-Effect Diagram: A cause-and-effect diagram is required for this question. It should be included clearly in the main body of the text, not hidden in the appendix. Missing or unclear diagrams will lead to a deduction. Correct Focus: The central effect in the diagram must be the key KPI identified in Question 1, not ROI or another unrelated metric. The diagram should show how that specific KPI is influenced. Clarity of Causal Relationships: The diagram should clearly distinguish between decisions made by the student's own VP, decisions made by other VPs, and any relevant external factors. External Factors External factors (e.g., CO₂, market volatility, lead time disruptions) must be included as potential causes. A complete diagram considers both internal and external influences on the KPI. 	__/10 pts

Individual Assignment Grading Rubric	
Criteria	Points
Part B. Question 3 <ul style="list-style-type: none"> Type of Decision-Making Identified: Most students correctly identified procedural rationality as their team's approach. The description of the decision-making type (procedural vs. political) should be accurate and supported with examples. Link to Key KPI: Students must clearly explain how the decision-making process (e.g., procedural or political) benefited or impacted their own key KPI (as defined in Question 1). This is one of the core elements we are evaluating in this question. A common mistake is to discuss general team performance or ROI without linking the decision-making process to the student's specific KPI. (Without this link, the maximum score for this question is 10.) Use of Literature: At least three academic references are expected. Literature should be used to support the analysis of decision-making approaches, not only to define them. Strong answers apply theory to actual team dynamics and outcomes. 	__/20 pts
Part B. Question 4 <ul style="list-style-type: none"> Focus on Activities: If we ask for activities, mention specific activities under your VP role that contribute to CO₂ emissions. Do not describe decisions: answer the question directly. Units for Numerical Data: When presenting CO₂ emission figures, do not forget to include the correct units (e.g., kg CO₂ per liter, etc.). Answers without units are incomplete. 	__/5 pts
Part B. Question 5 <ul style="list-style-type: none"> Focus on Decisions (Plural): The question asks for "decisions" made by your VP that influenced the carbon footprint. Therefore, we expect at least two distinct decisions. Mentioning only one decision is insufficient. Relevance to CO₂ Emissions: The decisions must be directly related to the carbon footprint. General decisions unrelated to emissions or vague operational choices will not meet the requirements of this question. 	__/10 pts
Part B. Question 6a <ul style="list-style-type: none"> More Than One Solution Required: The question asks for "solutions" (plural), so we expect at least two solutions. Providing only one solution is not sufficient for full credit. Consistency with Question 5: The solutions must be directly related to the decisions discussed in Question 5. Do not introduce new solutions that were not previously mentioned. 	__/2 pts
Part B. Question 6b <ul style="list-style-type: none"> Address Both Positive and Negative Impacts: The question requires you to discuss both positive and negative impacts of the solutions from Question 6a. Mentioning only one side is incomplete. Clarify Each Impact: Simply listing effects is not enough, and you need to clearly state whether each impact is positive or negative. 	__/3 pts
Structure of the assignment <ul style="list-style-type: none"> Adhere to Question Structure: Follow the structure of the assignment questions strictly. For example, do not combine answers to 6a and 6b into a single response. APA In-Text Citations: If APA in-text citations are not properly formatted (e.g., missing author-date format, no citation where needed), points will be deducted. APA Reference List: If the APA reference list is missing or not formatted correctly (e.g., inconsistent punctuation, missing italics, incorrect order), points will be deducted. Quantity and Quality of References: Students are expected to use at least 10 academic references. Among these, at least 5 must come from the recommended journal list provided in Appendix X of the course manual. 	__/0 pts

Team Assignment Grading Rubric	
Criteria	Points
Part A. Question 1	__/10 pts
Part A. Question 2	__/10 pts
Part A. Question 3	__/10 pts
Part B. Question 4	__/12 pts
Part B. Question 5	__/12 pts
Part C. Question 6	__/12 pts
Part C. Question 7	__/8 pts
Part C. Question 8	__/10 pts
Originality (out-of-the-box thinking)	__/8 pts
Applying theory to practice	__/8 pts
Structure of the assignment Did the teams adhere to the structure of the assignment?	__/0 pts

Note. For the team assignment, detailed sub-criteria are not specified for each question; instead, evaluative comments are provided when needed, especially for lower scores.

3.12

Simulation-Based Assessment in Data-Driven Design Approaches

Roland van de Kerkhof

Course Overview

Data-Driven Design Approaches in Supply Chain Management (DDDAiSCM, 325232-M-6) is a 6-ECTS (European Credit Transfer and Accumulation System) core course in the Master of Science in Supply Chain Management (MSCM) at Tilburg University. Each year, about 150 master's students participate, working in teams of five on a system dynamics project.

The course focuses on understanding and modeling supply chain dynamics. Through a mix of lectures, weekly coaching, and a seven-week project, students learn to build their own system dynamics simulations and use them to test scenarios and develop recommendations.

A unique aspect of the course is its annually updated real-world case. Each cohort engages with a new simulation context developed with an external partner. Previous cases have included hospital operations, ramp-up and quality strategies for high-tech manufacturers, and regulatory scenarios for the circular plastics industry in the Netherlands.

The course is taught by a three-member academic team, each coaching one-third of the project teams. By the end, students deliver a functioning simulation model and data-driven insights that support decision-making in complex supply chain environments.

Course Learning Goals

After successful completion of this course, students will be able to analyze complex supply chain problems and design solutions for these problems from an integral and holistic perspective.

To be specific, after successful completion of this course, students will be able to:

- CLG1: Identify when it is useful to apply a system dynamics perspective and/or simulation method to a supply chain problem;
- CLG2: Understand how underlying feedback structures generate specific instances of dynamic behavior in manufacturing and service supply chains;
- CLG3: Develop conceptual and quantitative System Dynamics simulation models of supply chain problems;
- CLG4: Know how the quality of conceptual and quantitative System Dynamics simulation models can be evaluated;
- CLG5: Analyze supply chain problems from an integral and holistic perspective by using conceptual and quantitative System Dynamics simulation models;

- CLG6: Design effective solutions for supply chain problems from an integral and holistic perspective by using quantitative System Dynamics simulation models;
- CLG7: Communicate the insights from the analysis and the solutions designed to managers.

Link to Program Learning Outcomes

DDDAiSCM contributes to several program learning outcomes in the MSc Supply Chain Management program, as indicated in Table 3.12.1.

Table 3.12.1
Program Learning Outcomes and Course Learning Goals

Intended Learning Outcomes Master SCM	Course Learning Goals
<i>Knowledge:</i> Critically evaluate specialized and advanced supply chain management concepts, models and theories, including their foundations and recent developments, at the interface of academia and professional practice.	CLG1, CLG2, CLG4
<i>Skills – Research:</i> Independently conduct a delineated research project compliant with acceptable methodological standards and relying on current theoretical insights.	
<i>Skills – Argumentation:</i> Develop a solid and objective argumentation.	CLG7
<i>Skills – Problem solving:</i> Identify and analyze complex problems in supply chain management practice and contribute to original and science-based solutions of such problems by identifying and using data.	CLG5, CLG6
<i>Skills – Learning:</i> Reflect on personal (learning) results, resulting in actions to continuously improve the learning process	CLG3
<i>Skills – Information:</i> Critically collect and analyze scientific and empirical information related to supply chain management in a responsible way.	CLG3
<i>Skills – Communication:</i> Communicate in a targeted way with peers, colleagues, supervisors and/or other specialists and non-specialists, according to academic and/or managerial conventions	CLG7
<i>Responsibility & Independence:</i> Effectively plan and pursue a project or learning process in complex dynamic environments, alone or in a team, and carry responsibility for the process and the outcome, for others and for society at large.	CLG3
<i>Responsibility & Independence – Judgments:</i> Formulate critical judgements based on incomplete/limited info, taking into account social, scientific and ethical responsibilities related to the application of own knowledge.	CLG5, CLG6, CLG7

In DDDAiSCM, learning takes place mainly during the 7-week project, which takes the students roughly 2 days per week. During this project, they have to develop a qualitative **assessment** of the case, identify potential solutions, develop a system dynamics simulation model that allows them to test their solutions, and present their insights and recommendations to the case owner. We believe students learn most about the dynamics of the system if they have to develop the simulation model themselves; students can only model something if they really understand how it functions. Every week, we coach each team for 20 minutes.

The course supports most of the program intended learning outcomes of the Supply Chain Management master. The lectures and project help in developing their supply chain management knowledge, and their skills and independence are developed during the project.

Teaching Activities

This course contains three types of instructions, namely:

- Lectures
- Guest lectures (for the project)
- Coaching sessions

All types of instructions help the student prepare for the final exam of this course. To help the students and their teams with the team assignment, the guest lectures and coaching sessions are organized. Table 3.12.2 is a more detailed overview of each type of instruction.

Table 3.12.2
Scheduling of Team Assignment

Week	Lecture	Project	Phase
1	<ul style="list-style-type: none"> • Course introduction • Method introduction 	<ul style="list-style-type: none"> • Case introduction 	1
2	<ul style="list-style-type: none"> • Conceptual modelling method • Supply chain dynamics (conceptual) 	<ul style="list-style-type: none"> • Coaching session 	
3	<ul style="list-style-type: none"> • Quantitative modelling method • Core building blocks project 	<ul style="list-style-type: none"> • Coaching session 	2
4	<ul style="list-style-type: none"> • Supply chain dynamics (quantitative) 	<ul style="list-style-type: none"> • Coaching session • Q&A session 	
5	<ul style="list-style-type: none"> • Model validation • Model initialization 	<ul style="list-style-type: none"> • Coaching session 	
6	<ul style="list-style-type: none"> • Policy design and testing • Policy communication 	<ul style="list-style-type: none"> • Coaching session 	3
7	<ul style="list-style-type: none"> • Looking back and forward • En-ROADS workshop 	<ul style="list-style-type: none"> • Coaching session • Presentation 	

Lectures

Every week starts with a lecture. During these lectures, the lecturers explain several main theoretical concepts and aspects of system (dynamic) approach to complex business situations (the methodology), and provide real-life examples from their own research and work experiences.

Guest Lectures (Project)

During the three guest lectures, one or multiple employees from the case organization will visit the university campus to discuss the case. The setup of the guest lectures is as following:

- *Case introduction*: 45 minutes case and company introduction, followed by a 45 minutes Q&A. Students are recommended to prepare well for this first session by reading the case description, making notes of what they do not understand right away and writing down the questions they might want to ask.
- *Q&A session*: 90 minutes Q&A with the case organization. This session provides the students the opportunity to ask anything they need to know to further develop their simulation, and about specific solutions they might want to test. All teams will be asked to write down their questions in a Google Sheet beforehand, such that the people from the case organization can come prepared.
- *Presentation*: the teams with the best presentations will be asked to present their findings and recommendations to the case organization during this third and final guest lecture. All students are invited to participate in the discussion following the presentations, such that we can provide a joint recommendation to the case organization.

Coaching Sessions

A total of six coaching sessions will be offered to each team. During these 20-minute coaching sessions, we discuss their analyses, the problems they run into when developing their model, the solutions they have in mind, etc. We call them coaching sessions because the aim is to let them find the answers and ensure that they can keep progressing in the right direction. Students are recommended to prepare a list of questions they'd like to discuss in each coaching session, we provide general guidelines on what to discuss in each of the six coaching sessions (see Table 3.12.3).

Table 3.12.3
Coaching Sessions

Phase	Session	Guidelines
1	1	Since you have only one week to complete the qualitative analysis, it is wise to have at least the first version of the causal loop diagram at the start of the coaching session. After the session, you can finalize your qualitative analysis.
2	2	As this is probably the first system dynamics model you've made, it is wise to (at least) discuss the overall design of your model and the development process. What are the key variables (and stocks), what are the main dynamics you want to include, where do you start, how do you work together, etc.
	3	Since you've been working on the simulation for slightly over a week, it is likely that you've run into some modelling problems. It is recommended to write down everything you run into and send it to your lecturer beforehand (latest 23:59 the day before), such that you can receive tips on each of these items during the coaching session.
	4	The simulation should be almost complete by now. During this coaching session, you can discuss the remaining questions and walk through the model together: are the units ok, are all processes incorporated, what assumptions did you make, what is the impact of these assumptions on your ability to analyze the effects of the solutions you'd like to test?
3	5	Two sessions remain: during the first of these, it is recommended to discuss all questions you have about the model we provided, as well as the setup of your presentation and your approach: what policies are you going to test, (how) are you going to perform a sensitivity analysis beforehand, what KPIs are you interested in, how are you going to test the robustness of your policies, and how are you going to present your findings.
	6	During the final coaching session, let's walk through your presentation together and discuss the interpretation of your findings: given these findings, what would you recommend the case organization?

Delivery Types

The seven lectures and the project-related lectures are offered physically, on-campus. For the six coaching sessions, each team can choose to either meet their lecturer physically on campus or online.

Assessment Overview

Cours learning goals 3, 5, 6, and 7 are assessed summatively in the project (40% of final grade), all course learning goals are assessed in the exam (60% of final grade) (see Table 3.12.4).

Table 3.12.4
Assessment Overview

Component	Weight	Type	Main focus
Project part 1: Qualitative analysis	10%	Summative (team)	Qualitative analysis of the dynamics in the case, the root cause(s) of the problem and potential solutions
Project part 2: Simulation model	15%	Summative (team)	Develop a system dynamics simulation model that can be used to test the potential solutions
Project part 3: Recommendations	15%	Summative (team)	Test the potential solutions, check their robustness and design a presentation to communicate your insights and recommendations
Exam	60%	Summative (individual)	All course learning goals

Justification of Weighting

In the Supply Chain Management master, we want to make sure that all students have gained a sufficiently deep understanding of and proficiency with the course materials. Therefore, we allocate 60% of the final grade to the exam and 40% to the project, even though students spend more time on the project than on studying for the exam. Luckily, if they've performed well on the project, they are also likely to get a high grade on the exam.

Since repeaters do not have to perform the project again (see Grading policies), we aim to align the weights for the project with the weights in the exam. For both, more than half of the grade for this course is directly related to the development and analysis of the system dynamics simulation models, as specified in Table 3.12.5.

Table 3.12.5
Grading Scheme

Examination of	Exam	Project
Develop causal loop diagram	10/90 points (constructed-response questions)	10% of final grade
Develop simulation model	20/90 points (constructed-response questions)	15% of final grade
Analyze simulation model and provide recommendations	20/90 points (constructed-response questions)	15% of final grade
Other learning goals	40/90 points (multiple-choice questions)	(involved in project)

Grading Policies

While the final grade is based on a weighted average of the four grades received, students have to score at least a 5.5 for the exam to be able to pass the course. In addition, the weighted average of their four grades needs to be higher than 5.5. If students participate in the resit, the resit grade overwrites the grade of the first exam. If students have not been able to successfully pass the course in the first year, they can participate in subsequent years as repeaters. Repeaters do not take part in the project again; for repeaters, the exam determines 100% of the final grade.

If students have an issue with the grade received for an assignment, they are asked to contact their lecturer via e-mail. There is a time limit: adjustments to assignment grades are made only in the first week the grades are provided. In this e-mail, they are asked to provide the **rubric item**, current score, expected score, argument(s), and evidence to support their argument(s).

Grading Procedures

Grading simulation models and policy recommendations is inherently subjective, as there is no single correct solution: the same system can be modelled in multiple valid ways, and different modelling choices can lead to equally defensible outcomes. For this reason, particular care is taken to ensure that grading is fair, transparent, and consistent across teams.

Each coach grades the work of their own team. This approach is intentional. Coaches provide ongoing, team-specific feedback throughout the project, and aligning grading with that coaching ensures that teams are assessed in relation to the guidance they received. At the same time, all grading is anchored in a shared grading rubric used by all coaches.

To promote consistency, the grading process includes several collective safeguards. The first coach makes explicit how each criterion of the rubric has been applied, clarifying how specific modelling choices or features are translated into grades. When a coach is uncertain about how to assess a particular aspect of a team's model or recommendations, this is discussed collectively with the other coaches. Similarly, cases in which a team may receive a very low or very high grade are discussed as a group before finalizing the assessment.

After each coach's grading is completed, we compare average grades across coaches, both for individual rubric components and for the overall grade. If discrepancies are identified, we jointly examine their causes and determine whether differences reflect legitimate variations in team performance or whether adjustments are needed to ensure consistency.

Project Design

In this course, we offer students the possibility to perform a real-life project, providing recommendations to an organization with a dynamically complex supply chain problem. Prior to the start of the course, we develop case descriptions of 8-10 pages together with the case company. During the first week of the course, people from the case organization introduce the problem, the processes involved, the solution space and – if relevant – the dataset. Then it is up to the students to convert this information into a system dynamics simulation model that allows them to analyze the problem and potential solutions, and to come up with a recommendation for the case company.

The project is split into three phases, each with a specific set of deliverables. After the deadline of each phase, we share our answers with the students. That is, if they haven't been able to create a proper causal loop diagram in the first phase, they can use ours as a basis for the simulation model. In phase 3, we all use the same simulation model (ours), such that we can collaboratively derive the main insights and recommendations (if all teams would be using their own simulation model, differences in insights are most likely caused by differences in the structure of their simulation models). In the remainder of this section, we elaborate on each of the phases.

Phase 1: Qualitative Analysis

Before the students start developing a quantitative model, it is important to understand and scope the problem. A good way of doing this for them is to create a causal loop diagram of the problem, only including the variables and relationships that they also want to include in their quantitative model in the next phase. At this stage, it is also good for students to start thinking about what solutions they might want to test in a later phase, as these have to be incorporated into the quantitative model.

Deliverables:

- A causal loop diagram, developed in the system dynamics software package Vensim PLE
- A written analysis of the (root) causes of the problem
- A list of solutions the students would like to test

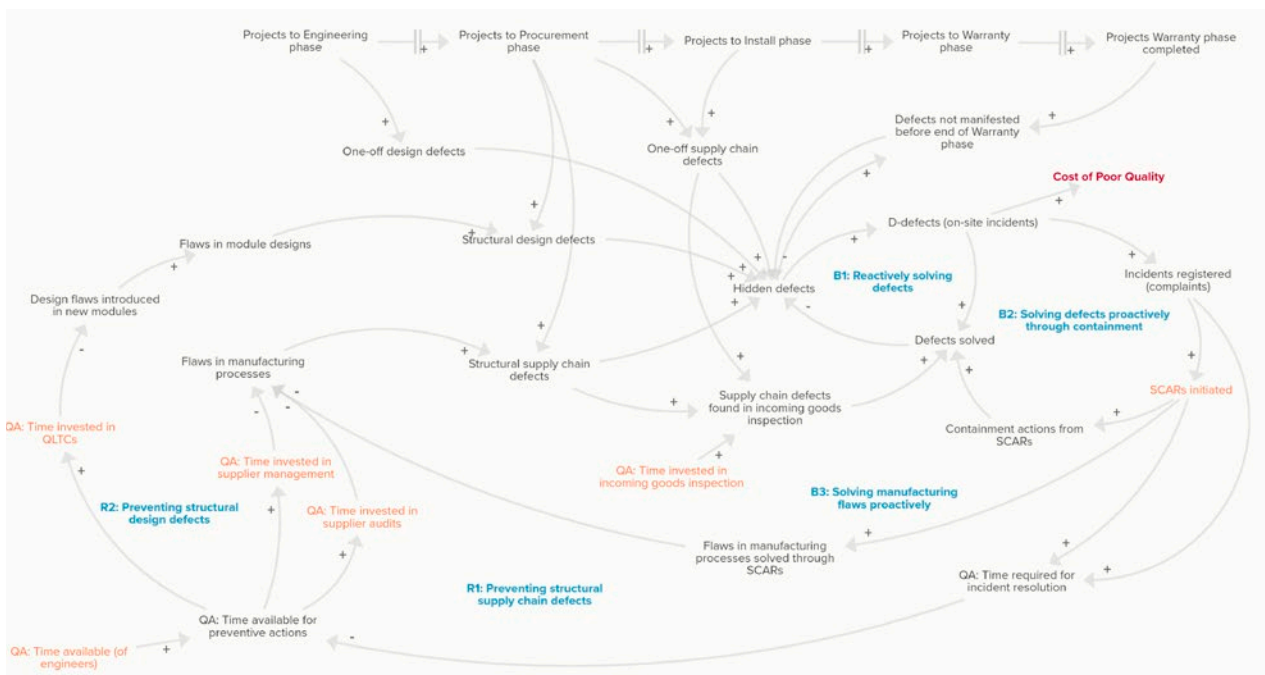


Figure 3.12.1
Illustration of Causal Loop Diagram

Phase 2: Simulation Model

Then it is up to the students to develop the smallest quantitative model possible that can explain the problem and allow them to test their potential solutions. Naturally, they have to make some assumptions in this process. No simulation model can perfectly mimic the real world, but some models are useful. Therefore, we ask them to be transparent about the assumptions they make and reflect on the impact of these assumptions on the analysis of the solutions.

Deliverables:

- A system dynamics model, developed in the system dynamics software package Silico
- A written description of the model, including an overview of the variables and relationships, a description of the main assumptions, and a reflection on the impact of these assumptions on the quality of the analyses

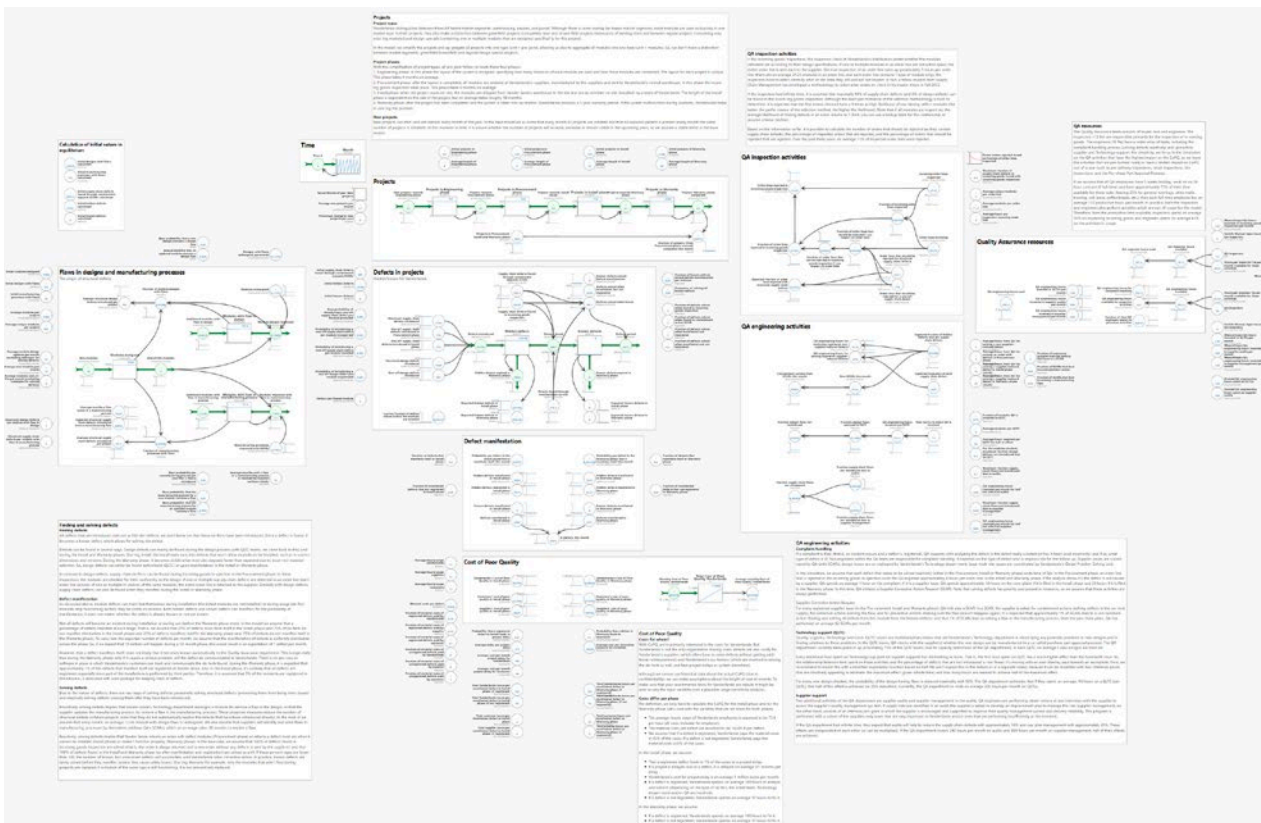


Figure 3.12.2
Illustration of System Dynamics Simulation Model

Phase 3: Recommendations

In the final phase, it is up to the students to test their solution designs, identify under which conditions these designs are indeed good solutions (how robust are the solutions), and present their insights and recommendations to the organization. Both insights and recommendations can be valuable. The teams with the best presentations are selected to present to the case organization. All presentations are sent to the case organization, together with a joint recommendation of all students (that we draft together during the final session).

Deliverables:

- A PowerPoint presentation that clearly communicates the students' insights and recommendations
- A 10-minute presentation to the case organization

Assessment Criteria

The project assignments are graded with the rubrics below (excluding the detailed descriptions of the criteria). Although we update the descriptions for the Case understanding criteria yearly (to align with the content of the case), the remainder of the rubrics remain stable across years (see Table 3.12.6).

Table 3.12.6
Assessment Rubrics

Dimensions	Criteria	Max points
Methodology	Variables are correct	2
	Relationships and feedback loops are correct	2
	Layout and format help understanding	2
Case understanding	Main variables included	2
	Sensible first analysis: problems & solutions	2

Simplified rubric of first phase

	Items	Max points
Methodology	Model structure makes sense	2
	Model behavior makes sense	2
	Recommended good practices used	2
Case understanding	Main variables used (model completeness)	2
	Quality of reflection on impact of assumptions	2

Simplified rubric of second phase

	Items	Max points
Methodology	Sound policy analysis	2
	Quality of robustness checks	2
	Clarity of communication	2
Case understanding	Quality of recommendations	2
	All impacts considered	2

Simplified rubric of third phase

Reflections on Challenges and Adaptations

Simulation Complexity

One of the key trade-offs we face (and intensively discuss) every year is the desired level of detail in the simulation model⁵⁸. On the one hand, we hope to aid the case organization, generating new insights and practicable recommendations during the project. This requires making the simulation realistic, such that it resembles the case's context and solution space sufficiently well. On the other hand, this is the first system dynamics simulation model the students make, so it should be doable for inexperienced teams to develop the simulation model in a handful of weeks.

Our solution is twofold: every year we try to create the smallest and simplest simulation model that can simulate the consequences of potential solutions sufficiently accurate (it does not have to be perfectly accurate). In addition,

⁵⁸ All models are simplifications of reality. In professional settings, models of complex systems such as factories or supply chains that aim to closely reflect real-world dynamics typically require on the order of 2,000 to 25,000 variables. By comparison, the models developed in this course, which usually contain approximately 100–200 variables, are therefore considered sufficiently realistic for educational and analytical purposes.

we encourage the students to develop the best simulation model they can make within the time given (in phase 2), but also explain to them that it is ok if some parts of the simulation model are simplified. At the start of phase 3, we provide them with our simulation model, which is typically more detailed than the ones the students have developed, and does allow for a more accurate analysis of the policies.

It Is Dynamic

Preparation is the key to success. While the course starts at the end of January or early February, we start with the preparations for the case in September or October. This includes finding a case organization, specifying the case, developing the simulation model, developing the case description, and performing a first qualitative analysis.

However, good preparation is not sufficient. Every year, students come up with creative ideas in phase 1, and we would like to be able to test these potential solutions in phase 3. In addition, we always receive new information during and after the students' Q&A sessions with the case organization. Therefore, we have to update our simulation model during the course.

3.13

Reflection and Logbook Assessment

Miranda Stienstra, Astrid Kramer, Louise van Hoek, and Ya-Ping (Amy) Hsiao

Introduction

In higher education, and more specifically at universities, learning begins with gaining knowledge. Therefore, courses are often structured in such a way that knowledge is the central course learning goal. However, knowledge alone is not enough. University degree programs also have a clear ambition in the development of skills. Embedding skills into a curriculum is more challenging, yet through carefully designed activities, such as case studies, tutorials, and group assignments, students can gradually learn to put theory into practice. Nevertheless, intended program **learning outcomes** of programs do not stop at acquiring knowledge and skills. The broader aim is to develop competences which are generally described as the combination of knowledge, skills, and attitude to solve a problem (Baartman et al., 2007).

Unlike knowledge and skills, competences are more difficult to foster within the traditional teaching formats like lectures and tutorials. They demand experiential learning, (critical) reflection, and practice in authentic settings that cannot easily be simulated in a standard classroom environment. Moreover, essential competencies such as self-confidence or the ability to reflect are difficult to measure and assess objectively, yet they are essential for both academic success and personal growth. That is why shifting the focus from grading outcomes to supporting the learning process itself often proves to be the most effective approach to fostering the development of competences.

At Tilburg University, the ambition to foster competences is closely connected to Tilburg University's educational profile (TEP): we want our students to develop knowledge, skills, and character (Tilburg University, 2017b). While knowledge and skills are strongly embedded in the curricula, the element of character is supported in all bachelor's program at the Tilburg School of Economics and Management (TiSEM) through the course Competence Compass. In this chapter, we describe how this course contributes to students' self-awareness regarding the extent to which they master a set of competencies by shifting the focus from grading outcomes to supporting students' learning processes through experiential and reflective activities.

Course Overview

Competence Compass is a mandatory 3-ECTS (European Credit Transfer and Accumulation System) course in the first block of the second year. It is included in all bachelor's programs at TiSEM, namely Bedrijfskunde, Economics, Econometrics and Operations Research, Economie en Bedrijfseconomie, Entrepreneurship and Business Innovation, Fiscale Economie, and International Business Administration.

This course aims to increase students' awareness of a set of competences, why they matter, and how they can be developed. The emphasis is not on reaching specific levels of competence, but on cultivating self-awareness that will help students succeed in their studies and build a meaningful professional and personal life. The competences are

divided into three clusters, and students choose from each cluster one competence, based on a self-reflection tool. The competencies are derived from a framework from Blijlevens (2019). The clusters are:

1. Resilience and self-regulation: 1) self-confidence, 2) setting and monitoring boundaries, 3) optimal balance study/work and personal development, 4) reflection on development and performance.
2. Goal setting and performance: 5) working in a goal-oriented manner, 6) focusing attention, 7) planning, 8) process-oriented working, 9) performing under pressure.
3. Problem solving and communication: 10) communicating, 11) making decisions, 12) problem-solving ability, 13) perseverance, 14) adaptability.

The course is structured as a personal learning journey in which students engage in experiences and reflect on them systematically. Reflective journaling forms the core of this process. Students are asked to write down their reflections and then articulate them in a short video. This cycle is repeated three times during the seven-week course, creating a rhythm of continuous learning and self-reflection. At the same time, we recognize that students vary in how they prefer to approach reflection. The course design therefore also includes additional challenges to encourage those who prefer not to follow a staged process to still engage meaningfully with reflection.

Course Learning Goals

Upon successful completion of this course, the student is able to:

1. Self-assess and reflect on his or her competence levels using specific tools (ILO4); and
2. Demonstrate and account for his or her current competence levels and development plan (ILO4 and 7).

The course learning goals (CLGs) contribute to two out of seven ILOs in the bachelor, namely:

- ILO4 – Pursue personal development to enhance professional competencies through guided reflection and institutional support, critically evaluating one's own qualities and interests and translating these insights into well-considered academic or professional choices.
- ILO7 – Take responsibility for oneself, close others, and broader society by making informed decisions, and by planning and executing projects, independently or collaboratively, in dynamic, unpredictable, and international environments, while sharing responsibility for managing complex processes and upholding societal and scientific obligations.

The CLGs in this course are assessed on a pass-or-fail basis across several assignments. The underlying assumption is that students learn most effectively through practice and repetition, which is why students are asked to engage in reflection three times during the course. Since the CLGs emphasize personal development and progress over time, a pass-or-fail system is most appropriate, as growth is difficult to quantify in a reliable way. The learning goals focus on each student's own development, aiming at Bloom levels apply and analyze (Ching & Da Silva, 2017). Table 3.13.1 shows how the course is assessed.

Table 3.13.1
Assessment Components

Assessment	Grading
Self-reflection tool	Pass or fail
Portfolio competency 1	Pass or fail
Portfolio competency 2	Pass or fail
Portfolio competency 3	Pass or fail
Portfolio of 3 competencies + additional reflection (only if you missed (one of the) earlier deadlines)	Pass or fail

Teaching, Supervision & Learning Activities

The course is designed as an online module with several checkpoints to guide students through a structured process of reflection and personal development. Learning activities are deliberately integrated with assessment components. The students are required to submit four intermediate assignments.

The course starts with an online module in the University's digital learning environment. In this module, students are introduced to the fourteen competencies. Here the goal is to gain knowledge on the competencies. For each competency information is given on 1) what the competency means, 2) three levels of the competency, and 3) how to develop the competency. To provide flexibility and accessibility, information is offered in multiple formats such as text, videos, and slides. Following this introduction, students complete the first assignment for which they have to use a self-reflection tool. For each competency, six descriptive statements are provided, from which students select the statement that best reflects their current situation. Table 3.13.2 shows the descriptive statements of competency self-confidence.

Table 3.13.2
Descriptive Statements of Self-Confidence

Statement		Choice
1	I often doubt my abilities and hesitate to start tasks or speak up	
2	After small wins I feel confident, but in new or demanding situations I quickly question myself.	
3	In familiar tasks I trust my skills, though I can still feel unsure when things get tough.	
4	I trust myself in most situations and take initiative, even when I face criticism or setbacks.	
5	In challenging situations or uncertain situations, I stay confident, accept mistakes, and remain assertive, though I may still feel vulnerable.	
6	Under pressure I remain confident, openly own mistakes, stay assertive, and handle tough situations without doubting my worth or skills.	

Based on the outcomes of this exercise, students select one competency from each cluster on which they will focus during the remainder of the course. For each chosen competency, students must provide a reasoned explanation. They are encouraged to select the competency that supports their personal development goals and future needs. This first stage should be completed within two weeks.

In the subsequent five weeks, students complete three portfolio assignments, one portfolio per competency. These portfolios are submitted in week 4, 6, and 7. For each portfolio, students select one from six possible assignments with the remark that each assignment can only be used once. This structure is designed to ensure diversity of learning experience and to encourage students to engage with their personal development from multiple perspectives. Table 3.13.3 gives an overview of the activities in the course and the estimated number of hours.

By following the schedule with the indicated deadlines, students keep an overview, learn throughout the process, carry that learning into the next competency, and avoid leaving everything until the last moment. This makes their development concrete, visible, and manageable. If the student does not submit anything at one of the interim deadlines, they are required to hand in one complete portfolio with an additional reflection at the end of the block.

Table 3.13.3*Estimated Time Investment*

Activity	Number of hours
Watch introduction video, read syllabus, plan deadlines, etc.	2
Read, watch, and/or listen to course materials and take notes and fill out self-reflection tool cluster 1	4
Read, watch, and/or listen to course materials and take notes and fill out self-reflection tool cluster 2	4
Read, watch, and/or listen to course materials and take notes and fill out self-reflection tool cluster 3	4
Online module information skills + assessment	6
Portfolio competency 1 from cluster 1	16
Portfolio competency 2 from cluster 2	16
Portfolio competency 3 from cluster 3	16
Additional reflection, resit, repair	16
Total	84 hours

Assessment Components

There are four assessment components: a self-reflection tool and three portfolios. In the portfolio, the students report on an assignment that fits the competency. There are six assignments:

1. Dare to try something new – practice a realistic and personal action that you find challenging but that you can carry out within one week. The activity is done while being observed by someone else, followed by reflection.
2. Planning in practice – creating a weekly schedule, living according to it, and reflecting on what went well and what was difficult.
3. Focus with the Pomodoro method – a technique that builds sustained focus through timed study cycles. Students practice and reflect on their experience.
4. Analyze your behavior – monitoring one's own behavior for five days, such as study habits or social media use, and reflecting on the outcomes.
5. Learning from practice – asking three people to share what they observed as personal strengths and areas for improvement in a past situation.
6. Design your future – speaking with three people who can provide insight into future options, such as studying abroad or choosing a master's program, and reflecting on what is learned.

Although the assignments given to students are diverse, building their portfolio follows a standardized process. Each portfolio consists of four phases: plan, action, analyze, and reflect. While the scaffolding questions in the planning and action phases differ across portfolios, the analyzing and reflecting phases remain the same. This set up ensures that while the practical experiences vary, the reflective practice itself develops in depth over time. In doing so, the course not only engages students in meaningful experiences but also systematically strengthens their capacity for reflection. In the analysis phase students answer the following questions for each portfolio:

- What similarities, differences, and recurring patterns do you observe based on the information you worked with?
- How did you feel during the process of completing the assignment, and what stood out to you (unexpected events, challenges, emotional reactions)?

In the reflection phase students answer the following questions for each portfolio:

- Based on your experiences while working on the assignment, which level of this competency best describes you, and why?
- What insight from this experience will you take with you in your future development?

Students who did follow the proposed schedule are required to submit an additional assignment to maintain a continuous learning path. This assignment mirrors the portfolio structure consisting of the four phases: plan, action, analyze, and reflect. In the plan and action phases, students work with slightly adapted scaffolding questions that prompt them to revisit their original self-reflection tool. This connects them with earlier insights and encourages evaluation of their own learning approach. Since the self-reflection was submitted in week two and the additional assignment is due in week seven, this creates a natural moment for reflection.

The assignment then looks forward: students select one competency they wish to further develop and create a personal development plan. This plan must include at least one activity within the university (for example, a workshop, project, or guidance session) and one activity outside the university (such as volunteer work, a sports club, a part-time job, or another personal challenge). Students discuss their plan with someone from their environment, seeking feedback on its feasibility, relevance, and possible improvements. These conversations deepen reflection, enhance self-awareness, and help to translate learning into concrete, future-focused action, restoring continuity in their reflective journey. The additional assignment concludes with the same analysis and reflection questions as the other portfolios.

Assessment Criteria

The course is assessed on a pass-or-fail basis. To complete the course, all separate assignments must be passed; if one assignment is not passed, the entire course is considered failed. To ensure **fairness** and continued learning, resit and repair opportunities are available, though these differ depending on the situation. Table 3.13.4 shows the **rubric** of the self-reflection tool. We focus on whether students have addressed the required points, and less on the quality of the content, because competence development is a personal process that cannot be strictly judged as right or wrong.

Table 3.13.4
Rubric Self-Reflection Tool

Criteria	Description	Rating		
		Complete	Incomplete	No submission
Self-reflection tool is fully filled out.	The statements are selected and there is an explanation why the competency is chosen.			

All portfolio assignments are evaluated with the same rubric displayed in Table 3.13.5. The portfolio serves as a structured reflection of the student's engagement with an assignment and their broader learning process. Assessment focuses on whether the student has addressed the required points and demonstrated a serious commitment to the task, rather than solely evaluating the quality of the content. This approach aligns with the educational goal of fostering competency development, which is an ongoing and personal process. By emphasizing effort, reflection, and meaningful engagement, the aim is to encourage students to take ownership of their learning and support their growth.

Each portfolio is scored on a scale from 0 to 2, where a score of 0 indicates that the portfolio is missing or that it cannot be considered as a serious attempt. When a portfolio receives a score of 1, it is considered a serious attempt, and the student may be required to submit a repair assignment for the specific insufficient element, while the remainder of the portfolio remains valid. A score of 2 is considered sufficient/pass. This approach allows students to continue progressing through the reflective learning process without interruption, thereby motivating them to engage meaningfully with their assignments from the start while also providing the flexibility to address minor shortcomings without compromising the overall development trajectory.

Table 3.13.5
Rubric Portfolio

Criteria	Description	Rating	
The portfolio contains all the required components.	If this criterion is not met, the assignment will not be reviewed further. This is a knock-out criterion in which we check for completeness of 1) certain length of answers, 2) requirements in assignment, 3) access to video, 4) inclusion of technology statement.	Fulfilled	Not fulfilled
Plan	The portfolio clearly explains why the competency was chosen.	Sufficient	Insufficient
Execution	The portfolio provides a clear picture of the activity undertaken.	Sufficient	Insufficient
Experience	The portfolio shows how the student experienced the activity and how this contributes to competency development.	Sufficient	Insufficient
Analysis	The portfolio examines underlying causes of the experiences and links them to the student's competency level.	Sufficient	Insufficient

Reflections on Challenges and Adaptations

This self-study course is students' first opportunity to work independently and responsibly, as well as to self-assess and reflect on their own development at the middle stage of the curriculum. Since this is a mandatory course in all TiSEM bachelor's programs, around 800 students participate in this course simultaneously. We were pleased to see a high level of engagement among students with the course materials. Many students submitted their assignments on the first attempt, demonstrating that they took the course seriously and met the requirements on time.

Given the large and diverse student group, clear communication about course requirements, process, and deadlines is particularly challenging. This diversity includes differences in nationality, cultural background, language proficiency, academic experience, and learning styles. Each student may interpret instructions differently based on these factors. While such diversity greatly enriches the learning environment, it also means that extra care must be taken to ensure clarity, consistency, and shared understanding of expectations so that all students can engage successfully with the course.

In this course, we address the issue of late submissions through a structured and transparent system. Rather than relying on subjective, negotiation-based decisions, students who submit work late are required to complete an additional reflection assignment. Our experience suggests that late submission policies should move away from lecturer discretion and toward clear, consistent rules that include both incentives and consequences. Such an approach ensures **transparency** and fairness, so all students know in advance what will happen if they miss a deadline, eliminating the need to request special consideration or extensions. This also reduces uncertainty and the perception of bias in grading or decision-making. Furthermore, the policy emphasizes student responsibility and autonomy. Students have the freedom to manage their own time and choose whether to submit on time or take an extension, understanding that doing so comes with an automatic, constructive cost (i.e., the additional reflection task). This system encourages student planning while maintaining fairness and accountability.

Students are allowed to use generative artificial intelligence (GenAI), and they need to report on their use in a technology statement. If GenAI is not used, they need to report on it as well. It is likely that students make use of AI tools, however, they do not explicitly recognize it as such. Some may believe that using certain AI-based resources does not count as AI use, for example spelling and grammar checkers, translation programs, and search engines. As a result, students neglect to provide a technology statement and a statement about AI use in their work and fail to meet the criteria for whether the portfolio contains all the required components. In general, this highlights the need

for clearer communication and guidance on what constitutes AI use and the importance of transparency, both for ethical reasons and for fostering responsible academic practices.

Grading reflection assignments brings examiners into the personal thinking and experience space of students which is a space we normally do not enter. Although this is not a problem, it is important to acknowledge, because it requires extra care, empathy, and respect for the vulnerability students may show in their reflections.

Student feedback revealed some unease. Students reported that they were accustomed to courses focusing primarily on knowledge, and therefore found this course, which is centered on reflection and personal development, unfamiliar. While they appreciated the clarity of the assignments and the manageable workload, some students felt that their experience so far in their studies did not fully meet their expectations. Our observations suggest that the issue lies not in the course structure or difficulty, but in its framing within the curriculum. Providing students with a clearer explanation of the value of reflection, its connection to their broader academic and professional development, and its role in supporting the university's mission of character formation may enhance their understanding and engagement.

3.14

Internship Assessment

Jolijn van Cruchten and Louise van Hoek

Introduction to the TiSEM Internship

Every year, approximately 100-120 students from Tilburg School of Economics and Management (TiSEM) choose to do a curricular internship. This so called *TiSEM internship course*, introduced in 2019, is a 24 ECTS (European Credit Transfer and Accumulation System) elective in the first semester of the third year of the Bachelor's programs in Business Administration, Economics, Economics and Business Economics, and International Business Administration. It is one of three options in the Mobility Window, alongside an international exchange and a minor. The internship offers students practical experience, preparing them for the job market and helping them choose a master's program. For many students, it is their first exposure to a real business environment.

The internship's six learning goals cover many of the program intended **learning outcomes** of the bachelor's programs. By analyzing company documents, students connect their academic knowledge to organizational reality and refine their information-literacy skills; by planning and executing day-to-day tasks, they practice responsibility. The project at the core of the internship provides methodological training in a business context, requiring students to frame questions, gather evidence, and formulate sound recommendations. Structured reflection converts experience into self-knowledge, while working in (often multicultural) teams deepens communicative competence. Together, these activities help students grow knowledge, skills, and character in a professional environment.

Teaching, Supervision & Learning Activities

Internship preparation starts in the semester before the course begins. Students submit an Internship Proposal outlining their intended project and internship placement. This proposal is assessed by the Internship Coordinator to ensure both the internship in general and the project meet key criteria, including practical relevance, opportunities to apply academic knowledge and skills, and a balance between day-to-day tasks and project work. Students can only start the internship once their proposal has been approved.

The internship itself spans approximately five months. Students transition from the university setting to the workplace while continuing to engage in academic and reflective learning activities. The course is supported by a structured supervision model, overseen by the Internship Coordinator:

- The *university supervisor* meets the student four times, focusing on the link between practice and theory, choosing appropriate research method(s) and monitoring process and progress.
- The *company supervisor* provides day-to-day coaching and completes two formal -not graded- evaluations.
- *Career Services* supports personal development and reflection through a Kick-Off session, an individual coaching meeting, and a workshop of choice.

From the first week, students participate in a highly blended learning format combining workplace learning, online modules, Canvas-based activities, and in-person sessions. One of the central pillars of the course is self-reflection, which is supported through online journaling every three weeks and the development of a personal development plan based on individual learning goals.

Throughout the internship, students produce several key deliverables: a Work Report, a Project Report, and a Self-Reflection Report. These deliverables are shaped by academic supervision, company feedback, and the student's personal growth, together forming the basis for assessing achievement of the course learning goals.

Formative assessment

Formative assessment is embedded throughout the internship course as a key mechanism to support students' academic, professional, and personal development. Feedback is not limited to one role or moment – it is built into the entire structure of the internship. The university supervisor provides structured feedback at four key moments, helping students define their project, monitor progress, and reflect on academic quality. Company supervisors provide ongoing on-site coaching and conduct formal formative evaluations at the midpoint and end of the internship, offering the student insights into their workplace performance. Career Services supports personal learning with feedback through coaching sessions and reflective development.

Self-reflection is a central component of formative assessment in this course. Students begin by selecting two personal learning goals using a Self-Assessment Tool. This tool supports students in identifying the most relevant goals from a predefined list, based on their current level of competence and the perceived value of improving those skills. The two selected goals serve as the foundation for all subsequent self-reflection and related assessment components.

Another key tool in this process is the Self-Reflection Journal, completed every three weeks via Canvas. This journal functions as a low-threshold, ongoing check-in, prompting students to evaluate their learning goals, challenges, and progress. It enables them to identify patterns in their behavior, make sense of workplace experiences, and adjust their approach where needed. While the journal is not graded and unseen by any supervisor, it provides valuable and personal insights for the student and supports the creation of the final Self-Reflection Report by offering a continuous thread of reflection and insight.

Summative Assessment

The TiSEM Internship course includes three **summative assessments** that together reflect the learning goals of the internship. Each **assessment** is a deliverable that contributes to the final grade based on a fixed weight: the Project Report (60%), the Self-Reflection Report (30%), and the Work Report (10%). These assignments are designed to match the effort students invest, the depth of learning achieved, and the type of competencies demonstrated.

Work Report (10%)

The Work Report is the first written deliverable, submitted after the first 4-6 weeks of the internship. In approximately 2000 words, students describe the host organization's structure, culture, market position, and their own responsibilities within the firm. Setting an early deadline ensures that the student is required to take a deep dive into the internship organization. This helps them understand the project within the broader context of the company, the market, the consumer, and the product or service. It also provides the university supervisor with essential insights to help shape a meaningful project – both for the student and the company.

Project Report (60%)

As the core academic component, the Project Report (approx. 4200 words) documents a problem or opportunity identified by the company, explored through academic methods, and resulting in practical recommendations. The structure includes context and relevance, research question(s), methodology, results and recommendations, and conclusions and limitations. A signed timesheet confirming at least 336 hours of project work is mandatory; without it, the report is considered incomplete. Students receive supervision during four academic meetings, with only the fourth meeting dedicated to feedback on a draft version of the report. Interim progress is typically shared through presentations or short updates. The final report is submitted at the end of the internship.

Self-Reflection Report (30%)

In no more than 1200 words, students reflect on their development toward two Personal Learning Goals (PLGs) chosen at the start of the internship via the Self-Assessment Tool. An example of a PLG could be:

“By the end of my internship, I want to have given at least 3 presentations, where the structure and visual aids have supported my story in such a way that the audience found the message of my presentation clear.”

Each goal is explored from two perspectives; both the attainment of the goal and the process of working towards the goal should be reflected upon. This should be done by using the STARR method (Situation, Task, Action, Result, Reflection), based on real experiences during the internship. Sources for reflection include journal entries, insights derived from the Career Services sessions, and the formal company evaluations. This deliverable demonstrates the student's ability to critically reflect, apply feedback, and articulate personal growth.

Authentic Assessment and the Use of Generative Artificial Intelligence (GenAI)

The three internship deliverables are designed as authentic assessments, meaning they replicate the kinds of tasks graduates are likely to perform in the workplace while meeting the course learning goals. Students work on real company projects, apply academic and professional skills, collaborate with stakeholders, and produce outputs similar to reports expected from junior professionals. This approach ensures that students demonstrate applied research skills, critical reflection, and professional communication in a real-world context.

GenAI presents challenges for many forms of assessment, but its impact here is limited. Each deliverable requires evidence and insights that can only be gathered through the student's direct involvement, such as internal documents, first-hand observations, interviews with staff, and personal reflections from private journals. Furthermore, the nature of the projects is highly practical, which often requires direct implementation within the company. Close supervision by the university supervisor, company mentor, and Career Services further safeguards the integrity of the work.

That said, AI can still influence certain parts of the process, such as writing support or idea generation, if used without **transparency**. Students are therefore encouraged to use AI critically and responsibly, and to acknowledge its use where relevant. In fact, AI is playing an increasingly prominent role in many of our students' projects, sometimes even becoming the main topic. For more and more companies, the key question is how to effectively integrate AI into their operations and workflows.

Determining the Grade

All summative assignments in the internship course are assessed using the same 4-10 analytic **rubric** (see Table 3.14.1). Since university supervisors typically oversee at least five students, and sometimes as many as fifteen, report assessment can be a significant time commitment. To focus effort where it is most valuable, written comments are mandatory for outstanding (9-10) or insufficient (4-5) performances, and optional for grades between 6 and 8. This cuts feedback time by about a third yet still gives actionable notes where they matter most.

The final grade for each deliverable is based on the weighted average of its individual criterion scores. Each component must be passed individually (minimum 5.5) and missing knock-out appendices (timesheet, personal development plan/PDP or company evaluation) results in a fail. The final grade of the course is a weighted average of the three components. One resubmission per assignment is allowed, but resit grades are capped at 6.0.

Table 3.14.1
Grading Scale for All Internship Rubrics

10	9	8	7	6	5	4
Excellent	Very Good	Good	More than sufficient	Sufficient	Insufficient	Incomplete
Meets all descriptors at an extremely high level; extremely clear and well developed	Meets all descriptors at a high level; very clear and well developed	Meets most descriptors at a high level; clear and well developed	Meets all descriptors at a sufficient level; some lapses in clarity and development	Meets most descriptors at a sufficient level; sometimes unclear or under-developed	Does not meet descriptors at a sufficient level; often unclear or undeveloped	Meets no descriptors or the required element is missing

The final grade for each deliverable is based on the weighted average of its individual criterion scores. Each component must be passed individually (minimum 5.5) and missing knock-out appendices (timesheet, PDP or company evaluation) results in a fail. The final grade of the course is a weighted average of the three components. One resubmission per assignment is allowed, but resit grades are capped at 6.0.

Criteria for the Work Report

The Work Report is graded as the average of equally weighted criteria across three areas (see Table 3.14.2):

- a critical description of the company's origins, activities, markets, structure, strategy, competitors, and culture;
- an analysis of the department's role, activities, interdepartmental relations, and communication patterns;
- and a short account of the interns' non-project tasks, linked to departmental and company goals.

Key criteria include the ability to extract essential information from relevant sources and to recognize patterns in communication and decision-making. Strong reports integrate evidence from multiple sources into a coherent and critical analysis, showing how the company and department operate within their wider context. Weaker submissions tend to remain descriptive, miss essential details, or rely too heavily on uncritical source material, limiting the depth of insight.

Additionally, students may attach documents (e.g., organizational scheme, print screen of company website), but be selective: only those attachments that students think are useful to get a good understanding of the company should be included.

Table 3.14.2*Required Elements and Assessment Criteria of Work Reports*

Required elements	Assessment criteria
A critical description of the company, its origins, activities and markets the company is operating in, the industry or sector, number of employees, organizational structure, company's mission and/or strategy, main competitors, organizational culture, number of nationalities working in the company.	<ul style="list-style-type: none"> • Identify documents/sources relevant for analyzing the organization context. • Extract essential information from various formal sources to characterize the company's role, rules and purposes and its relations to other companies in the external environment.
The department where you did your internship: your critically describe and analyze its role and activities, the relations with other departments in the company, the organizational structure.	<p>Analysis based on (1) formal company documents and (2) observations and interacting with others</p> <ul style="list-style-type: none"> • Identify documents relevant for analyzing the department context. • Extract essential information from various formal sources to characterize the department's role, rules and purposes and its relations to other departments in the external environment. • Recognize the patterns in communication and decision-making.
A description of your non-project activities in the organization. Give a short description of the activities that you (are planning to) conduct, especially those that are not part of your project.	<p>Role and responsibilities of an intern</p> <ul style="list-style-type: none"> • Identify the general role and tasks as intern in the (1) department and (2) company. • Relate this role and these tasks to the overall goal of the department/company.

Criteria for the Project Report

The Project Report is graded as the weighted total of different criterion scores (see Table 3.14.3). Assessment covers the project's background and context within the organization, the clarity and measurability of its goal and central question, the appropriateness and justification of the chosen methodology, the quality and applicability of results and recommendations, the conclusions and limitations, and the integration of academic theory with practice. Key criteria include the ability to clearly position the project in the company's structure and strategy, to justify methodological choices, and to provide evidence-based recommendations of practical relevance.

High-quality reports present a logically structured, methodologically sound investigation with conclusions supported by robust evidence and clear links to both company needs and academic learning. Weaker reports tend to offer vague project framing, insufficient justification of methods, limited analysis, or recommendations lacking practical applicability.

Additionally, a knock-out criterion is set. Students need to submit a Timesheet showing: (1) at least 336 hours were spent on the project, and (2) 272 'extra' hours were spent on the project or on other work related activities.

Table 3.14.3
Required Elements and Assessment Criteria of Project Reports

Required elements	Assessment criteria
Describe the background and context of the project, as well as positioning of the project within the organization. (20%)	<ul style="list-style-type: none"> Clearly articulate the background and context that led to the identification of the project, including a detailed description of the circumstances or factors that gave rise to the 'problem' being addressed. Explain the project's position within the organizational structure, its strategic importance, and justify the urgency or significance in addressing organizational needs or goals.
The goal of the project and the central question. (5%)	The project goal and the central question are formulated in a clear, well-defined and measurable manner, from both company's and academic perspective (i.e., the project and central question can be linked to the study program).
Describe the methodology used for the project, detailing steps taken to achieve the central goal. Justify research methodologies chosen (e.g., desk research, stakeholder interviews, quantitative methods) based on relevance and suitability. Outline the roles of involved parties and collaboration mechanisms used for project progression and goal attainment. (25%)	<ul style="list-style-type: none"> The methodology used is clearly described with a logical sequence of steps contributing to the project's central goal. Research methodologies chosen are clearly justified based on their relevance and suitability to the project's goal. Roles of involved parties are clearly defined, specifying their contributions to project progression and goal attainment.
The current status of the project: what were the research outcomes? What advice and/or final results were provided? (25%)	<ul style="list-style-type: none"> Assess research outcomes with sufficient detail to convey key insights and conclusions. Provide clear, relevant advice supported by evidence and integrating academic insights effectively. Recommendations should demonstrate practical applicability, illustrating their impact and relevance to stakeholders or the project's context.
Conclusions & Limitations. Did the project meet the expected outcomes? (15%)	Evaluate the attainment of the project goal by identifying contributing activities, tasks, and any flaws in the approach used.
Link theory to practice: A description of the project in relation to the curriculum of your study program. To which courses did it relate? During your internship, did you observe things that were similar or rather very different from what you were taught in class? (10%)	<ul style="list-style-type: none"> Analyze how the project applies academic knowledge from relevant courses in the study program. Evaluate the practical implementation of theoretical concepts and academic skills during the internship, noting any observed deviations from classroom learning.

Criteria for the Self-Reflection Report

The Self-Reflection Report is also graded as the weighted total of criterion scores (see Table 3.14.4). Assessment focuses on the overall level of critical reflection, the use of the STARR method to describe at least two situations linked to the student's PLGs, the integration of insights and feedback into personal development, and the final evaluation of progress on each PLG. Key criteria include the ability to analyze experiences in depth, explain behaviors or outcomes using personal insight and prior learning, and demonstrate the application of feedback from Career Services, company supervisors, and official evaluations.

Strong reports provide a clear, deep reflective narrative that links specific experiences to measurable personal growth, showing how feedback and insights were actively incorporated. Weaker submissions may remain superficial and factual, fail to connect experiences to learning goals, or overlook how feedback informs development.

Additionally, a knock-out criterion is set. Students must submit a (1) development plan and (2) company evaluation form.

Table 3.14.4*Required Elements and Assessment Criteria of Self-Reflection Reports*

Required elements	Assessment criteria
Overall level of critical reflection	<ul style="list-style-type: none"> explores or analyses events or experiences, asks questions and looks for answers, considers alternatives, speculates or hypothesis about why something has happened. explains their own or other's behavior or feelings using personal insight, inferences, experiences or previous learning, demonstrating a depth of understanding.
A clear description – by the means of the STARR method – of at least 2 specific situations or events, related to the personal learning goals, that have impacted the student's self-development	<ul style="list-style-type: none"> describes why the situation is relevant to the learning goal. outlines how the situation contributes to the overall development-process of this specific learning goal, describing the actions that led to this situation and/or actions that followed as a result.
A clear description of how specific insights have impacted the student's self-development on each learning goal	<ul style="list-style-type: none"> describes how they have applied insights from the three Self-Reflection sessions to their work context. has integrated the feedback received from others (e.g., peers, colleagues, supervisors). incorporates the official company feedback, taking appropriate actions to deal with it, which contributes to the final evaluation in line with the feedback received.
The final evaluation	<ul style="list-style-type: none"> evaluates their final performance against the established learning goals, reflecting on both

Reflections on Challenges and Adaptations

A 24-Credit Course: Balancing Deliverables and Deadlines

An internship of 24 credits equals 672 hours, making it 'heavier' in weight than a thesis.

The possibility of failing this course can feel highly consequential for students. However, in practice, course failure is rare due to close supervision and continuous monitoring throughout the internship period. When students do fail, it is typically because they discontinue the internship or become so immersed in the company's day-to-day operations that they lose focus on their project.

To help ensure that students are on the right track from the outset, early assessment components have been built into the course. The internship proposal students must submit prior to the semester starts, makes sure that students already have at least a broad idea of the project that they will be working on. Stronger project descriptions (well demarcated, clear focus and steps) tend to translate into better performance throughout the internship. Therefore, a second stage 'early' assessment has been added to the course.

In the first month, students submit their work report (10% of total grade) and project proposal (formative). The work-report pushes students to deep dive into their organization at the start, in order to accelerate knowledge and understanding of company stakeholders, processes, tools, etc. Grades are typically around 7, which is perceived by students as 'low'. Students tend to underestimate this 10% report. Therefore, it is often seen as a 'wake-up call' and sets the bar of the university's expectations.

The project proposal should reflect the project as discussed with the company and the university supervisor. In the first meeting, the university supervisor helps the student to clearly demarcate the project, define steps, and discuss possible research element(s). If a student receives a "no go" on their project proposal – even after incorporating feedback – they may still proceed, but this functions as an early signal that the project does not yet meet academic standards.

Although the course is designed as such that students receive early feedback, many students experience the number of deadlines as quite intense. While students are required to submit three summative reports, the total number of

deliverables is much higher, including items such as the personal development plan, self-reflective journaling, company evaluations, and time sheets.

In course evaluations, for example, students indicated that while they appreciated the journaling activity, it felt excessive. As a result, the journaling frequency has been reduced from bi-weekly to every three weeks.

Increasing Inter-Rater Reliability

Inter-rater reliability among university supervisors remains a key concern, given the large number of supervisors and diverse internship projects. Many academic examiners feel uncertain about their assessments, as they are often less familiar with evaluating more practical reports. This makes calibration sessions essential (Bloxham et al., 2016; Hsiao et al., 2023).

Before the final assessment, calibration sessions are held with the goal of establishing a shared baseline for grading. Before attending, university supervisors are asked to complete a short homework assignment: grading three old (anonymous) reports using the existing rubric and providing comments to support their scores.

These assessments are collected in advance, and commonalities and discrepancies are identified as a basis for discussion. During the session, these points are reviewed, and agreements are made on how to interpret specific criteria. All decisions and interpretations are documented in an Assessment Guideline.

External Supervisors and Grading

According to general university policy, external supervisors are not allowed to officially grade student performance. This is due to concerns about the **reliability** and consistency of grading, as company supervisors come from widely varying backgrounds – even though we require them to hold either (1) a master's degree or (2) at least ten years of relevant work experience.

Despite this policy, company supervisors often have the clearest view of the student's actual performance, as they work with the intern on a daily basis. They observe real behavior and outcomes, rather than relying solely on written reports. In light of this, many students perceive the grading policy as unfair.

Currently, company evaluations are used as input for the self-reflection report, but ideally, we would involve the assessment of the company in a more meaningful way. Calibration sessions could help address concerns around consistency. However, with such a large number of company supervisors involved, organizing these sessions remains a practical challenge still to be solved.

3.15

Bachelor Thesis Assessment – Creating a collaborative research experience

Astrid Kramer and Miranda Stienstra

Introduction

Universities in Dutch higher education increasingly face challenges in bachelor's thesis **assessment**, a difficulty intensified by the rise of GenAI. The bachelor's thesis serves not only as the culmination of a student's academic development but also as an indicator of whether the **learning outcomes** of a bachelor's program have been achieved (NVAO, 2018). Despite its pivotal role in demonstrating academic competence and program quality, institutions struggle to design efficient and effective thesis trajectories and assessment frameworks. This challenge is exacerbated by growing uncertainty regarding the impact of generative artificial intelligence (GenAI) on the **validity** and originality of student work. Moreover, the absence of clear guidance from the NVAO on thesis assessment leaves institutions to interpret quality criteria independently (Ministry of Education, 2021).

Students often experience writing a bachelor's thesis as demanding and overwhelming, as it is their first substantial independent research project. Developing a well-defined problem statement and translating theoretical insights into a coherent research design require a solid understanding of existing literature, methodological awareness, and critical thinking. Consequently, students may struggle to see how their work aligns with the intended relevance, rigor, or quality standards of academic research.

To address these challenges, this chapter proposes a redesigned bachelor's thesis process in the bachelor Business Administration at Tilburg School of Economics and Management at Tilburg University. The new approach conceptualizes the thesis as a collective developmental process, supported by peers and mentors within structured thesis circles. The redesign aims to strengthen students' capacity to co-create relevant knowledge, mitigate common group-work pitfalls such as free-riding and uneven contributions, and position AI as a partner in co-creation rather than a substitute for learning. The redesign pursues three objectives:

1. To align the capstone project more closely with bachelor-level outcomes by shifting from solitary work to collaborative research, thereby reducing uncertainty and enhancing feedback opportunities.
2. To increase relevance through supervisor-initiated research projects.
3. To preserve learning authenticity by acknowledging the GenAI-rich environment and adapting assessment criteria to evaluate students' own reasoning and contributions.

Course Overview

The BSc Thesis Business Administration (12 ECTS, European Credit Transfer and Accumulation System) is the capstone project in the bachelor Business Administration and is typically written in the last semester of the program. The graduation project has a coordinator and academic staff as supervisors. Supervisors should have at least a master's degree. Around 200 students participate in the course and are supervised in groups of five students. Each student writes an individual thesis, with peer and supervisor feedback distributed across four meetings. Students are allowed to start with the BSc Thesis Business Administration if they have passed all courses in the first and second year of the bachelor which equals 120 ECTS.

Course Learning Goals

A bachelor's thesis represents the culminating academic experience of a degree program. Beyond exploring a specific topic in depth, its primary purpose is to enable students to apply previously acquired knowledge while developing problem-solving skills, learning and development skills, and communication and information skills. The learning goals are formulated as follows:

On successful completion of this course, students will be able to:

1. Design a research project that addresses a real-world business or societal problem, applying relevant theories and methodologies.
2. Locate or generate appropriate data, select and apply suitable analytical techniques, and interpret findings in relation to the research questions.
3. Produce a coherent written thesis, present and defend the underlying arguments in a small-group session, and provide constructive peer feedback.
4. Critically evaluate their role in collaboration, including data collection and feedback while demonstrating responsible and ethical use of GenAI.

The course learning goals (CLGs) contribute to all program intended learning outcomes (ILOs) in the bachelor and are displayed in Table 3.15.1. The course learning goals are to foster academic and professional skills. Executing the bachelor's thesis that addresses a real-world business or societal problem demonstrates the integration of disciplinary, methodological, and contextual knowledge (ILO 1) and the ability to apply this knowledge to analyze complex, unpredictable challenges (ILO 2, 3). It also reflects an awareness of ethical and societal responsibility (ILO 7). CLG 2 emphasizes methodological proficiency and the capacity to conduct scientifically rigorous and ethically sound research (ILO 1, 2, 3, 5). Producing a coherent written thesis, presenting and defending arguments, and engaging in peer feedback aligned with scholarly synthesis and communication skills (ILO 5, 6), strengthening critical reflection and collaborative competence. Finally, critically evaluating one's own role in collaboration, including the responsible and ethical use of generative artificial intelligence (GenAI), promotes self-awareness, professional growth, and adherence to integrity and accountability in research and teamwork (ILO 1, 4, 6, 7).

Table 3.15.1*Link Course Learning Outcomes to Intended Learning Outcomes*

Category	ILO	CLG
Knowledge	1. Demonstrate cohesive insight by organizing and integrating different types of advanced knowledge in the following areas: <ul style="list-style-type: none"> • Disciplinary knowledge: Accounting, finance, human resource management, information management, marketing, organizational behavior, strategy, and supply chain management. • Methodological knowledge: qualitative and quantitative research methods (e.g., survey design, OLS regression, interviews, case studies). • Real-world context and contemporary issues: corporate law, economics, ethics, and philosophy. • Scholarly and ethical standards: character strengths (honesty, transparency, integrity, fairness and non-discrimination laws and standards in business and professional ethics (including responsible use of AI). 	1, 2, 4
Skills: applying knowledge	2. Apply constructed knowledge along with established methodologies, techniques, and procedures to analyze and solve business cases within professional or research contexts.	1, 2
Skills: problem-solving skills	3. In unfamiliar and dynamic environments, identify and analyze complex and unpredictable problems in disciplinary or professional practice, and solve them without a predefined plan by developing tactical and strategic approaches and effectively selecting and using relevant data.	1, 2
Skills: learning and development skills	4. Pursue personal development to enhance professional competencies through guided reflection and institutional support, critically evaluating one's own qualities and interests and translating these insights into well-considered academic or professional choices.	4
Skills: information skills	5. Synthesize quantitative and qualitative information from a focused range of (non-) academic sources in a scientifically responsible, ethical, and legal manner to investigate and understand phenomena in business administration and effectively deliver the synthesized results through diverse presentation modes suited to the intended audience's needs.	2, 3
Skills: communication skills	6. Communicate effectively in Dutch/English with peers, colleagues, supervisors, specialists and non-specialists, using appropriate academic, managerial, or professional conventions relevant to the context.	3, 4
Responsibility and independence	7. Take responsibility for oneself, close others, and broader society by making informed decisions, and by planning and executing projects, independently or collaboratively, in dynamic, unpredictable, and international environments, while sharing responsibility for managing complex processes and upholding societal and scientific obligations.	1, 4

Integrated Approach to Supervision, Assessment, and Feedback

In the bachelor's thesis Business Administration, students formulate a problem statement, conduct research, and reflect on their conclusions in a written report. In one semester (14 study weeks), which equals a workload of 20 hours per week, each student works under the supervision of a faculty member.

Supervision takes place in thesis circles. Thesis circles offer a structured collaborative setting in which students, coached by an academic staff member, work on a thesis project within a shared research theme (Romme & Nijhuis, 2002). The coach aims not only at guiding individual thesis writing but also at fostering teamwork, peer feedback, and collective progress. Students depend on one another's work, function as a self-managing team, organize sessions, plan deadlines, and support each other. This approach transforms the bachelor's thesis from a solitary task into a collaborative, structured, and meaningful research experience for both students and their coaches.

To maintain fairness, thesis circles require a shared framework between supervisors. For example, in projects based on surveys, students work with a standardized and approved questionnaire prepared by the supervisor, as most

bachelor students lack sufficient expertise in instrument design. Similarly, in qualitative projects, students may analyze existing interview transcripts rather than conduct interviews themselves, though participation in selected interviews can be considered. Exemplary cases are displayed on the next two pages.

In all cases, supervisors provide clear formats and guidelines for data collection and analysis. Students find research methods challenging, because these skills are not always developed fully in the earlier curriculum (which focuses mostly on knowledge and applying knowledge, and this knowledge is more about theoretical concepts, not procedural knowledge of methods how-to). Each thesis follows the structure of a scientific paper: introduction, theoretical framework, methods, analysis, and discussion. While students share the same project and dataset, they develop individual research questions and models, ensuring both originality and comparability.

Exemplary Cases Bachelor Thesis

Julia participates in questionnaire research on entrepreneurial well being

Julia, a Business Administration student, joins her supervisor's long-term research on entrepreneurial wellbeing. The project has 500 survey responses, but another 250 are needed, and each student collects 50. Julia develops her own conceptual model with an independent, dependent, and moderator variable plus controls. Access to the full anonymized dataset allows her to test hypotheses and gain hands-on research experience, contributing to the supervisor's larger project while collaborating with peers.

Sean analyses company documents on CSR reporting

Sean, an Economics and Business Economics student, joins his supervisor's EU-funded research on CSR reporting. He analyzes 50 annual reports from Spain, while peers cover other European countries. Using these reports, Sean develops his own conceptual model and examines variables such as environmental initiatives, gender, and GRI alignment. Contributing to the shared dataset, he tests hypotheses and gains hands-on experience, learning how to conduct research while supporting a larger, collaborative project.

Lara analyses interviews on patient's experiences in a healthcare pathway

Lara, an International Business Administration student, joins her supervisor's PhD research on patient experiences in a healthcare pathway. She analyzes 10 anonymized interview transcripts, focusing on satisfaction, communication, care efficiency, and emotional support. Contributing to this qualitative dataset, Lara develops her insights and gains analytical skills within a meaningful, collaborative research experience.

Sven evaluates tax filings

Sven, a Tax Economics student, joins his supervisor's research on corporate tax practices across Europe. He analyzes 50 annual reports and tax filings, examining variables like effective tax rate, tax incentives, and international subsidiaries. By contributing to the shared dataset, Sven interprets patterns and supports the larger project, gaining practical research experience.

Maya develops a model

Maya, an Econometrics and Operations Research student, joins her supervisor's research on food supply chains for processing tomato residues into soup. Using datasets from retailers, wholesalers, and logistics providers, she develops a quantitative model to predict bottlenecks and optimize distribution of surplus tomatoes. By contributing to this ongoing project, Maya refines her model while supporting the supervisor's larger study. Working within the shared dataset gives her hands-on research experience, enhancing modeling skills in a meaningful, collaborative project.

Students are supervised in groups of five across four meetings and receive feedback on their input from their supervisor and peers. In addition, they receive guidance on data collection, analysis, and writing up their findings. Meetings also include reflection on progress, collaboration, and responsible use of AI. An overview of the sessions:

- Meeting 1: Introduction to the research topic and supervisor, explanation of the chosen method.
- Meeting 2: Discussion about the students' individual areas and methods within the data collection.
- Meeting 3: Discussion of the students' results based on the shared data.
- Meeting 4: Feedback on thesis draft.

Producing the bachelor's thesis is a structured yet demanding research experience, requiring students to manage their time, integrate feedback, and work independently. Students meet weekly in self-organized groups to keep minutes, provide peer feedback, and write brief reflections, which are then discussed with their supervisor. To integrate the GenAI component, students are encouraged to critically reflect on their use of AI through a set of structured scaffolding questions. These reflections are discussed during student-led group meetings, allowing participants to share insights, challenges, and strategies. Key outcomes from these discussions are subsequently communicated to the supervisor to inform guidance and support.

Assessment Components

The bachelor's thesis has two assessed components. First, the bachelor's thesis is a scholarly report of the research and is graded on a 1–10 scale. Second, the process and reflection component, which evaluates the student's contribution to the thesis circle, the quality of peer feedback (providing at least one strong point and three areas for improvement), communication with the supervisor (e.g., ask early, listen well, turn feedback into progress), and responsible use of GenAI, is assessed on pass-or-fail basis and must be passed to complete the course.

For the scholarly report, there is a repair opportunity if the initial assessment is insufficient. However, in the resit, the maximum grade is 6.0. This cap reflects the added benefit of receiving additional feedback in the resit. The pass-or-fail component carries substantial weight. Failure to contribute to the team effort cannot be compensated by submitting an additional reflection afterwards. Students must recognize that the process cannot be retroactively demonstrated. The pass-or-fail requirement ensures sustained and genuine engagement throughout the thesis trajectory.

Assessment Criteria

The thesis is graded, with explicit attention to the specific context of each thesis circle. For instance, a circle working with raw interview data and coding complex conversations requires a different evaluation approach than one focused on recruiting sufficient participants for a standardized survey. Both tasks demand effort, creativity, and skills, but how these qualities translate into a final grade will naturally differ. The structure of the thesis circle must not influence a student's potential grade. Students analyzing interview transcripts receive the same grading opportunities as those conducting questionnaires, making it essential that all circles are carefully designed to ensure fairness and a level playing field. The BSc Thesis Business Administration **rubric** contains three weighted elements. Each criterion is graded on a scale of 4–10 and the final grade is a weighted average of the criterion scores (see Table 3.15.2).

Table 3.15.2*Bachelor Thesis Assessment: Criteria and Descriptors*

Criteria	Description	Weight %
1. Problem framing and theoretical argumentation	The student derives a focused problem statement and well-formulated research questions, and develops a coherent, literature-based argument that motivates clear expectations or hypotheses relevant to the real-world setting.	35
2. Empirical execution: data collection and analysis	The student carries out data collection and analysis to a professional standard, taking responsibility for dependencies on others in the circle (e.g., shared data-gathering) and demonstrating understanding of methodological assumptions, limitations, and checks that go beyond the basics.	30
3. Interpretation & applicability	Findings are interpreted in direct relation to the research questions and theoretical expectations, with realistic conclusions, clear limitations, and articulated implications for practice (business/organizational decisions) and for further academic research.	35
	Pass	Fail
4. Contribution to the circle (process and responsibility)	Attends consistently; takes ownership of tasks; delivers on time; contribution evident in minutes/artefacts; shared data-collection responsibilities documented.	Repeated absence or non-delivery; contribution not evidenced or misleading; no trace in minutes/artefacts.
5. Quality of peer feedback	Feedback is specific, evidence-based, action-oriented; professional tone; ≥ 2 substantive examples pointing to text/analysis with concrete next steps.	Vague/unsubstantiated comments; only praise or only criticism; unprofessional tone; insufficient or absent examples.
6. Reflection quality (insight & action)	Identifies what went well/poorly and why; formulates concrete improvement actions that are (being) implemented.	Merely descriptive/defensive; no cause analysis; no actionable steps.
7. Responsible AI use (disclosure & boundaries)	Transparent about what/why/how AI was used; safeguards quality/ethics/privacy; states what was not delegated; complies with course/faculty rules.	No disclosure; inappropriate/undocumented use; privacy/ethics breach; substitution of own reasoning/learning.

Reflections on Challenges and Adaptations

The bachelor's thesis circles effectively bridge education and research, strengthening the connection between student learning and academic output. In this model, students act as junior research assistants, contributing to ongoing projects while developing key academic and professional skills. The following reflections outline the main benefits and considerations for students, supervisors, and the development of AI literacy.

First, bachelor's thesis circles provide authentic research experience by engaging students in data collection and analysis that may contribute to publications or impact reports. This approach fosters methodological, analytical, and conceptual skills while promoting collaboration, accountability, and peer learning. Students also strengthen communication and critical reasoning through feedback, presentations, and reflection on research decisions. Challenges remain, as students are still developing research skills and may encounter difficulties in data collection, analysis, or conceptualization. Balancing independence with group responsibilities and coordinating within shared timelines can be demanding. However, these challenges are integral to learning, fostering resilience, practical research experience, and collaboration, preparing students for future academic and professional endeavors.

Second, supervisors involving students as junior research assistants offer clear benefits but also require careful management. Students enable larger scale data collection, enrich ongoing projects, and align teaching with research, allowing supervisors to serve both purposes. Students may also bring fresh perspectives and help to identify promising candidates for future research projects. Embedding students in ongoing projects that may result in data and

possibly conference presentations and publications strengthening the supervisor's academic career. Yet, as bachelor's students are still developing their skills, close guidance remains important. When well managed thesis circles effectively combine educational and research objectives. Although not all academic staff may wish to participate in the proposed setup due to their position and role, it remains important to integrate the university's dual mission of education and research within collaborative initiatives. Active engagement in research enriches teaching quality, fosters innovation, and strengthens the academic community. While it is possible for lecturers to focus primarily on teaching, this separation is undesirable in the long term. Therefore, alternative mechanisms should be established to ensure that all staff remain academically engaged and contribute to the continuous development of disciplinary knowledge and pedagogical excellence.

Finally, removing the bachelor's thesis to prevent potential misuse of GenAI may seem like a safeguard, but it would undermine an essential part of academic learning. Instead of avoiding the challenge, students should be guided to use AI tools responsibly and transparently within their research. When thesis topics align closely with a supervisor's area of expertise, it becomes much easier to assess authenticity and detect inappropriate or excessive AI use.

3.16

Doing Capstone Projects Differently at Entrepreneurship and Business Innovation

Wynand Bodewes and Werner Liebrechts

Introduction to the Dutch Higher Education Context

In the Netherlands, the Dutch-Flemish Accreditation Organization (NVAO, 2018) has established the assessment framework for the national higher education accreditation system. Standard 4 specifies that a program must demonstrate that its program intended **learning outcomes** (ILOs) have been achieved. Evidence for this includes assessment results, the quality of final work, and the extent to which graduates perform successfully in professional practice or in further study.

A non-exhaustive list of possible final works includes a thesis, a portfolio, a professional product, one or more examinations, an article, an artistic performance, or a combination of these. At Dutch research universities, final projects are mostly presented as a thesis reporting on the execution of an academic research project. However, when employability is considered, as reflected in graduates' functioning in professional practice or subsequent programs, a traditional thesis based solely on a literature-identified research gap may not always best demonstrate the required professional competencies.

In the Netherlands, business degree programs are often accredited not only by the NVAO but also by the Association to Advance Collegiate Schools of Business (AACSB)(<https://www.aacsb.edu/>). The AACSB likewise does not prescribe a specific format for graduation projects. Instead, it emphasizes that the chosen **assessments** must demonstrate the program's learning competencies in a valid way across all locations and delivery modes.

The Role of Capstone Projects in Dutch Higher Education

Around the world, the term capstone project is commonly used to refer to graduation projects. A capstone project is a culminating assignment that integrates the knowledge, skills, and competencies students have acquired throughout their program. Completed near the end of a degree, it serves both as a demonstration of academic mastery and as a bridge between university study and professional practice.

Often, capstone projects are assignments in which students are supposed to display mastery of most of the ILOs. Capstone projects vary across disciplines. In the social sciences, they often take the form of independent research. In the natural and life sciences, students typically join ongoing research projects aligned with their supervisor's

work. Because social science students usually choose their own topics, their projects may be misaligned with the supervisor's research pipeline. Given that not all graduates will pursue academic research careers, alternative capstone formats may better reflect the skills required in the labor market.

Additionally, capstone projects can be individual or team based. In engineering programs, they often involve industry collaboration, exposing students to real-world challenges, and allowing programs to assess professional readiness (Konrad et al., 2018). When capstones address authentic professional problems, they also help students develop a clearer professional identity (Kwan & Lo, 2023).

Capstone projects not only serve as learning experiences but also as tools for program evaluation, offering evidence of educational effectiveness and graduate preparedness (Ahmad & Alammery, 2022; Wefald et al., 2022). In Europe, where the master's degree is generally seen as the final degree, academic bachelor's capstone projects may signal readiness for a master's program. This helps explain why Dutch university bachelor's programs frequently require an academic thesis, because it reflects supervised research activities.

Capstone projects are central to undergraduate education, linking academic learning with professional application. They build essential skills and provide programs with valuable feedback on curriculum effectiveness. As higher education evolves, capstone projects continue to play a vital role in preparing graduates for the demands of their future professions.

Capstone Projects in the Bachelor Entrepreneurship and Business Innovation

In many academic fields, the bachelor's thesis is modeled on an empirical research paper, or, at times, a research proposal. This is common not only in business programs but also in economics, psychology, and sociology. The rationale is that graduates should demonstrate mastery of scholarly research, albeit at a level below that of senior researchers. This contrasts with fields such as engineering, where capstone projects often involve designing a product, machine, or instrument, emphasizing applied, practical outcomes.

The BSc Entrepreneurship and Business Innovation (E&BI) at Tilburg University aspires to educate future entrepreneurs and business innovators, yet combines that with a clear academic orientation, emphasizing research skills, critical thinking, and analytical abilities. Students engage with scientific literature, explore theoretical frameworks, and critically evaluate research, gaining a deep understanding of entrepreneurship and business innovations fields of research as an evidence-based for entrepreneurial endeavors (Intended Learning Outcome (ILO 1, 5).

Research skills are further developed through the application of research methods and analytical techniques, enabling students to conduct rigorous empirical investigations and draw meaningful conclusions (ILO 2, 3). Critical thinking is fostered by questioning assumptions, evaluating arguments, and analyzing complex issues from multiple perspectives (ILO 2, 3, 6, 7).

The program also emphasizes problem-solving, teaching students to identify unmet needs, validate value propositions, design innovative products or services, and develop actionable business models and strategies. By integrating academic knowledge with practical insights, graduates are assisted to develop relevant competences instrumental to implement effective solutions in dynamic business environments and reflecting on their own development (ILO 3, 4, 5).

Following the principle of **constructive alignment** (Biggs & Tang, 2011) at Tilburg University, the capstone is designed to align closely with the program intended learning outcomes. However, unlike other bachelor's programs at the Tilburg School of Economics and Management (TiSEM), which generally require students to conduct empirical research or develop a research proposal, E&BI capstone projects can take multiple forms: (1) Academic research (i.e., validating theoretical concepts or models), (2) applied research (i.e., addressing business challenges faced by external organizations), or (3) case development research (i.e., developing a business case including teaching notes).

Designing E&BI's Capstone Course

E&BI aims to actively involve stakeholders in curriculum design. In close collaboration with the Vice Dean of Education and the Associate Deans of Education of TiSEM, a 12 ECTS (European Credit Transfer and Accumulation System) course was developed, in which students design, execute, and report on their capstone project. The course assessment overview (see Table 3.16.1) outlines the course's learning goals and how they relate to the program intended learning outcomes.

Table 3.16.1
Course Assessment Overview

Course learning goals (CLGs) <i>On successful completion of this course, students will be able to:</i>	Intended learning outcomes	Proposal (0%) ☒ Pass required?	Thesis (100%) ☒ Pass required?
1. Present fitting conceptual frameworks or theoretical models by reviewing relevant theories and empirical evidence	1, 5	X	X
2. Judge methodological options to design a research project that would allow you to draw valid conclusions and/or develop sound solutions.	1, 2	X	X
3. Apply relevant qualitative and/or quantitative methods and analytical techniques to answer one or more research questions.	1, 2		X
4. Build on existing and one's own research to create insights that improve the theory and/or practice of entrepreneurship and/or business innovation	3, 5		X
5. Format deliverables in compliance with the style guide for student papers, including in-text citations and references, as specified by APA Style 7	6	X	X
6. Display a critical mindset that balances sometimes conflicting perspectives from relevant stakeholders in the execution of a research project.	4, 7	X	X

As mentioned, E&BI's capstone project takes one of the following three forms:

1. Academic thesis: Students conduct research that is to result in a deliverable that informs academics (i.e., the primary aim is to make a theoretical contribution).
2. Applied thesis: Students conduct research that is to result in a deliverable that informs practitioners (i.e., the primary aim is to make a practical/managerial contribution).
3. Case development thesis: Students conduct research that is to result in a deliverable that informs future E&BI students (i.e., the primary aim is to make a didactical contribution).

The course design is structured around four key themes, each representing a set of principles intended to address the issues identified in our analyses in the accreditation of the initial program design:

1. Scholarship and capstone types

- There are different types of scholarship, and each can represent an academic contribution that signals mastery of the program's learning outcomes.
- In the capstone course of the E&BI program, students can opt for three of Boyer's (1997) notions of the scholarship of discovery, the scholarship of engagement, and the scholarship of teaching.

Supervisors can choose from academic research (primarily informing academics), applied research (primarily informing practitioners), and case development research (primarily informing educators), reflecting the three types of scholarships mentioned under the previous point.

2. Supervision and circle structure

- Supervision is organized in circles of about ten students.
- Supervisors select the type of research projects to be conducted in that circle (academic, applied, or case development).
- Circles meet four times during a semester (in person), and no additional in-person supervision is provided beyond those four moments of contact.

3. Project methodology and domain

- Each graduation project is an individual assignment and requires independent work.
- Each type of research has a methodological textbook that the students must consult when designing and executing a project.
- Supervisors choose whether the students in their circle will use qualitative research methods, quantitative research methods, or a combination of the two.
- Supervisors define a domain within which the project should be framed, reflected in seminal papers that can serve as a starting point for a literature search.

4. Assessment & Student Assignment

- To ensure that the same ILOs and learning goals are assessed, all these are to be assessed using a single **rubric**, irrespective of the form of the capstone project.
- Students indicate their preferences for the complete set of circles provided during the semester, and an Excel-based tool assigns students to a circle while considering all relative preferences. As a result, almost all students are assigned to a circle of their first or second preference.
- When indicating these preferences, students do not know which faculty member will be supervising which circle. This ensures that students indicate their preferences based on the type of research and thematic focus of the circles, not personal knowledge of the supervisor.

The design principles mentioned above ensure that supervisors oversee projects that they are comfortable with (type of research, research methods, and research domain). This is a key difference from other capstone courses at TiSEM, in which students are typically given almost complete freedom to propose what and how they would like to execute their capstone project. The course coordinator is to ensure that, overall, the availability of circles matches the students' appetite (especially regarding the mix of academic and applied research-based projects). Usually, the cohort splits equally between academic and applied research circles, with the case development research-based circle being the least popular. However, those who choose the latter type of research are often highly committed to it.

Each thesis circle will have a clear thematic focus. That focus aligns with the expertise and research interests of the faculty member supervising that circle. Each circle presents a set of core readings that all students in that circle must read critically to kick off their research project. Each circle shows which courses in the E&BI program it links to. Students in a circle need to ensure that they stay aligned with that focus. That will make it easier for a supervisor to provide feedback and guidance, and for the student to give and receive constructive feedback (including peer feedback). Consequently, the number of thematic options differs each time the Capstone course runs. The availability of supervisors, their thematic preferences, and their choice of thesis type determine which circles will be available. This limits student choice, yet it will ensure competent supervision can be provided.

Each semester, a new set of supervisors is asked to propose a circle, using a template that ensures that all circle outlines are complete and consistent. Usually, most supervisors have supervised an E&BI circle before, but some have no prior experience with supervising E&BI students. Most have supervised bachelor's and/or master's theses as part of other programs at TiSEM before and thus need to adjust to the distinctive setup of this program's capstone course. This requires clear and timely communication from the course coordinator to the new supervisors.

During the course, the students will first develop a research proposal that will be assessed by the circle supervisor on a pass-or-fail basis. During the semester, four circle meetings are scheduled in fixed weeks. Attending and preparing for these meetings is expected (and monitored). The supervisor and a second examiner will eventually assess the thesis. The second examiner is supervising another circle in the same semester. Matches between first and second examiners are made based on the research type, method and domain as much as possible.

There are strict requirements regarding the form and length of a thesis (for example, regarding adherence to APA 7 Style for a student paper, and minimum and maximum word counts).

Capstone Projects in the Time of Generative Artificial Intelligence (GenAI)

A key requirement for assignment deliverables to reflect student learning and mastery is that the work is produced by the student. The hiring of ghostwriters has long been a challenge to detect, and, in a way, GenAI tools have democratized the usage of ghostwriting. AI can be used as a co-pilot to enhance the quality of student work, but students must take responsibility and not outsource entire sections (or the thesis altogether). Both students and supervisors are informed of this AI policy, including during the introductory lecture at the start of the semester.

Circle meetings serve as opportunities to test students' verbal representation of their work. Students must be able to explain in their own words what they have done, how, and why. Supervisors may schedule additional meetings if they—or the second examiner—deem it necessary to verify the authenticity of a student's work. Students receive more detailed instructions on one of the course's Canvas pages (see Box 3.16.1).

Box 3.16.1

The AI Policy of E&BI's Capstone Course

N.B.: The text below applies to both the proposal and the thesis.

This assignment (**AI – Index – Level 3**) allows you to use tools such as ChatGPT to refine ideas, structure your assignment, and paraphrase texts. AI can serve as support in improving your assignment, but you remain responsible for its content and the conclusions. Because AI models such as ChatGPT do not have a transparent search function and it is often unclear where information comes from, searching for literature via academic databases such as Web of Science is strongly recommended. For other types of information, we also recommend using authentic and verifiable sources. For data analyses, AI may assist with calculations, visualization, or code suggestions, but it cannot replace your own interpretation, verification, or final analysis of results. Within the assignment, you may use AI tools for:

Checking spelling, grammar, coherence, syntax and referencing: Use AI to identify and correct language and formatting errors, and to enhance readability. Instruct the AI to highlight each error (in bold) with the suggested correction in [brackets], and to avoid automatically rewriting the entire text.

Editing:

- Check narration or subtitles in videos/audio for grammar, clarity, and flow.
- Review instructions, labels, or UI text in prototypes for clarity in the prototypes.
- Check code comments, variable names, or textual descriptions in scripts, programming, or visualizations for readability and overall coherence.

Inspiration and brainstorming: AI may be used to explore ideas. Verify all AI-generated information with reliable sources and reference these in your assignment. **Incorrect, fabricated sources or missing references will not be tolerated.**

It is not allowed to have AI generate complete texts, tables, or graphs instead of making them yourself. AI tools may only be used for supportive purposes.

Transparency: Clearly document for each tool what you used it for in the technology statement. You must use the Technology statement student(s) (see text below). **Incomplete transparency or presenting AI output as your own work is not permitted.**

Data confidentiality: Never enter sensitive or confidential data into AI platforms, as these may not comply with Tilburg University's or external organizations' privacy guidelines.

Improper or non-transparent use of AI within the assignment will lead to sanctions and may result in your assignment being declared invalid by the Examination Board.

Technology statement student(s)

For the assignment, you must add the technology statement, using the text below. Replace the text in capital letters with the requested information.

During the preparation of this work, I/We used **[NAME TOOL/SERVICE/VERSION OF AI TOOL]** in order to **[REASON]**. The following parts of the assignment were affected/generated by AI tool usage: **[INTRODUCTION/METHODS/DISCUSSION]**. After using this tool/service, **[NAME STUDENT(S)]** evaluated the **validity** of the tool's outputs, including the sources that GenAI tools have used, and edited the content as needed. As a consequence, **[NAME STUDENT(S)]** take(s) full responsibility for the content of their work.

Assessment Components

Proposal

Students are required to submit a research proposal during the first part of the graduation project. The proposal indicates the extent to which students have achieved part of the CLGs (see Table 3.16.2) and is assessed on a pass-or-fail basis by the supervisor.

A "pass" allows students to continue working on their project according to the proposed plan. A "fail" provides an opportunity to submit a revised proposal, which will again be assessed on a pass-or-fail basis. If the revised proposal gets a "pass", students may continue their project, albeit with a delay. If it gets a "fail" again, students are not permitted to continue and will fail the course.

Incomplete, late, or missing submissions automatically result in a "fail" and will not receive written feedback.

The criteria for proposal assessment are specified in a rubric (see Table 3.16.2). Students are expected to familiarize themselves with this rubric well in advance of submission.

Final Product

The final product (i.e., the thesis) is the only graded assignment in the course and demonstrates how well students have achieved the CLGs. It counts for 100% of students' course grade.

A rounded grade of 6.0 or higher (i.e., 5.5 or higher) is required to pass the course. Students scoring below 5.5 at the first attempt may submit a revised version. The grade for this resubmission cannot exceed 6.0, even if the final product was not submitted or submitted late at the first attempt. Students who fail to achieve at least 5.5 on the resubmission will fail the course.

Incomplete or late submissions are not graded and do not receive written feedback.

Two examiners independently graded the final product, i.e., of the circle (first examiner) and another qualified faculty member (second examiner). The course coordinator determines the second examiner based on expertise, interests, and availability. The assessment criteria are specified in a rubric (see Table 3.16.3), which students are expected to review well before submission.

Guidelines for Submissions

- Proposals must be between 2,000 and 3,000 words (excluding cover page, reference list, and appendices). Submissions outside this range automatically receive a “fail.” The word count must be stated on the cover page.
- Theses must be between 6,000 and 9,000 words (excluding cover page, abstract, acknowledgements, reference list, and appendices). Submissions outside this range will not be graded. The word count must be stated on the cover page.

Both proposals and theses must follow APA Style 7 for a Student Paper, ensuring consistent formatting, including cover page, line spacing, in-text citations, and references. Students should refer to Tilburg University’s APA Manual for detailed guidance.

Revised proposals or theses must be submitted as PDFs with visible track changes using MS Word’s track change’s function. Submissions that do not comply with this rule will not be graded, resulting in failure of the course.

Grading Rubrics

Rationales and Background of the Rubric Design

From the start, all capstone projects were required to be assessed using a single rubric. Since all projects are built on empirical research, the appropriateness of data collection and analysis can be considered while grading. However, the role and emphasis of empirical research may differ across projects, and the assessment must account for this. The capstone course is assigned to help students demonstrate achievement of the program’s intended learning outcomes, translated to course learning goals. Across three types of projects, theory and empirical data analyses are used differently and for different goals which must be reflected in the assessment criteria and standards. Differences in emphasis on rigor, comprehensiveness of literature use or empirical data analysis should not be seen as problematic but as inherent to the project type.

A rubric provides a framework for qualitatively assessing complex or authentic student work (Jonsson & Svingby, 2007). The initial rubric used for our capstone course was co-created by the academic director of the program, the Vice-Dean of Education and the Associate Deans of Education, with the assistance of an assessment specialist. It was standards-based (Jonsson & Svingby, 2007) to ensure valid, transparent, and reliable scoring. Each program-level learning outcome is translated into at least one course learning goal, resulting in multiple criteria to assess capstones that cover all the program’s learning outcomes.

Complex, performance-based assessments (Linn et al., 1991) require careful consideration to ensure that grades accurately reflect what students have learned. Rubrics help examiners provide valid, reliable, and transparent judgments aligned with course learning goals and program learning outcomes. While assessment remains inherently subjective (Sadler, 2005), using criterion- and standards-based rubrics reduces subjectivity and increases transparency compared with norm-referenced approaches.

In the most recent site visit, the panel recommended reducing the number of criteria by grouping them into fewer, more abstract standards and suggested moving from a standards-based to a norm-based rubric. While peer-reviewed literature offers little support for norm-referenced rubrics for theses, Bloxham et al. (2016) note that detailed rubrics with many criteria do not guarantee marker agreement and may benefit from fewer criteria for complex assignments. Detailed standards-based rubrics require examiners to make multiple complex, intuitive judgments, making variability inevitable (Verhavert et al., 2019). Therefore, we have simplified the earlier rubrics for both proposals and theses in Table 3.16.2 and 3.16.3.

Table 3.16.2*Rubric Assessment Criteria – Proposal*

Criterion (relevant CLGs)	Description
Problem statement and research questions (CLG6)	<ul style="list-style-type: none"> • The proposal presents a clear problem statement, objectives, and (an) appropriate research question(s) • The problem statement and research question(s) are well motivated and well formulated • The research question(s) is/are relevant from a theoretical, practical and/or didactical point of view • The proposal clearly discusses the identified gap(s) in the existing literature • The originality and relevance of the proposed research are clearly addressed
Theory and conceptual framework (CLG1)	<ul style="list-style-type: none"> • The proposal clearly describes what we already know about the topic, what not, what we still need to know, and why • Relevant literature is briefly introduced, and this preliminary literature review shows some degree of synthesis • The proposal introduces relevant theories and concepts in a clear manner • Key concepts and the relationships between them are clearly explained, and placed into a correct and appealing conceptual framework • Possible hypotheses are well formulated and based on relevant theories and sound theoretical argumentation
Data and methods (CLG2)	<ul style="list-style-type: none"> • The proposal describes the intended research approach (e.g., data sources, sample, sample size, participants, equipment, materials) with a sufficient level of detail • The proposal clearly describes how data will be collected and preprocessed for further analysis, and how the resulting dataset will be analyzed • It is clear what research method(s) will be used to obtain (solutions and recommendations from) the empirical findings • The proposed research method(s) reflect(s) the state of the literature, and enable(s) to answer the research question(s) • The proposal includes a clear description and convincing justification of the research method(s) that will be used • It is clear how all key concepts will be measured, this is feasible (within the remaining time), and these measures are valid and reliable
Formatting and referencing (CLG5)	<ul style="list-style-type: none"> • The formatting of the proposal complies with the style guide for student papers as specified by APA Style 7 • The text is presented in a structured manner, and follows a logical ordering principle (i.e., chapters, sections, and paragraphs) • Tables and figures are formatted in a consistent manner, and in line with academic standards • In-text citations and references are correct and complete, and in compliance with APA Style 7 standards • The proposal contains only plain text (no bullet points), and is free from (English) spelling and grammar mistakes

Note. Grading instructions for students: Each criterion is assessed as either a “pass” or a “fail”. Only if you receive a “pass” on each of the four criteria will your entire proposal receive a “pass”. In any other case, so even if just one of the four criteria receive a “fail”, your entire proposal will receive a “fail.”

Table 3.16.3*Rubric Assessment Criteria – Final product*

Criterion relevant CLGs)	Description
Problem statement and research questions (CLG6) (15%)	<ul style="list-style-type: none"> • The thesis presents a clear problem statement, objectives, and (an) appropriate research question(s) • The problem statement and research question(s) are well motivated and well formulated • The research question(s) is/are relevant from a theoretical, practical and/or didactical point of view • The thesis clearly discusses the identified gap(s) in the existing literature • The originality and relevance of the research are clearly addressed
Theory and conceptual framework (CLG1) (20%)	<ul style="list-style-type: none"> • The thesis clearly describes what we already know about the topic, what not, what we still need to know, and why • Relevant literature is critically and thoroughly discussed, and the literature review shows a high degree of synthesis • The thesis correctly discusses relevant theories, and describes relevant concepts in a clear manner • Key concepts and the relationships between them are clearly explained, and placed into a correct and appealing conceptual framework • Hypotheses (if any) are well formulated and based on relevant theories and sound theoretical argumentation
Data and methods (CLG2) (15%)	<ul style="list-style-type: none"> • The thesis describes the research approach (e.g., data sources, sample, sample size, participants, equipment, materials) with a level of detail that enables others to replicate the research • The thesis clearly describes how data has been collected and preprocessed for further analysis, and how the resulting dataset has been analyzed • It is clear what research method(s) has/have been used to obtain (solutions and recommendations from) the empirical findings • The research method(s) reflect(s) the state of the literature, and enable(s) to answer the research question(s) • The thesis includes a clear description and convincing justification of the research method(s) that has/have been used • It is clear how all key concepts have been measured, and these measures are valid and reliable
Results (CLG3) (20%)	<ul style="list-style-type: none"> • The data collection and preprocessing methods have been performed correctly and effectively • The method(s) to analyze the data has/have been applied correctly and effectively • The research findings are presented in a clear and structured manner • The robustness of the findings has been checked sufficiently • The results are based on a correct and complete analysis, showing a deep understanding of the research question(s) and relevant analytical techniques to answer it/them • The conceptual framework and empirical findings have been used to methodically develop a solution to the project's problem
Conclusions and discussion (CLG4, CLG6) (20%)	<ul style="list-style-type: none"> • The conclusions presented in the thesis can be drawn from the results and are related to the research question(s) • The thesis includes an in-depth discussion of the theoretical, practical and/or didactical implications of the research findings • The thesis discusses limitations of the research, and offers interesting and relevant recommendations for future research • The implications and recommendations logically follow from the results and the corresponding conclusions • The thesis clearly discusses who may benefit from the findings (e.g., academics, firms, policymakers, society) and how
Formatting and referencing (CLG5) (10%)	<ul style="list-style-type: none"> • The formatting of the thesis complies with the style guide for student papers as specified by APA Style 7 • The text is presented in a structured manner, and follows a logical ordering principle (i.e., chapters, sections, and paragraphs) • Tables and figures are formatted in a consistent manner, and in line with academic standards • In-text citations and references are correct and complete, and in compliance with APA Style 7 standards • The thesis contains only plain text (no bullet points), and is free from (English) spelling and grammar mistakes

Note. Grading instructions for students: Each criterion can be scored from 1 to 10 (0 if absent). The grade is then determined by adding the weighted scores. Only the first examiner has the discretionary power to adjust that grade by up to 0.5 points (down or up) based on desired behavior during the semester. This includes, but is not limited to, good preparation for and active participation in the circle meetings, provision of valuable peer feedback, effective use of (peer) feedback, professional treatment of participants, responsible use of data and other scientific sources, staying on schedule, and meeting deadlines.

Grading Procedures

Independent Assessment

Two examiners (the thesis circle chair and a second faculty member) independently grade the final product using the second rubric above.

Comparison and Resolution of Grades

The examiners meet to compare their assessments and discuss any differences.

If both examiners award the same grade, that grade is final.

If the grades differ:

- When the difference is 1.0 point or less, the final grade is the average of the two grades.
- When the difference is greater than 1.0 point, or when one grade is below 5.5 and the other is 5.5 or higher, a third examiner is appointed to grade the work independently.

Role of the Third Examiner

The third examiner assesses the work without knowledge of the previous grades.

- If the third grade is equal to or lower than the lowest of the original two grades, the final grade is the lowest original grade.
- If the third grade is equal to or higher than the highest of the original two grades, the final grade is the highest original grade.
- If the third grade falls between the two original grades, the final grade is the average of all three grades.

Calibrations to Improve Inter-Rater Reliability

In practice, examiners may interpret assessment criteria or standards differently, leading to variation in scoring. With multiple criteria, this can increase measurement error, even with careful calibration amongst examiners. Thesis assessment requires a third examiner when the grade difference between two examiners exceeds 1.0 point on a 10-point scale. The accreditation panels also review whether such disparities occur and whether the rubric is applied appropriately. During the last re-accreditation, the panel observed that despite using a rubric with clear standards, the difference between the grades given by the two examiners was often too large. To address this, the program organizes annual calibration sessions for examiners with varying levels of experience with thesis supervision in general, and E&BI capstones in particular. During these sessions, examiners grade the same two student products and discuss their findings. Differences in grading between experienced and less experienced examiners are identified and resolved during the session, helping to improve consistency before the grading period begins.

It was also decided to replace the standards-based rubric with norm-based rubric with fewer criteria, hoping that this would result in a more holistic assessment of thesis quality and hence in a more aligned grading.

This aligns with Bloxham et al., (2016) who argue that a detailed rubric with many criteria does not guarantee marker agreement, and suggest opting for fewer criteria when grading more complex student deliverables.

Reflection

Relationship Between Grading of Proposals and Theses

Observations from calibration sessions have provided several insights into the grading relationship between research proposals and final products within the E&BI program.

The E&BI proposal is more substantial than those in many other programs at the TiSEM. It is summatively assessed on a pass-or-fail basis, with clear consequences; failure after a resit or non-compliance with submission rules results in course failure. The proposal already covers approximately one third of the thesis content, thereby addressing four

key goals, namely: literature review and gap identification, research questions, methods, and data. However, these sections are generally less elaborate than in the full final products, reflecting a lower assessment bar appropriate for this stage.

Personal observations suggest that the formalized proposal structure contributes to timely completion of final products. The requirements encourage students to think through essential components, particularly data collection, the correct measurement of variables, and alignment between methods and the introduction/conceptual framework. This structured approach also aids examiners in evaluating the feasibility of proposed research.

The standards applied to proposals and final products differ significantly. A proposal deemed sufficient does not automatically imply that the corresponding sections in the final products will meet the higher standards required for a final assessment. Proposals are graded on a pass-or-fail basis, whereas final products are graded on a numerical scale with weighted criteria. For example, when rating the thesis, theory and conceptual framework (CLG1) and other components carry different weights than in the proposal, reflecting greater expectations for scope, depth, and analytical rigor.

Examiners are instructed to grade each submission independently. Progress or improvements between the proposal and the final products should not influence grading; all assessments must strictly follow the rubric. The proposal requires four criteria to be deemed sufficient, whereas the final product should meet additional criteria, with some criteria carrying higher weights. Only the first examiner has the authority to adjust the final grade based on observed student behavior and process engagement (with max. 0.5 points, up or down)

What to Keep and What to Change in the Future?

The program aims to retain the option for students to choose between different types of research to accommodate diverse preferences and learning paths. At a minimum, academic and applied research will continue to be available. While there is no objection to maintaining case development research as a third option, in practice there appears to be limited interest from both faculty and students. Consequently, this option may gradually disappear unless actively promoted.

Looking ahead, and in line with the initial accreditation documentation of the E&BI program (in 2019), we plan to introduce the option of “creating and validating a new commercial or social venture” from the 2026/2027 academic year onwards. There is also consideration for allowing students to undertake this type of research in small teams, while ensuring that each individual’s contributions are clearly documented and that all CLGs are achieved on an individual basis.

These startup research projects must extend beyond the scope of prior coursework. For instance, as applied research must go beyond simply applying Porter’s Five Forces to an external organization, startup research must similarly go beyond conducting basic industry and competitor analyses. The academic directors of E&BI are currently preparing a document outlining this possible new type of research option. This document will be used to integrate recent considerations and feedback, and eventually to seek approval from E&BI’s Education Committee.

3.17

Master Thesis Organization and Management Studies

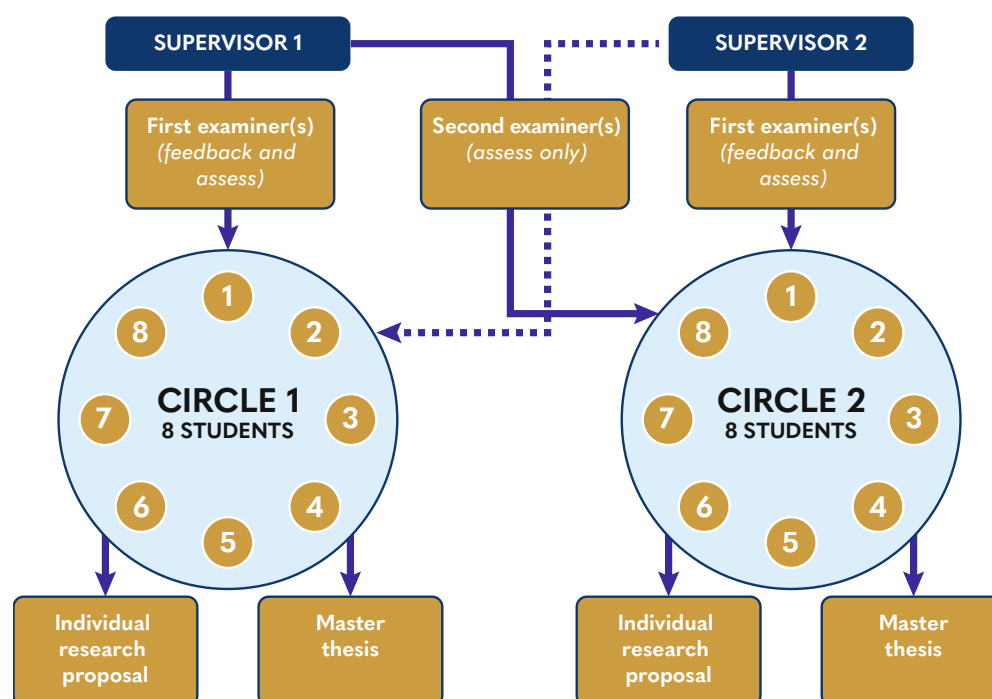
Roel Rutten and Lonneke Schellekens

Course Overview

The Master Thesis Organization and Management Studies (MTOMS – 24 EC) is the final part of the Organization and Management Studies (OMS) master program in which students write an individual master's thesis. This course follows up on the course 'Master Seminar Organization and Management Studies' (MSOMS – 3EC) in which students write an individual research proposal (IRP) to prepare for their master's thesis. Students are assigned to supervisors in the process based on the program specified procedures. Supervisors are also thesis examiners, next to the second reader. After the supervisors have approved their IRP, students continue to master's thesis.

For the master's thesis, students individually do a literature study, collect and analyze data, and report the findings in their master's thesis. Throughout MSOMS and MTOMS, students are supervised by OS staff members in so-called master thesis circles. Circles are groups of no more than eight students who generally write their master's thesis on a similar topic (on an organizational topic on which their supervisor is knowledgeable) and they benefit from giving feedback on each other's work.

Each teaching staff member participates in two circles (see Figure 3.17.1): their own circle (first circle), where they serve as supervisor/first examiner, and a second circle, where they act as second examiner. Each student first writes an IRP, and if they pass, they proceed to write a thesis. In their own circle, supervisors/first examiners assess two products per student: the IRP (in January) and the thesis (in June). Across both circles, a teaching staff assesses up to $2 \times 8 = 16$ IPRs and $2 \times 8 = 16$ Master Theses per academic year. To ensure independence and neutrality, each teaching staff only assesses the IRPs and Master Theses in their second circle, with whom they have no supervisory involvement.

**Figure 3.17.1**

How Supervisors (1st Examiners) and Second Examiners Are Assigned to Circles

Thesis Learning Goals

During MTOMS students develop the following thesis learning goals:

- Designing a theoretically and methodologically coherent, plausible and feasible academic research (program learning outcomes/PLOs 1.4; 2.2; 2.3; 2.6).
- Conduct a theoretic research (PLOs 2.4; 2.5; 3.3; 3.6; 5.1; 5.2).
- Collecting and analyzing relevant data (PLOs 3.2; 3.4; 3.5).
- Reporting research findings in a systematic and accessible way (PLOs 4.1; 4.2; 4.3; 4.4; 5.3; 5.4).

The thesis is a comprehensive exercise in which all program **learning outcomes** are integrated. The thesis learning goals are directly aligned with the PLOs, ensuring that students consolidate and apply the full range of competencies developed throughout the program (see Table 3.17.1). First, students design theoretically and methodologically coherent research projects, thereby demonstrating their understanding of scientific knowledge development, their ability to independently contribute to theory building, and their capacity to address organizational complexity. Second, they conduct theoretical research, applying advanced methods and ethical standards, critically evaluating the work of others, and situating their study in ongoing academic and professional debates. Third, they collect and analyze data, which requires them to adapt their perspectives to new insights, articulate uncertainty, and assess contributions from related disciplines. Finally, students report their findings in a systematic and accessible way, demonstrating mastery of academic English, the ability to address diverse audiences, and constructive engagement in scholarly exchange. This process also cultivates independent learning, the continuous updating of knowledge, and a commitment to lifelong learning.

Table 3.17.1*Accountability Table*

Program Learning Objectives served by the courses	Individual Research Proposal	Master Thesis
Knowledge and Understanding	Independently writing a research proposal. The proposal is assessed on the criteria as specified in the assessment form.	Independently writing a master's thesis. The thesis is assessed on the criteria as specified in the assessment form.
1.4: Insight into how scientific knowledge is developed.		
Applying Knowledge and Understanding		
2.2: Able to, independently, make an original contribution to the development of knowledge in organization studies including theory development and the development of knowledge in relevant subfields.		
2.3: Able to design and conduct independent research to address questions relating to complexity and dynamics in and between organizations.		
2.4: Able to apply current advanced methods and statistics in independent research to diagnose organizational problems.		
2.5: Able to apply insights derived from research to design organizational interventions.		
2.6: Able to apply the knowledge of intra-organizational dynamics and organizational change to lead a team in the process of organizational change implementation.		
Making judgements		
3.2: Able and willing to change his/her opinion and perspective as new knowledge is acquired or the new empirical facts emerge.		
3.3: Able to integrate ethical norms and values into his/her own professional (also research) behavior.		
3.4: Able to deal with and articulate uncertainty, including in complex situations with limited or incomplete data.		
3.5: Able to recognize other scientific fields and parties and to assess their (scientific) quality and limitations.		
3.6: Able to judge the work of peers.		
Communication		
4.1: Are able to communicate clearly orally and in writing about research findings, new developments, and solutions in organization studies and their implications for organizational practice and society.		
4.2: Able to craft his/her communication content and style to effectively communicate with various audiences and stakeholders, specialists, non-specialists.		
4.3: Able to communicate orally and in writing in academic English.		
4.4: Able to provide and receive feedback in a critical but constructive manner.		
Learning Skills		
5.1: Able to collect the necessary information on organizational problems.		
5.2: Able to follow the developments in organization studies relevant for the execution of professional activities at a national and international level.		
5.3: Able and willing to assess, revise, update, and extend his/her knowledge base.		
5.4: Aware of the continuous process of knowledge development and willing to engage in lifelong learning.		
Grade	One grade, 100%	One grade, 100%

Teaching, Supervision & Learning Activities

Introductory Lecture

The introductory lecture introduces the MSOMS and MTOMS courses and explains what is expected of students and what students may expect of their supervisors. After the lecture, students indicate their first, second and third preferences to the course coordinator. As much as possible, the course coordinator will assign students to a circle of their choice.

MSOMS-Writing the IRP

During MSOMS, students write the individual research proposal (IRP). The IRP demonstrates the feasibility of the research project the student intends to conduct. It provides evidence that the student possesses the required academic level and methodological competence to carry out the study successfully. Furthermore, the IRP offers supervisors the opportunity to provide early feedback and guidance, ensuring that potential flaws or weaknesses in the research design are addressed before the student proceeds to the full thesis stage.

During MSOMS, students will meet in four circle meetings with each meeting lasting no more than two hours. The first meeting will be organized by the 1st supervisor, organizing the next meetings is the responsibility of the students themselves. During these four meetings, supervisors will guide students through writing their IRPs. Between circle meetings, students may contact the 1st supervisors with short questions only, which they will answer. Lengthier questions and the actual supervision take place in the circle meetings to allow all students to benefit from the 1st supervisor's comments. Students also read each other's work and give feedback to each other during the circle meetings.

During the IRP assessment meeting the 1st and 2nd supervisors discuss the strengths and weaknesses of each individual IRP. Students may answer questions for clarification when their IRP is assessed. At the end of the meeting all students will receive an **assessment** of their IRP. The assessment can be: 1 (no revision), 2 (minor revision), 3 (major revision) or 4 (rewrite). An assessment of 1 or 2 is a Pass, which means that students continue to MTOMS to write their master's thesis. An assessment of 3 or 4 is a Fail, which means that students have to revise and resubmit their IRPs. In all cases, the 1st and 2nd supervisors give clear feedback on the strength and weaknesses of each IRP and clearly indicate which part or parts of an IRP must be revised, when necessary. The program follows the school provisions of thesis assessment to deal with disagreements between two examiners. If examiners cannot agree, a third examiner is involved. The third examiner decides the final grade based on (i) assessment form of 1st and 2nd examiners and (ii) their own reading/assessment of the thesis.

MTOMS-Writing the Thesis

During four two-hour circle meetings students are supervised while writing drafts of their master's thesis. Students are responsible for organizing the circle meetings. The 1st supervisor gives feedback on draft chapters that students submit before the circle meetings. Also, during the writing of the master's thesis students give feedback on each other's work. Supervision, written peer and lecturer feedback, and narrative lecturer and peer feedback are mostly organized during the circle meetings. Collective supervision and feedback ensure (i) a more efficient use of supervisor time because important things need to be explained only once, (ii) that students benefit from the feedback to others, (iii) that students can feedback each other's work, and (iv) that supervisors work on topics on which they are experts.

Assessment Components

The master's thesis comprises a single grade (100%). Students have two opportunities to submit their final thesis: the first and second rounds. To use any one of the opportunities, students must submit a complete and final thesis. Incomplete theses, e.g., with missing appendices, cannot be submitted. Students who do not have a final thesis for the first round must use the second round. The final grade of the thesis will be determined after the defense. During

the defense meeting students face a round of questions from their 1st and 2nd supervisors. The purpose of the defense is: 1) to clarify issues and address concerns related to the thesis; and 2) checking students' authorship/ownership of their own work.. Successfully clarifying issues and addressing concerns may result in a higher grade while failing to do so may result in a lower grade.

A grade of 5.5 or higher is a *pass*, a grade lower than 5.5 is a *fail*. Students who receive a pass have finished the master's thesis course. It is not possible to revise and resubmit a thesis that has received a pass to attempt to get a higher grade. To get a higher grade after a pass, students must submit a new thesis on a different topic for the second opportunity – keeping in mind that the last grade counts. Students who received a fail in the first round must revise and resubmit their thesis in the second round, based on the feedback they received from their supervisors. Students who have not submitted a final thesis by the second round automatically fail the master's thesis. The second round is identical to the first round. Students who receive a fail in the second round must write a new thesis in the next academic year. They may do so on the basis of the same IRP because this IRP has received a pass.

Use of Large Language Models and Generative Artificial Intelligence (GenAI)

An IRP and a master's thesis are *original works* from a student. Letting someone or something else write (part of) the thesis is regarded as fraud and punishable in the same way as plagiarism. The use of Large Language Models (LLMs) and GenAI is limited to idea generation and structuring, corresponding to Level 3 of the Tilburg University AI-Index, which means that students can use AI tools to develop or refine initial ideas, materials (e.g., interview guides, stimuli for experiments), paraphrasing, structures, or outlines. This level is chosen because the thesis must remain the student's original work. Students may therefore use LLMs and GenAI as supportive resources to inspire their own idea generation, thinking, and writing. However, AI tools cannot substitute for independent analysis, critical reflection, or academic writing. Any AI-generated output may serve only as input for the student's original work and must not be directly included in the thesis. Ultimately, students are fully responsible for ensuring that their thesis reflects their own ideas, methodological choices, and academic integrity.

To allow supervisors to make a valid assessment of an IRP and master's thesis, students must be transparent about their use of LLMs and GenAI. Only under the condition of **transparency** can students demonstrate their mastering the required critical thinking skills. Accordingly, all use of LLMs and GenAI must be:

- declared to the first supervisor,
- discussed with the first supervisor,
- approved by the first supervisor.

This transparency allows supervisors to guide students in the process of learning and writing. It allows students to claim genuine authorship of their work, also when they use LLMs and GenAI as a resource.

Assessment Criteria

The assessment criteria of the thesis are:

1. Introduction and research question
 - Students clearly introduce the research problem and explain why this is a (theoretical, methodological, practical) problem (depending on the nature of the research).
 - Students clearly specify what will be explained (outcome) and on the basis of what factors (conditions, causes) the explanation will be developed (in agreement with the good-practice rules of the method they chose).
 - Students develop the research problem into a clear research question and introduce the research context (empirical setting, object of analysis).
 - Students clearly explain why the research method chosen is appropriate for the proposed research problem, research question and research context.

2. Theory and literature
 - Students clearly identify and discuss a body of literature that is relevant for their research.
 - Students discuss the state of the art in this literature and explain how this potentially answers their research question and/or informs their empirical research.
 - Students demonstrate the ability to go beyond summarizing the literature, for example by identifying gaps in the literature, connecting different literatures, drawing novel conclusions, or developing novel propositions (hypotheses).
3. Design and methods
 - Students clearly explain the design of their study and how they will get from raw data, to analyzable data, to answering their research question.
 - Students clearly explain why their design is (most) appropriate given their research question and how theory/ literature inform the research design.
 - Students clearly explain how they will collect their data and why these data will answer the research question.
 - Students clearly explain how they will select respondents.
 - Students clearly explain which method(s) of data analysis they use and, in connection to that, deal with issues such as operationalization and measurement in accordance to the rules of good practice of their method.
 - Students discuss the strengths and weakness of their design and explain how they deal with ethical issues concerning data collection.
4. Analysis and results
 - Students present, discuss and analyze their data in accordance to the rules of good practice of their method.
 - Students present and report their empirical findings in accordance to the rules of good practice of their method.
 - Students explain on the basis of what criteria they draw inferences from their data. For example, for statistical analysis they may discuss significance levels; for narratives they may explain the plausibility of their inferences by weighing confirming and disconfirming evidence.
 - Students explain clearly what the empirical findings of their study.
5. Discussion and conclusion
 - Students specifically answer the research question and discuss the strengths and limitations of their answer in a way that clearly goes beyond a repetition of the results.
 - Students reflect on the implications of the findings and limitations of their (theoretical, methodological and empirical) research. They also discuss how the implications of their findings for (broader discussions in) the literature, methodology and practice.
6. Coherence and analytical depth
 - In addition to the above, students prepare a data package. The data package must contain the raw data and must explain how the data were processed and analyzed. Students submit the data package together with their thesis.
7. Form and style
 - Students write their thesis in accordance with the form and style requirements specified in the course manual and with the good practice standards of the scientific community in general.
 - In any case, theses must not have systematic grammatical or other language errors; the lay out and presentation of the work should be clear and easy to follow.
 - Artwork must be reproduced in a clear way.
 - Appropriate references must be made throughout the work and a complete list of references in a consistent format must be provided at the end of the work.
 - Non-essential materials must be presented in logically ordered and structured annexes.
8. Attitude and independence
 - Students demonstrate an ambition to produce high-quality work that meets ethical standards, and take initiative to find literature, collect data, develop analytical skills, etc.

- In case thesis writing takes place in circles, students engage with circle members in a productive and constructive manner, in particular by giving constructive feedback on each other's work. Independence means that students engage with supervisors to reflect on their own ideas rather than depend on supervisors to develop ideas.
- Students demonstrate that they are able to develop original and creative ideas concerning their research. Note that only the first supervisor assesses this criterion.

Master's Thesis Defense

The first and second supervisors identify the key issues they want to discuss with the student. They formulate these issues into three to six questions that the student can answer within the maximum twenty minutes of the thesis defense. The supervisor assesses how the student performed in the defense on three criteria: exhaustive, convincing and concise. Based on the student's performance, supervisors can: (i) lower their initial grade of the thesis with 1.0 or 0.5 point, (ii) leave their initial grade unchanged, or (iii) raise their initial grade with 0.5 or 1.0 point.

- **Exhaustive:** It was evident that the student understood all questions. The student answered all questions and addressed all the issues raised. In their answers, the students prioritized the most important issues.
- **Convincing:** The student answered evidenced that they comprehend the concepts/ theories/ methods/ etc. discussed in their thesis. The student successfully clarified the issues that the supervisors raised.
- **Concise:** The student answered the questions in a confident and straightforward manner and did not stray into irrelevant issues. Answers were to the point and not evasive.

Rubric Thesis Assessment

A thesis fails in case it scores a 'Poor' on at least one criterion. A thesis also fails when it scores an 'Insufficient' on at least two criteria. Higher scores on other criteria cannot compensate 'poor' and 'insufficient' scores (see Table 3.17.2). The final grade of the thesis is the weighted average of the grades for the individual criteria as displayed in the grading table in Table 3.17.2 and adjusted for the performance during the thesis defense.

Table 3.17.2
Rubric *Thesis Assessment*

CRITERION	Label	Excellent	Very good	Good	More than sufficient	Sufficient	Insufficient	Poor	Weight
	Grade								
1. Introduction and research question Students clearly introduce the research question and explain which underlying research problem (theoretically, methodologically, practically) it addresses and why it is important.									15
2. Theory and literature Students clearly identify and systematically discuss a body of literature that is relevant for their research, and explicate the implications of the literature for their study.									15
3. Design and methods Students clearly explain how the study is designed and how they will get from raw data to analyzable data to answering their research question.									15
4. Analysis and results Students present, discuss and analyze their data in accordance to the rules of good practice of their method.									15
5. Discussion and conclusion Students specifically answer the research question and discuss the strengths and limitations of their answer. The implications for theory, methodology, and practice are discussed clearly. Students makes clear suggestions for future research.									15
6. Coherence and analytical depth Coherence means that all parts of the thesis are clearly connected. Analytical depth means that the discussion and argumentation throughout the paper must be rigorous, clear and concise.									15
7. Form and style Students write their thesis in accordance with the form and style requirements specified in the course manual and with the good practice standards of the scientific community in general.									5
8. Attitude and independence Students demonstrate an ambition to produce high-quality, replicable work that meets ethical standards, and take initiative. Independence means that students engage with supervisors to reflect on their own ideas rather than depend on supervisors to develop ideas.									5

Rubric Defense Assessment

The thesis defense is a new component, specific to the OMS program. It allows for specific instructions and criteria tailored to the assessment of an oral examination (see Table 3.17.3). Supervisors establish an initial grade for the thesis. Following a student's performance during the defense, this grade may change as follows.

Table 3.17.3

Rubric Defense Assessment

Criterion	plus 1.0	plus 0.5	0	minus 0.5	minus 1.0
Exhaustive: It was evident that the student understood all questions. The student answered all questions and addressed all the issues raised. In their answers, the students prioritized the most important issues.	The student clearly understood all questions and addressed all the issues raised in balanced way. Spending more time on the most important ones.	The student clearly understood all questions and addressed almost all the issues raised.	The student understood the questions but the answers did not address or clarify the issues beyond what the thesis says.	The student had some difficulty understanding the questions and addressed the majority of the issues raised.	The student struggled to understand the questions and addressed only a few of the issues raised.
Convincing: The student's answers evidenced that they comprehend the concepts/ theories/ methods/ etc. discussed in their thesis. The student successfully clarified the issues that the supervisors raised.	The student's answers evidenced a strong command of their thesis and substantially clarified (almost) all of the issues raised.	The student's answers evidenced a good command of their thesis and substantially clarified most of the issues raised.	The student's answers confirmed supervisors' initial assessment of the student's command of their thesis and did not substantially clarify the issues raised.	The student's answers raised some doubts about their command of their thesis and clarified only a few of the issues raised.	The student's answers raise further questions about their command of their thesis and failed to clarify the issues raised.
Concise: The student answered the questions in a confident and straightforward manner and did not stray into irrelevant issues. Answers were to the point and not evasive.	The student's answers were very focused and very straightforward and no irrelevant issues were touched.	The student's answers were focused and straightforward and no irrelevant issues were touched.	The student's answers were neither particularly straightforward nor particularly evasive. No irrelevant issues were touched.	Some of the student's answers were unfocused and evasive and touched on some irrelevant issues.	The student's answers were unfocused and evasive and touched multiple irrelevant issues.

Reflections on Challenges and Adaptations

Maintaining Consistency Across Supervisors

Up to 12 supervisors are involved in thesis supervision, one for each circle. Supervisors differ substantially in their backgrounds (some are more qualitatively oriented researchers, others more quantitatively) and experience (some are newly appointed assistant professors who just finished their PhD, others are senior associate professors). However, the same standards of supervision must be maintained across supervisors. Allowing for variations in their personal style, supervisors must have the same basic approach to giving feedback and must follow the same procedures. To ensure a reliable and valid process of supervision and assessment across supervisors, a fairly detailed “handout for supervisors” was written. Also, before the start of the thesis supervision, the thesis coordinator calls a meeting with all supervisors to talk them through the procedures. Note also that each supervisor serves as a second reader in another circle. As much as possible experienced and unexperienced supervisors are paired.

Changing the Thesis Assessment

It used to be the case that students submit their draft thesis, defend their thesis and then have the opportunity to make changes to their master’s thesis. Students would then submit their revised master’s thesis to receive their final grade. However, this meant that supervisors had to assess each thesis twice. First to prepare questions for the thesis defense and then to establish the final grade. This was considered to be too time consuming. Anticipating also substantial budget cuts in the years ahead, time savings had to be made in thesis supervision. Therefore, students now submit their final thesis for the defense and can no longer make changes after the defense. This time-saving operation required designing a new protocol to assess students’ performance during the defense and establish how it affects the final grade.

Appendix 1.

Technology Statements and Declaration based on AI index Tilburg

	Level in AI use(s) in product(s)	Question	Yes/No	List the GenAI tools/services (if applicable), by completing the statement
1	No AI	Do you confirm that you did not use any GenAI tools/services at any point during the assessment ?		Not applicable
2	AI-assisted editing	Did you use any GenAI tools/services to help you in editing, including spelling, punctuation, grammar, syntax, and referencing?		During the preparation of this work, the student(s) used [NAME TOOL / SERVICE / VERSION OF AI TOOL] in order to [REASON]. The following parts of the assignment were affected/generated by AI tool usage: [INTRODUCTION / METHODS / RESULTS / DISCUSSION]. After using this tool/service, [NAME STUDENT(S)] evaluated the validity of the tool's outputs, including the sources that GenAI tools have used, and edited the content as needed. As a consequence, [NAME STUDENT (S)] take(s) full responsibility for the content of their work.
3	AI-assisted idea generation and structuring	Did you use any GenAI tools to develop or refine initial ideas, materials (e.g., interview guides, stimuli for experiments), paraphrasing, structures, or outlines?		
4	AI task completion, human evaluation	Did you use any GenAI tools to complete certain elements of the task (e.g., literature search, data analyses, audio transcription)?		
5	Full AI	Did you use any GenAI tools as a 'co-pilot' to complete your work?		

In addition to this statement, a lecturer may ask students to include references in the text, for example using the APA method: <https://auteursrechten.nl/wp-content/uploads/2024/10/The-APA-guidelines-explained-Generative-AI.pdf>

GLOSSARY

Assessment

Any systematic method of obtaining information on student learning to support inferences about knowledge, abilities or skills. It is used either to make an official judgement or decision about what has been achieved (summative assessment), such as assigning grades or awarding certification, and/or to give feedback to support and improve future learning (formative assessment). High-quality assessment requires alignment with learning outcomes and learning activities. See also **Formative assessment** and **Summative assessment**.

Assessment literacy

The knowledge and skills required to design, administer, analyze, and interpret assessments effectively. Assessment Literacy includes understanding of test design, reliability, validity, biases, accommodations, and appropriate use of test results. Assessment-literate educators support valid score interpretations, mitigate threats to validity, and make data-driven decisions about instruction and policy (adapted from NCME glossary, 2026).

Classical test theory (CTT)

A foundational psychometric framework that models an observed score as the sum of a true score and random measurement error. It provides practical indices such as item difficulty (p-value), item discrimination (item-total correlation), and reliability. CTT is widely used in Dutch higher education because it requires modest sample sizes and simple computations. However, its parameters are sample-dependent. The same assessment will have a higher mean p-value with a student group of high mean ability as compared with a student group of lower mean ability.

Consequential validity

The extent to which the effects of the use of an assessment—both intended and unintended—support or undermine the validity of score interpretations. Consequential validity includes issues such as fairness, equity, washback, and social impact. These issues are integral to the validity argument.: If an assessment produces harmful or inequitable outcomes unrelated to the construct, its use is invalid.

Content validity

The representativeness and relevance of the content of an assessment relative to the intended domain of knowledge or skills. It is established through systematic domain analysis, assessment blueprints, and expert judgment. High content validity ensures that the assessment samples the full breadth and depth of the construct, avoiding both underrepresentation (too narrow) and contamination (irrelevant content).

Construct

A theoretical attribute—such as critical thinking, reading comprehension, or self-efficacy—that cannot be observed directly but is inferred from data: registrations of behavior or response in assessments. A construct should be conceptually (what it is) and operationally defined (how it is observed/measured). Poorly defined constructs lead to construct underrepresentation or contamination.

Construct validity

The degree to which evidence supports the claim that an assessment measures the intended construct and not something else. It integrates multiple sources of evidence: internal structure (factor analysis), response processes (cognitive interviews), relations to other variables (convergent/discriminant validity), and consequences of use. Construct validity is not a “type” of validity, but the unifying concept underlying all validity evidence.

Constructive alignment

An approach to course and assessment design in which intended learning outcomes, teaching activities, and assessments are coherently developed and maintained with the purpose of having students learn exactly what the course intends them to learn and to ascertain that they reach the level of performance or standards they are expected to reach after completing a course. Strong alignment enhances validity and instructional quality and supports the PDCA-cycle.

Correction for guessing

An adjustment of individual test scores or pass-or-fail thresholds in multiple-choice tests to account for the influence of guessing. Please note the following:

Corrected test scores for guessing

A formula for calculating the score corrected for guessing in multiple-choice tests is:

$$X_C = X - \frac{F}{(a - 1)}$$

Here, X_C is the score corrected for guessing, X is the original score, F is the number of incorrect answers, and a is the number of answer alternatives. Using this formula, the number of correct answers resulting from guessing is estimated based on the number of incorrect answers. The formula assumes that all incorrect answers are the result of pure guessing. This is a very strict and, in most cases, unrealistic assumption, making the correction difficult to justify. Furthermore, using this correction will reduce the reliability of the assessment scores and will also have a negative impact on the validity of the assessment. Risk-averse candidates, of which a relatively large part are women, tend to skip more items than other candidates. So, an assessment using this type of correction is biased against risk-averse candidates.

Corrected cut-off score for guessing

A better method is to take guessing into account when determining the cut-off score/ pass-or-fail threshold. The expected score students would achieve if they guessed everything is subtracted from the maximum score and a threshold is set on the remaining points. Suppose a test contains 40 four-choice items, and the threshold is set at 70% true knowledge. The guessing score then would be 10 ($\frac{1}{4}$ times 40). This leaves 30 score points of which 70% has to be achieved, which is 21 points. The cut-off score is then set at $10 + 21 = 31$.

Criterion validity

The degree to which assessment scores relate to an external criterion that represents the same or a related construct. It includes predictive validity (future performance) and concurrent validity (current performance). Strong criterion validity supports practical utility.

Cut-off score

A numerical threshold used to classify examinees into categories (e.g., pass-or-fail, basic/proficient/advanced). A cut-off score must be defensible, transparent, and linked to a formal standard-setting procedure. Poorly set cut-off scores distort decisions, inflate error rates, and undermine fairness. See also **Passing score**.

Diagnostic assessment

Diagnostic assessment identifies specific misconceptions, skill gaps, or learning needs. It is granular, low-stakes, and designed to inform targeted instruction. Diagnostic assessments require high content validity and fine-grained task design.

Distractors

The incorrect options on multiple-choice items. Distractor analysis provides information on the attraction of the incorrect options. Functional distractors attract low-performing students more than high-performing ones. Non-functioning distractors reduce the reliability of a multiple-choice test.

Fairness

Providing equal opportunity to all examinees to demonstrate the construct being measured by the assessment. It includes accessibility, absence of bias, equitable consequences, and appropriate accommodations. Fairness is a core requirement for validity.

Formative Assessment

The process of gathering and interpreting evidence of student learning during instruction to inform the next steps in teaching and learning. Its purpose is to provide feedback that helps students understand where they are in their learning, where they need to go, and how to close the gap. Formative assessment supports self-regulation and guides instructional adjustments. It is not intended for grading. Effective formative assessment requires clear learning goals, understandable success criteria, and actionable feedback.

Generalizability Theory

A psychometric framework that decomposes measurement error into multiple so-called facets (items, raters, occasions), providing a richer understanding of reliability than CTT. Like IRT, generalizability theory requires large samples, specialized software, and technical expertise. It is therefore highly valuable for large-scale or high-stakes assessments, but seldom practical or necessary for assessments in a university setting.

Inter-rater reliability

The correlation between the scores of independent raters marking the same performance, essential for rating essays, portfolios, and performance tasks. Training, calibration, and clear rubrics improve inter-rater reliability. See also **Rubric**. Note that there is an essential difference between inter-rater reliability and inter-rater agreement. The first one refers to measures of consistency in rank-ordering or relative scoring, rather than strict accordance. The latter to exact sameness of scores.

Intra-rater reliability

The correlation between scores by the same rater marking the same performance across time. It reflects how consistent one assessor is over time when applying the same criteria. Drift, fatigue, and bias can reduce intra-rater reliability. Monitoring and recalibration are essential to warrant an acceptable intra-rater reliability. Note that there is an essential difference between intra-rater reliability and intra-rater agreement. The first one refers to measures

of consistency in rank-ordering or relative scoring. The latter to measures of exact sameness of scores for an individual assessor with identical responses.

Item response theory (IRT)

A psychometric framework that models the probability of a correct response to an item as a function of a student's ability and characteristics of the item, such as its difficulty and how well it distinguishes between stronger and weaker students. By focusing on how individual items behave, IRT shows how much information each item provides about student ability. Although it is more powerful than CTT, IRT requires large samples, specialized software, and technical expertise. It is therefore highly valuable for large-scale or high-stakes assessments, but seldom practical or necessary for assessments in a university setting.

Item analysis

An activity that provides quantitative information on item difficulty, discrimination, and distractor functioning, with the aim of checking and improving test quality. It identifies flawed items and supports test refinement.

Item difficulty

The proportion of examinees answering an item correctly, when the maximum score on an item equals 1. With maximum scores above 1, the item difficulty is the mean score divided by the maximum score. Note that item difficulty as defined in CTT depends on the mean ability in the candidate population. The higher the mean ability in the candidate population, the higher the item difficulty. Note that p-value is another term used and that 'easiness' would be a better term, because the direction in which this statistic runs is counterintuitive. See also **P-value**.

Item discrimination

A measure from CTT that indicates how well an item differentiates between high- and low-ability examinees.

Items

The smallest scoring units in an assessment, such as questions or tasks. Item quality directly affects reliability, validity, and fairness.

Learning outcomes

A set of descriptions of the specific knowledge, abilities, or skills that students are expected to gain from a course. They make clear what a student has to be able to do after completing the course. A well-written outcome typically includes three elements. It contains an action verb that describes observable behavior. It also contains a concrete and clear description of what the student will be able to do under what conditions. And it describes the level of performance or standards a student is expected to reach after completing the course.

P-value

The proportion of candidates answering a dichotomous (binary-score; 0-1) item correctly, divided by the total number of candidates.

Passing score

A numerical threshold used to classify examinees into the categories pass and fail, basic/proficient/advanced). A passing score must be defensible, transparent, and linked to a formal standard-setting procedure. Poorly set passing scores distort decisions, inflate error rates, and undermine fairness. See also Cut-off score.

Reliability

The extent to which assessment scores are free from random measurement error. Reliable scores are dependable, stable, and reproducible across replications of the measurement process. Because measurement error can arise from multiple sources—such as items, raters, or testing occasions—reliability is evaluated through evidence of consistency across administrations, across raters, and across different but equivalent sets of items or tasks. See also **Standard error of measurement (SEM)**.

Rubric

A structured scoring guide describing performance levels on the dimensions of a task. A rubric supports consistency and transparency of ratings, and provides concrete feedback to students.

Summative assessment

Summative assessment is a method of evaluating students' learning at the end of a unit, term, or course to determine the knowledge, skills, and attitudes they have attained based on learning goals or outcomes. It is used to measure overall performance and make formal decisions about students' achievement, such as assigning grades, determining course completion, or awarding certificates.

Standard error of measurement (SEM)

The quantification of the amount of random measurement error in the assessment score of a student. It is defined as the standard deviation of a student's observed scores from repeated administrations of the same assessment (or parallel forms of an assessment) under identical conditions. Because such repeated measurements are rarely feasible, the SEM is typically estimated from group data through the calculation of a reliability coefficient. The SEM indicates how much an observed score is expected to vary around the examinee's true score. See also **Reliability** and **Classical test theory (CTT)**.

Standard setting

A procedure for determining cut-off scores for performance levels for a specific assessment. In standard setting, experts judge what a minimally competent student should be able to achieve on this particular assessment. This approach is preferable to using a fixed percentage-correct rule, because it adjusts the cut-off score on the assessment to its actual difficulty and therefore yields a more valid and defensible standard.

Transparency

Accessibility and clarity of information about assessment purposes, design, administration, scoring, and interpretation. Transparent assessment practices ensure that students, educators, and other stakeholders understand what is being assessed, why it is being assessed, how performance will be judged, and how results will be used. While some technical details may be restricted for security reasons, transparency requires that the goals of assessment, the criteria for success, and the meaning of scores are communicated clearly and accurately.

Validity

The degree to which accumulated evidence and theory support the intended interpretation of assessment scores for a given use. If multiple interpretations of an assessment score for different uses are intended, validity evidence for each interpretation is needed, because score interpretations may vary across uses.

REFERENCES

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Agarwal, P., & Roediger, H. (2011). Expectancy of an open-book test decreases performance on a delayed closed-book test. *Memory*, 19, 836-852. <https://doi.org/10.1080/09658211.2011.613840>
- Ahern, A. (2007). What are the perceptions of lecturers towards using cooperative learning in civil engineering? *European Journal of Engineering Education*, 32(5), 517-526. <https://doi.org/10.1080/03043790701433152>
- Ahmad, E., & Alammery, A. (2022). Towards unified management of software capstone projects in Saudi universities: A survey-based study. *Arab Gulf Journal of Scientific Research*, 40(2), 118-138. <https://doi.org/10.1108/AGJSR-04-2022-0037>
- Ajjawi, R., & Boud, D. (2018). Examining the nature and effects of feedback dialogue. *Assessment & Evaluation in Higher Education*, 43(7), 1106-1119. <https://doi.org/10.1080/02602938.2018.1434128>
- Ajjawi, R., Tai, J., Huu Nghia, T. L., Boud, D., Johnson, L., & Patrick, C.-J. (2020). Aligning assessment with the needs of work-integrated learning: The challenges of authentic assessment in a complex context. *Assessment & Evaluation in Higher Education*, 45(2), 304-316. <https://doi.org/10.1080/02602938.2019.1639613>
- Akimov, A., & Malin, M. (2020). When old becomes new: A case study of oral examination as an online assessment tool. *Assessment & Evaluation in Higher Education*, 45(8), 1205-1221. <https://doi.org/10.1080/02602938.2020.1730301>
- Alinier, G., & Oriot, D. (2022). Simulation-based education: Deceiving learners with good intent. *Advances in Simulation*, 7(1), 8. <https://doi.org/10.1186/s41077-022-00206-3>
- Allal, L. (2020). Assessment and the co-regulation of learning in the classroom. *Assessment in Education: Principles, Policy & Practice*, 27(4), 332-349. <https://doi.org/10.1080/0969594X.2019.1609411>
- Almond, R. J. (2009). Group assessment: Comparing group and individual undergraduate module marks. *Assessment & Evaluation in Higher Education*, 34(2), 141-148.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Arifiyanti, F., Soeharto, S., Amukune, S., Van Nguyen, S., Aburezeq, K., Hidayatullah, A., & Sarimanah, E. (2023). Investigating rater-student interaction, gender bias, and major bias in the assessment of research seminar presentation. *Heliyon*, 9(6), e16548. <https://doi.org/10.1016/j.heliyon.2023.e16548>
- Association of American Colleges and Universities. (2009a). *Critical thinking VALUE rubric*. <https://www.aacu.org/initiatives/value-initiative/value-rubrics/value-rubrics-critical-thinking>
- Association of American Colleges and Universities. (2009b). *Valid assessment of learning in undergraduate education (VALUE)*. <https://www.aacu.org/initiatives/value>
- Augar, N., Woodley, C. J., Whitefield, D., & Winchester, M. (2016). Exploring academics' approaches to managing team assessment. *International Journal of Educational Management*, 30(6), 1150-1162. <https://doi.org/10.1108/ijem-06-2015-0087>
- Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & van der Vleuten, C. P. M. (2007). Evaluating assessment quality in competence-based education: A qualitative comparison of two frameworks. *Educational Research Review*, 2(2), 114-129. <https://doi.org/10.1016/j.edurev.2007.06.001>

- Bain, J., Ballantyne, R., Mills, C., & Lester, N. (2002). *Reflecting on practice: Student teachers' perspectives*. Post Pressed.
- Balloo, K., Evans, C., Hughes, A., Zhu, X., & Winstone, N. (2018). Transparency isn't spoon-feeding: How a transformative approach to the use of explicit assessment criteria can support student self-regulation. *Frontiers in Education*, 3, 69. <https://doi.org/10.3389/feduc.2018.00069>
- Bamber, M. (2015). The impact on stakeholder confidence of increased transparency in the examination assessment process. *Assessment & Evaluation in Higher Education*, 40(4), 471-487. <https://doi.org/10.1080/02602938.2014.921662>
- Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on Psychological Science*, 1(2), 164-180. <https://doi.org/10.1111/j.1745-6916.2006.00011.x>
- Bastola, M. N., & Hu, G. (2024). Engagement with supervisory feedback on master's theses: Do supervisors and students see eye to eye? *Assessing Writing*, 60, 100841. <https://doi.org/10.1016/j.asw.2024.100841>
- Baume, D. (2001). *A briefing on assessment of portfolios: Assessment series No. 6*. LTSN Generic Centre.
- Bayley, T., Maclean, K. D., & Weidner, T. (2024). Back to the future: Implementing large-scale oral exams. *Management Teaching Review*, 11(1), 159-170. <https://doi.org/10.1177/23792981241267744>
- Bearman, M., & Ajjawi, R. (2018). From "seeing through" to "seeing with": Assessment criteria and the myths of transparency. *Frontiers in Education*, 3, 96.
- Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*. Advance online publication. <https://doi.org/10.1080/02602938.2024.2335321>
- Beller, M., & Gafni, N. (2000). Can item format (multiple choice vs. open-ended) account for gender differences in mathematics achievement? *Sex Roles*, 42(1), 1-21.
- Bennion, J., Cannon, B., Hill, B., Nelson, R., & Ricks, M. (2020). Asking the right questions: Using reflective essays for experiential assessment. *Journal of Experiential Education*, 43(1), 37-54.
- Berk, R. A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research*, 56(1), 137-172. <https://doi.org/10.3102/00346543056001137>
- Bezcioglu-Göktolga, I., Hsiao, Y.-P., & Grootjans, L. (2025). Transforming group work: The impact of generative AI integration in an introductory research methods course. *Journal of Research in Innovative Teaching & Learning*. Advance online publication. <https://doi.org/10.1108/jrit-08-2025-0198>
- Biesta, G. (2015). *Good education in an age of measurement: Ethics, politics, democracy*. Taylor & Francis.
- Biesta, G. (2021). *World-centred education: A view for the present*. Taylor & Francis.
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university: What the student does*. Open University Press.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. Academic Press.
- Biggs, J. B., & Tang, K. C. C. (1998). Assessment by portfolio: Constructing learning and designing teaching. In *Curriculum and assessment for Hong Kong: Two components, one system* (pp. 443-462). Open University of Hong Kong Press.
- Black, P., Harrison, C., & Lee, C. (2003). *Assessment for learning: Putting it into practice*. McGraw-Hill Education.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31. <https://doi.org/10.1007/s11092-008-9068-5>
- Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81-90.
- Blijlevens, S. (2019). *Performance behaviour in elite sports*. University of Groningen.
- Blokvoort, S., Drissen, Y., & Jacobs, H. (Eds.). (2025). *Towards a neuro-inclusive assessment of presentation skills: Empowering authentic presenters*. Tilburg University.

- Bloxham, S. (2012). 'You can see the quality in front of your eyes': Grounding academic standards between rationality and interpretation. *Quality in Higher Education*, 18(2), 185-204. <https://doi.org/10.1080/13538322.2012.711071>
- Bloxham, S., & Boyd, P. (2007). Providing effective feedback. In *Developing effective assessment in higher education: A practical guide* (pp. 103-116). McGraw-Hill Education. <http://ebookcentral.proquest.com/lib/uvtilburg-ebooks/detail.action?docID=332673>
- Bloxham, S., & Campbell, L. (2010). Generating dialogue in assessment feedback: Exploring the use of interactive cover sheets. *Assessment & Evaluation in Higher Education*, 35(3), 291-300. <https://doi.org/10.1080/02602931003650045>
- Bloxham, S., den-Outer, B., Hudson, J., & Price, M. (2016). Let's stop the pretence of consistent marking: Exploring the multiple limitations of assessment criteria. *Assessment & Evaluation in Higher Education*, 41(3), 466-481. <https://doi.org/10.1080/02602938.2015.1024607>
- Bologna Working Group on Qualifications Frameworks. (2005). *A framework for qualifications of the European higher education area*. https://ehea.info/media.ehea.info/file/WG_Frameworks_qualification/71/0/050218_QF_EHEA_580710.pdf
- Bosco, A. M., & Ferns, S. (2014). Embedding of authentic assessment in work-integrated learning curriculum. *Asia-Pacific Journal of Cooperative Education*, 15(4), 281-290.
- Bosher, S. (2003). Barriers to creating a more culturally diverse nursing profession: Linguistic bias in multiple-choice nursing exams. *Nursing Education Perspectives*, 24(1), 25-34. https://journals.lww.com/neponline/full-text/2003/01000/barriers_to_creating_a_more_culturally_diverse.9.aspx
- Bosmans, A. M. M. (2026). *MSc thesis marketing management*. Tilburg University. <https://uvt.osiris-student.nl/onderwijscatalogus/extern/examenprogramma/36640/1M300-2025?taal=en>
- Boud, D. (1995). *Enhancing learning through self-assessment*. Kogan Page.
- Boyer Commission on Educating Undergraduates. (1998). *Reinventing undergraduate education: A blueprint for American research universities*.
- Boyer, E. L. (1997). *Scholarship reconsidered: Priorities of the professoriate*. Wiley.
- Branch, W. T. J., & Paranjape, A. (2002). Feedback and reflection: Teaching methods for clinical settings. *Academic Medicine*, 77(12 Part 1), 1185-1188. <https://doi.org/10.1097/00001888-200212000-00005>
- Bridges, P., Cooper, A., Evanson, P., Haines, C., Jenkins, D., Scurry, D., Woolf, H., & Yorke, M. (2002). Coursework marks high, examination marks low: Discuss. *Assessment & Evaluation in Higher Education*, 27(1), 35-48. <https://doi.org/10.1080/02602930120105045>
- Brokerhof, I. M., Bal, P. M., Jansen, P. G., & Solinger, O. N. (2018). Fictional narratives and identity change: Three pathways through which stories influence the dialogical self. In M. M. Puchalska-Wasy, P. K. Oleś, & H. J. M. Hermans (Eds.), *Dialogical self: Inspirations, considerations and research* (pp. 29-57). Towarzystwo Naukowe Katolickiego Uniwersytetu Lubelskiego Jana Pawła II.
- Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education*, 3. <https://doi.org/10.3389/feduc.2018.00022>
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5-12. <https://doi.org/10.1111/j.1745-3992.2003.tb00139.x>
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803-848. <https://doi.org/10.3102/0034654316672069>
- Brothen, T. (2012). Time limits on tests: Updating the 1-minute rule. *Teaching of Psychology*, 39(4), 288-292. <https://doi.org/10.1177/0098628312456630>
- Brown, G. T. L., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: Directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, 22(4), 444-457. <https://doi.org/10.1080/0969594X.2014.996523>

- Bruce, I. (2008). *Academic writing and genre: A systematic analysis*. Bloomsbury Academic.
- Bruster, B. G., & Peterson, B. R. (2013). Using critical incidents in teaching to promote reflective practice. *Reflective Practice*, 14(2), 170-182.
- Buehler, M. J., & Schneider, L. U. (2009). Speak up! Oral examinations and political science. *Journal of Political Science Education*, 5(4), 315-331. <https://doi.org/10.1080/15512160903253277>
- Burton, S. J., Sudweeks, R. R., Merrill, P. F., & Wood, B. (1990). *How to prepare better multiple-choice test items: Guidelines for university faculty*. Brigham Young University.
- Buxton, M., Phillippi, J. C., & Collins, M. R. (2015). Simulation: A new approach to teaching ethics. *Journal of Midwifery & Women's Health*, 60(1), 70-74.
- Capelo, C., Pereira, R., & Dias, J. F. (2021). Teaching the dynamics of the growth of a business venture through transparent simulations. *The International Journal of Management Education*, 19(3), 100549. <https://doi.org/10.1016/j.ijme.2021.100549>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315-1325. <https://doi.org/10.1080/02602938.2018.1463354>
- CAST. (2026). *CAST releases new UDL guidelines, adding readability considerations*. <https://readabilitymatters.org/articles/cast-releases-new-udl-guidelines-adding-readability-considerations>
- Charlton, N., & Newsham-West, R. (2023). Program-level assessment planning in Australia: The considerations and practices of university academics. *Assessment & Evaluation in Higher Education*, 48(6), 820-833. <https://doi.org/10.1080/02602938.2022.2134553>
- Cheng, B., Wang, M., Yang, S. J., & Peng, J. (2011). Acceptance of competency-based workplace e-learning systems: Effects of individual and peer learning support. *Computers & Education*, 57(1), 1317-1333.
- Chester, M. D. (2003). Multiple measures and high-stakes decisions: A framework for combining measures. *Educational Measurement: Issues and Practice*, 22(2), 32-41.
- Ching, H., & Da Silva, E. (2017). The use of Bloom's taxonomy to develop competences in students of a business undergrad course. *Journal of International Business Education*, 12, 107-126.
- Chou, C.-Y., & Zou, N.-B. (2020). An analysis of internal and external feedback in self-regulated learning activities mediated by self-regulated learning tools and open learner models. *International Journal of Educational Technology in Higher Education*, 17(1), 55. <https://doi.org/10.1186/s41239-020-00233-y>
- Chust-Hernández, P., Martínez-Arnau, F. M., Alcalá-Dávalos, L., & Pérez-Ros, P. (2025). An online workshop is effective in improving the oral defense of the bachelor's thesis in nursing students: A real-world cohort study. *Nurse Education in Practice*, 89, 104624. <https://doi.org/10.1016/j.nepr.2025.104624>
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Sage Publications.
- Coffey, M., & Gibbs, G. (2001). The evaluation of the Student Evaluation of Educational Quality Questionnaire (SEEQ) in UK higher education. *Assessment & Evaluation in Higher Education*, 26(1), 89-93. <https://doi.org/10.1080/02602930020022318>
- Cohen-Schotanus, J., & Van der Vleuten, C. (2010). A standard setting method with the best performing students as point of reference: Practical and affordable. *Medical Teacher*, 32, 154-160. <https://doi.org/10.3109/01421590903196979>
- Connell, R. (2019). *The good university: What universities actually do and why it's time for radical change*. Bloomsbury Academic.
- Connelly, V., Dockrell, J. E., & Barnett, J. (2005). The slow handwriting of undergraduate students constrains overall performance in exam essays. *Educational Psychology*, 25(1), 99-107. <https://doi.org/10.1080/0144341042000294912>
- Cook, D. A., Zendejas, B., Hamstra, S. J., Hatala, R., & Brydges, R. (2014). What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Advances in Health Sciences Education*, 19(2), 233-250. <https://doi.org/10.1007/s10459-013-9458-4>

- Corbin, T., Bearman, M., Boud, D., & Dawson, P. (2026). The wicked problem of AI and assessment. *Assessment & Evaluation in Higher Education*, 51(4), 736-752. <https://doi.org/https://doi.org/10.1080/02602938.2025.2553340>
- Corbin, T., Walton, J., Bannister, P., & Deranty, J.-P. (2026). On the essay in a time of GenAI. *Educational Philosophy and Theory*, 58(3), 198-210. <https://doi.org/10.1080/00131857.2025.2572802>
- Cremer, P. (2005). Should student learning journals be assessed? *Assessment & Evaluation in Higher Education*, 30(3), 287-296. <https://doi.org/10.1080/02602930500063850>
- Curtis, G. J. (2025). The two-lane road to hell is paved with good intentions: Why an all-or-none approach to generative AI, integrity, and assessment is insupportable. *Higher Education Research & Development*, 44(8), 2151-2158. <https://doi.org/https://doi.org/10.1080/07294360.2025.2476516>
- Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, 21(2), 149-166.
- Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, 49(7), 1005-1016. <https://doi.org/10.1080/02602938.2024.2386662>
- De Groot, A. (1970). Some badly needed non-statistical concepts in applied psychometrics. *Nederlands Tijdschrift voor de Psychologie*, 25, 360-376.
- de Jong, T., Lazonder, A., Pedaste, M., & Zacharia, Z. (2018). Simulations, games, and modeling tools for learning. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 256-266). Routledge.
- De Jonge Akademie. (2021). *A smarter academic year*. <https://www.dejongeakademie.nl/en/projects/2043281.aspx>
- Deming, W. E. (2023). *PDSA cycle*. The W. Edwards Deming Institute. <https://deming.org>
- Dobson, J. L. (2008). The use of formative online quizzes to enhance class preparation and scores on summative exams. *Advances in Physiology Education*, 32(4), 297-302. <https://doi.org/10.1152/advan.90162.2008>
- Dobson, S. R., & Fudiyartanto, F. A. (2023). Challenging the culture of formative assessment: A critical appreciation of the work of Royce Sadler. In S. R. Dobson & F. A. Fudiyartanto (Eds.), *Transforming assessment in education: The hidden world of language games* (pp. 143-163). Springer International Publishing. https://doi.org/10.1007/978-3-031-26991-2_6
- dos Santos, J. A., & Gomide, T. R. (2013). Merging the case method and simulation in management education: Is it possible? *Developments in Business Simulation and Experiential Learning*, 40, 61-67.
- Dotson, W. H. (2010). *Investigating the variables in a mock exam study session designed to improve student exam performance in an undergraduate behavior modification and therapy course*. University of Kansas.
- Doyle, E., & Buckley, P. (2022). The impact of co-creation: An analysis of the effectiveness of student authored multiple choice questions on achievement of learning outcomes. *Interactive Learning Environments*, 30(9), 1726-1735. <https://doi.org/10.1080/10494820.2020.1777166>
- DuBois, P. H. (1970). *A history of psychological testing*. Allyn and Bacon.
- Duchi, L., Wijnia, L., Loyens, S., & Paas, F. (2025). Redefining education: The role of study crafting in enhancing students' well-being, self-direction, and academic achievement. *Learning and Individual Differences*, 118, 102608. <https://doi.org/10.1016/j.lindif.2024.102608>
- Dutch Education Inspectorate. (2025). *Further ahead: Examination boards in a changing higher education (short version)* [Verder vooruit: Examencommissies in een veranderend hoger onderwijs (Verkorte versie)]. Ministry of Education, Culture and Science [Ministerie van Onderwijs, Cultuur en Wetenschap (OCW)].
- Dwyer, S. C. (2001). Constructing portfolios together: Promoting student/faculty metacognition. *Journal of Teaching and Learning*, 1(1), 19-29.
- Eaton, S. E. (2024). Decolonizing academic integrity: Knowledge caretaking as ethical practice. *Assessment & Evaluation in Higher Education*, 49(7), 962-977. <https://doi.org/10.1080/02602938.2024.2312918>
- Edgeworth, F. Y. (1888). The statistics of examinations. *Journal of the Royal Statistical Society*, 51(3), 599-635.

- Elander, J., Harrington, K., Norton, L., Robinson, H., & Reddy, P. (2006). Complex skills and academic writing: A review of evidence about the types of learning required to meet core assessment criteria. *Assessment & Evaluation in Higher Education*, 31(1), 71-90. <https://doi.org/10.1080/02602930500262379>
- Ellington, H., Gordon, M., & Fowle, J. (1998). *Using games & simulations in the classroom*. Kogan Page.
- Ellis, C., & Murdoch, K. (2024). The educational integrity enforcement pyramid: A new framework for challenging and responding to student cheating. *Assessment & Evaluation in Higher Education*, 49(7), 924-934. <https://doi.org/https://doi.org/10.1080/02602938.2024.2329167>
- Elsamanoudy, A., Shehata, M., Almarabheh, A., & Alrefaie, Z. (2024). Evaluation of modified essay questions (MEQs) as an assessment tool in third-year medical students' modular summative assessment. *BMC Medical Education*, 24(1), 1445. <https://doi.org/10.1186/s12909-024-06469-w>
- Ennis, R. H. (1993). Critical thinking assessment. *Theory Into Practice*, 32(3), 179-186. <https://doi.org/10.1080/00405849309543594>
- Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology*, 70(1), 113-136. <https://doi.org/10.1348/000709900158001>
- Eraut, M. (2009). Transfer of knowledge between education and workplace settings. In H. Daniels, H. Lauder, & J. Porter (Eds.), *Knowledge, values and educational policy* (pp. 217-237). Routledge.
- Evans, J., & Kerridge, C. (2021). How to develop assessments based around teaching simulations. In C. Elliott, J. Guest, & E. Vettraino (Eds.), *Games, simulations and playful learning in business education* (pp. 88-99). Elgaronline. <https://doi.org/10.4337/9781800372702.00015>
- Facione, P. (1989). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*.
- Felix, M. S., & Smith, I. (2019). *A practical guide to dissertation and thesis writing*. Cambridge Scholars Publishing. <https://doi.org/10.65325/EB7714>
- Fellenz, M. R. (2004). Using assessment to support higher level learning: The multiple choice item development assignment. *Assessment & Evaluation in Higher Education*, 29(6), 703-719. <https://doi.org/10.1080/0260293042000227245>
- Feller, M. (1994). Open-book testing and education for the future. *Studies in Educational Evaluation*, 20(2), 235-238. [https://doi.org/10.1016/0191-491X\(94\)90010-8](https://doi.org/10.1016/0191-491X(94)90010-8)
- Fenton, A. (2025). Reconsidering the use of oral exams and assessments: An old way to move into a new future. *Educational Researcher*, 54(7), 430-436. <https://doi.org/10.3102/0013189x251333638>
- Fernsten, L., & Fernsten, J. (2005). Portfolio assessment and reflection: Enhancing learning through effective practice. *Reflective Practice*, 6(2), 303-309. <https://doi.org/10.1080/14623940500106542>
- Field, F., & de Neufville, R. (2021). *Thesis definition and preparation: Some general guidelines*. Massachusetts Institute of Technology Cambridge.
- Fisher, D., & Frey, N. (2014). *Checking for understanding: Formative assessment techniques for your classroom*. Association for Supervision & Curriculum Development.
- Florida State University. (2002). *Sequence learning objectives: How to*. https://static.aminer.org/pdf/PDF/000/272/539/place_sequence_learning_for_navigation.pdf
- Foot, L. (2016). *Oral exams: Preparing for and passing candidacy, qualifying, and graduate defenses*. Academic Press. <http://site.ebrary.com/id/11117965>
- Förster, C. E., & Rojas-Barahona, C. A. (2023). Quality criteria for developing assessment tools. In C. E. Förster (Ed.), *The power of assessment in the classroom: Improving decisions to promote learning* (pp. 207-227). Springer International Publishing. https://doi.org/10.1007/978-3-031-45838-5_9
- Francis, N., Allen, M., & Thomas, J. (2022). Using group work for assessment—an academic's perspective. <https://www.advance-he.ac.uk/sites/default/files/2022-03/Using%20group%20work%20for%20assessment%20%E2%80%93%20an%20academic%E2%80%99s%20perspective.pdf>

- Frary, R. B. (1988). Formula scoring of multiple-choice tests (correction for guessing). *Educational Measurement: Issues and Practice*, 7(2), 33-38. <https://doi.org/10.1111/j.1745-3992.1988.tb00434.x>
- Freire, P. (2020). Pedagogy of the oppressed. In J. Beck, C. Jenks, N. Keddie, & M. F. D. Young (Eds.), *Toward a sociology of education* (pp. 374-386). Routledge.
- French, S., Dickerson, A., & Mulder, R. A. (2024). A review of the benefits and drawbacks of high-stakes final examinations in higher education. *Higher Education*, 88(3), 893-918. <https://doi.org/10.1007/s10734-023-01148-z>
- Frisbie, D. A. (1988). Reliability of scores from teacher-made tests. *Educational Measurement: Issues and Practice*, 7(1), 25-35.
- Gaines, T., & Burrows, N. L. (2024). Exploring alternative assessments during COVID: Instructor experiences using oral exams. *Frontiers in Education*, 9, 1379886. <https://doi.org/10.3389/feduc.2024.1379886>
- Gannon, T. A., Olver, M. E., Mallion, J. S., & James, M. (2019). Does specialized psychological treatment for offending reduce recidivism? A meta-analysis examining staff and program variables as predictors of treatment effectiveness. *Clinical Psychology Review*, 73, 101752. <https://doi.org/10.1016/j.cpr.2019.101752>
- Garg, N., & Lee, E. (2016). The reliability of mixed-format exams in higher education. In J. Domenech, J. Lloret, M. C. Vincent-Vela, E. d. l. Poza, & E. Zuriaga (Eds.), *Advances in higher education* (pp. 105-126). Editorial Universitat Politècnica de València.
- Gehringer, E. (2004). Reuse of homework and test questions: When, why, and how to maintain security? 34th Annual Frontiers in Education, Savannah, GA, USA.
- Gellman, E. S., & Berkowitz, M. (1993). Test-item type: What students prefer and why. *College Student Journal*, 27(1), 17-26.
- Gerbner, G. (1998). Cultivation analysis: An overview. *Mass Communication and Society*, 1(3-4), 175-194. <https://doi.org/10.1080/15205436.1998.9677855>
- Gerlt, M. S., von Platten, J., Klöckner, M., Larsson, V., & Cedervall, T. (2023). Reducing the impact of bias in oral assessments. *Pedagogiska inspirationskonferensen-Genombrottet*, 12, 3-3.
- Gibbs, G., & Andrew, C. (2001). *Learning by doing: A guide to teaching and learning methods*. Geography Discipline Network.
- Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, 1, 3-31. <https://doi.org/10.1007/978-3-8348-9837-1>
- Gillespie, C. S., Ford, K. L., Gillespie, R. D., & Leavell, A. G. (1996). Portfolio assessment: Some questions, some answers, some recommendations. *Journal of Adolescent & Adult Literacy*, 39(6), 480-491. <http://www.jstor.org/stable/40014037>
- Girdharry, K., & Khachatryan, D. (2023). Meaningful writing in the age of generative artificial intelligence. *Double Helix: A Journal of Critical Thinking and Writing*, 11(1), 1-10. <https://doi.org/10.37514/DBH-J.2023.11.1.04>
- Glendinning, I. (2024). Detecting is just the start: Deciding whether to take action and generating evidence is harder. In S. Mahmud (Ed.), *Academic integrity in the age of artificial intelligence* (pp. 220-240). IGI Global. <https://doi.org/10.4018/979-8-3693-0240-8.ch012>
- Golding, C., Sharmini, S., & Lazarovitch, A. (2014). What examiners do: What thesis students should know. *Assessment & Evaluation in Higher Education*, 39(5), 563-576. <https://doi.org/10.1080/02602938.2013.859230>
- Gonsalvez, C. J., & Freestone, J. (2007). Field supervisors' assessments of trainee performance: Are they reliable and valid? *Australian Psychologist*, 42(1), 23-32. <https://doi.org/10.1080/00050060600827615>
- Gosling, D., & D'Andrea, V.-M. (2001). Quality development: A new concept for higher education. *Quality in Higher Education*, 7(1), 7-17. <https://doi.org/10.1080/13538320120045049>
- Grady, D. A. (2017). Indirect measures: Internships, careers, and competitions. In W. Christ (Ed.), *Assessing media education* (pp. 349-372). Routledge.
- Gravett, K., & Lygo-Baker, S. (2025). Affective encounters in higher education. *Studies in Higher Education*, 50(2), 228-238. <https://doi.org/10.1080/03075079.2024.2332427>

- Gravett, K., Taylor, C. A., & Fairchild, N. (2024). Pedagogies of mattering: Re-conceptualising relational pedagogies in higher education. *Teaching in Higher Education*, 29(2), 388-403. <https://doi.org/10.1080/13562517.2021.1989580>
- Grierson, L., Winemaker, S., Taniguchi, A., Howard, M., Marshall, D., & Zazulak, J. (2020). The reliability characteristics of the REFLECT rubric for assessing reflective capacity through expressive writing assignments: A replication study. *Perspectives on Medical Education*, 9(5), 281-285. <https://doi.org/10.1007/s40037-020-00611-2>
- Grund, C. K., & Schelkle, M. (2020). Developing serious games with integrated debriefing: Findings from a business intelligence context. *Business & Information Systems Engineering*, 62(2), 87-101.
- Guetzkow, J., Lamont, M., & Mallard, G. (2004). What is originality in the humanities and the social sciences? *American Sociological Review*, 69(2), 190-212. <https://doi.org/10.1177/000312240406900203>
- Guskey, T. R., Frey, N., & Fisher, D. (2024). *Grading with integrity: A research-based approach grounded in honesty, transparency, accuracy, and equity*. SAGE Publications. <https://books.google.com.tw/books?id=EfsJEQAAQBAJ>
- Haase, E. (2019). *Analyzing change in future thought with letters from the future*. University of Twente.
- Haber, E., Jemielniak, D., Kurasinski, A., & Przegalinska, A. (2025). *Using AI in academic writing and research*. Springer.
- Hacker, D. J., Keener, M. C., & Kircher, J. C. (2009). Writing is applied metacognition. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 154-172). Routledge/Taylor & Francis Group.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Routledge.
- Haladyna, T. M. (2015). Item analysis for selected-response test items. In *Handbook of test development* (pp. 392-409). Routledge.
- Haladyna, T. M. (2019). Assigning a valid and reliable grade in a course. *IDEA*, 79, 1-15.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334. https://doi.org/10.1207/S15324818AME1503_5
- Hand, L., & Clewes, D. (2000). Marking the difference: An investigation of the criteria used for assessing undergraduate dissertations in a business school. *Assessment & Evaluation in Higher Education*, 25(1), 5-21. <https://doi.org/10.1080/713611416>
- Harvey, L. (2004–2026). *Analytic Quality Glossary*. Quality Research International. <https://www.qualityresearchinternational.com/glossary/#assurance>
- Harvey, P. D. (2019). Domains of cognition and their assessment. *Dialogues Clinical Neurosciences*, 21(3), 227-237. <https://doi.org/10.31887/DCNS.2019.21.3/pharvey>
- Hatfield, S. (2009). Assessing your program-level assessment plan. *IDEA Paper*, 45, 1-9.
- Hattie, J. (2015). The applicability of visible learning to higher education. *Scholarship of Teaching and Learning in Psychology*, 1(1), 79-91.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112. <https://doi.org/10.3102/003465430298487>
- Hayes, J. R. (1980). *Identifying the organization of writing processes*. Lawrence Erlbaum Associates.
- Hayes, T. A. I. B., Fry, J., Gambrell, E., Moulds, A., & Young, K. (1981). Modified essay question or MEQ. In T. A. I. B. Hayes, J. Fry, E. Gambrell, A. Moulds, & K. Young (Eds.), *The MRCGP study book: Tests and self-assessment exercises devised by MRCGP examiners for those preparing for the exam* (pp. 49-124). Springer Netherlands. https://doi.org/10.1007/978-94-015-7174-6_3
- Heritage, M. (2016). Assessment for learning: Co-regulation in and as student–teacher interaction. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 327-343). Springer International Publishing. https://doi.org/10.1007/978-3-319-39211-0_19
- Heron, M., Baker, S., Gravett, K., & Irwin, E. (2023). Scoping academic oracy in higher education: Knotting together forgotten connections to equity and academic literacies. *Higher Education Research & Development*, 42(1), 62-77. <https://doi.org/10.1080/07294360.2022.2048635>

- Heron, M., Medland, E., Winstone, N., & Pitt, E. (2023). Developing the relational in teacher feedback literacy: Exploring feedback talk. *Assessment & Evaluation in Higher Education*, 48(2), 172-185.
<https://doi.org/10.1080/02602938.2021.1932735>
- Hertel, J. P., & Millis, B. (2023). *Using simulations to promote learning in higher education: An introduction*. Taylor & Francis.
- Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 37-56). Springer.
https://doi.org/10.1007/978-94-017-9395-7_2
- Hoekstra, R., Vugteveen, J., Warrens, M. J., & Kruyen, P. M. (2019). An empirical analysis of alleged misunderstandings of coefficient alpha. *International Journal of Social Research Methodology*, 22(4), 351-364.
<https://doi.org/10.1080/13645579.2018.1547523>
- Hogan, T. P., & Murphy, G. (2007). Recommendations for preparing and scoring constructed-response items: What the experts say. *Applied Measurement in Education*, 20(4), 427-441. <https://doi.org/10.1080/08957340701580736>
- Hsiao, Y. P., Chiu, M. S., & Van de Watering, G. (2020). Making the most of student reflections for internship supervision and assessment. EARLI SIG6-7 Online.
- Hsiao, Y. P., den Ouden, B., Venhuizen, N., & Engelen, J. (2025). *AI Index Tilburg University based on Perkins et al. (2024)'s Artificial Intelligence Assessment Scale (AIAS)*. Tilburg University.
- Hsiao, Y. P., Klijn, N., & Chiu, M. S. (2023). Developing a framework to re-design writing assignment assessment for the era of large language models. *Learning: Research and Practice*. Advance online publication.
<https://doi.org/10.1080/23735082.2023.2257234>
- Hsiao, Y. P., van de Watering, G., Heitbrink, M., Vlas, H., & Chiu, M. S. (2023). Ensuring bachelor's thesis assessment quality: A case study at one Dutch research university. *Higher Education Evaluation and Development*, 18(1), 2-16.
<https://doi.org/10.1108/HEED-08-2022-0033>
- Hsiao, Y. P., Wolff, C. E., & Rajagopal, K. (2025). Slow and connected: Deliberate GenAI-enhanced instructional design fostering the essence of critical thinking. In S. Gregory (Ed.), *Effective instructional design informed by AI* (pp. 115-146). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3693-6527-4.ch004>
- Hudson, J. N., Rienits, H., Corrin, L., & Olmos, M. (2012). An innovative OSCE clinical log station: A quantitative study of its influence on log use by medical students. *BMC Medical Education*, 12(1), 111.
<https://doi.org/10.1186/1472-6920-12-111>
- Hundey, B., Anstey, L., Cruickshank, H., & Watson, G. P. L. (2020). Mentoring faculty online: a literature review and recommendations for web-based programs. *International Journal for Academic Development*, 25(3), 232-246.
<https://doi.org/10.1080/1360144X.2020.1731815>
- Hurford, D. (2020). *Strategies for criteria-aligned, fair and inclusive oral exams*.
<https://tidsskrift.dk/dut/article/download/122579/169793/257030>
- Jackson, J., Jones, M., Steele, W., & Coiacetto, E. (2017). How best to assess students taking work placements? An empirical investigation from Australian urban and regional planning. *Higher Education Pedagogies*, 2(1), 131-150.
<https://doi.org/10.1080/23752696.2017.1394167>
- Jankowski, N., & Slotnick, R. (2015). The five essential roles of assessment practitioners. *Journal of Assessment and Institutional Effectiveness*, 5, 78-100. <https://doi.org/10.5325/jasseinsteffe.5.1.0078>
- Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4(1), 53-71.
https://doi.org/10.1207/S15327566IJCE0401_04
- Joncas, S. X., St-Onge, C., Bourque, S., & Farand, P. (2018). Re-using questions in classroom-based assessment: An exploratory study at the undergraduate medical education level. *Perspectives on Medical Education*, 7(6), 373-378.
<https://doi.org/10.1007/s40037-018-0482-1>

- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130-144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- Joughin, G. (1998). Dimensions of oral assessment. *Assessment & Evaluation in Higher Education*, 23(4), 367-378. <https://doi.org/10.1080/0260293980230404>
- Joughin, G. (2010). *A short guide to oral assessment*. Leeds Met Press in association with University of Wollongong.
- Joughin, G., Boud, D., & Dawson, P. (2019). Threats to student evaluative judgement and their management. *Higher Education Research & Development*, 38(3), 537-549. <https://doi.org/10.1080/07294360.2018.1544227>
- Kaipa, R. M. (2021). Multiple choice questions and essay questions in curriculum. *Journal of Applied Research in Higher Education*, 13(1), 16-32. <https://doi.org/10.1108/JARHE-01-2020-0011>
- Kane, M. (2013). The argument-based approach to validation. *School Psychology Review*, 42(4), 448-457. <https://doi.org/10.1080/02796015.2013.12087465>
- Kember, D., Leung, D. Y. P., Jones, A., Loke, A. Y., McKay, J., Sinclair, K., Tse, H., Webb, C., Yuet Wong, F. K., Wong, M., & Yeung, E. (2000). Development of a questionnaire to measure the level of reflective thinking. *Assessment & Evaluation in Higher Education*, 25(4), 381-395. <https://doi.org/10.1080/713611442>
- Kennette, L. N., & McGuckin, D. (2025). Best practice for online tests: How long do students actually need? *Teaching and Learning Inquiry*, 13, 1-11. <https://doi.org/10.20343/teachlearninqu.13.35>
- Kesselman-Turkel, J., & Peterson, F. (2004). *Test-taking strategies*. University of Wisconsin Press.
- King, N. (2016). Case-based exams for learning and assessment: Experiences in an information systems course. 2016 IEEE Global Engineering Education Conference (EDUCON), Abu Dhabi, United Arab Emirates.
- Kingston, N., & Tiemann, G. (2010). Spearman–Brown Prophecy Formula. In *Encyclopedia of Research Design* (pp. 1403-1404). SAGE Publications. <https://doi.org/https://doi.org/10.4135/9781412961288.n427>
- Kirschner, F., Paas, F., & Kirschner, P. A. (2009). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review*, 21(1), 31-42. <https://doi.org/10.1007/s10648-008-9095-2>
- Knight, P., & Yorke, M. (2003). *Assessment, learning and employability*. Society for Research into Higher Education & Open University Press.
- Konrad, R., Hall-Phillips, A., & Vila-Parrish, A. R. (2018). Are our students prepared? The impact of capstone design pedagogical approaches on student skill development during industry-sponsored fieldwork. *INFORMS Transactions on Education*, 18(3), 183-193. <https://doi.org/10.1287/ited.2018.0198>
- Korthagen, F., & Vasalos, A. (2005). Levels in reflection: Core reflection as a means to enhance professional growth. *Teachers and Teaching*, 11(1), 47-71. <https://doi.org/10.1080/1354060042000337093>
- Korthagen, F. A. J. (2014). Promoting core reflection in teacher education: Deepening professional growth. In *International teacher education: Promising pedagogies (Part A)* (Vol. 22, pp. 73-89). Emerald Group Publishing. <https://doi.org/10.1108/S1479-368720140000022007>
- Kuh, G. D. (2008). High-impact educational practices: What they are, who has access to them, and why they matter. In G. D. Kuh (Ed.), *High-impact educational practices: What they are, who has access to them, and why they matter* (pp. 13-22). Association of American Colleges and Universities.
- Kun, A. I., Farkas, J., & Juhász, C. (2023). Dunning–Kruger effect in knowledge management examination of BSc level business students. *International Journal of Engineering and Management Sciences*, 8(1), 14-21. <https://doi.org/10.21791/IJEMS.2023.1.3>
- Kwan, C., & Lo, C. K. M. (2023). Evaluating the portfolio as a social work capstone project: A case study in Hong Kong. *Social Work Education*, 42(1), 145-160. <https://doi.org/10.1080/02615479.2022.2053099>
- Lam, R. (2018). *Portfolio assessment for the teaching and learning of writing*. Springer Nature Singapore.
- Lean, J., Moizer, J., Towler, M., & Abbey, C. (2006). Simulations and games: Use and barriers in higher education. *Active Learning in Higher Education*, 7(3), 227-242. <https://doi.org/10.1177/1469787406069056>
- Leenheer, J., Hsiao, Y.-P. A., Beekmans, N., & Hille, T. (2026, June 5). Comparative judgement of co-intelligence products. Festival Leren van Toetsen, Rotterdam, The Netherlands.

- Lejk, M., Wyvill, M., & Farrow, S. (1996). A survey of methods of deriving individual grades from group assessments. *Assessment & Evaluation in Higher Education*, 21(3), 267-280. <https://doi.org/10.1080/0260293960210306>
- Leyland, R., Heath, M., Collett, T., Drake, E., & Neve, H. (2024). 'How far I've come': The value of reviewing past written reflections. *Reflective Practice*, 25(6), 826-838. <https://doi.org/10.1080/14623943.2024.2406969>
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20(8), 15-21. <https://doi.org/10.2307/1176232>
- Lins de Holanda Coelho, G., Hanel, P. H. P., & Wolf, L. J. (2020). The very efficient assessment of need for cognition: Developing a six-item version. *Assessment*, 27(8), 1870-1885. <https://doi.org/10.1177/107319118793208>
- Lipson, C. (2007). *How to write a BA thesis: A practical guide from your first ideas to your finished paper*. University of Chicago Press.
- Liu, O. L., Frankel, L., & Roohr, K. C. (2014). Assessing critical thinking in higher education: Current state and directions for next-generation assessment. *ETS Research Report Series*, 2014(1), 1-23. <https://doi.org/10.1002/ets2.12009>
- Liu, Y., Xu, G., Yuan, S., Zhou, C., & Wang, C. (2024). Assessment as learning: Evidence based on meta-analysis and quantitative ethnography research. *Studies in Educational Evaluation*, 83, 101423. <https://doi.org/10.1016/j.stueduc.2024.101423>
- Livingston, S. (2009). Constructed-response test questions: Why we use them; how we score them. *R&D Connections*(11), 1-8.
- Lovitts, B. E. (2023). *Making the implicit explicit: Creating performance expectations for the dissertation*. Taylor & Francis.
- Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the "originality" of students' work. *Assessment & Evaluation in Higher Education*, 49(5), 651-664. <https://doi.org/10.1080/02602938.2024.2309963>
- Lyons, N. (2009). Approaches to portfolio assessment of complex evidence of reflection and reflective practice. In N. Lyons (Ed.), *Handbook of reflection and reflective inquiry: Mapping a way of knowing for professional reflective inquiry* (pp. 473-488). Springer.
- M'Baya, A., Laval, J., Moalla, N., Ouzrout, Y., & Bouras, A. (2016). Ontology based system to guide internship assignment process. 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Naples, Italy.
- MacKenzie, A. (2024, October 22-24). Incorporating reflection questions on final exams. Teaching Professor Online Conference, Online.
- Maclellan, E. (2008). The significance of motivation in student-centred learning: A reflective case study. *Teaching in Higher Education*, 13(4), 411-421. <https://doi.org/10.1080/13562510802169681>
- Madaus, G. F., & O'Dwyer, L. M. (1999). A short history of performance assessment: Lessons learned. *Phi Delta Kappan*, 80(9), 688.
- Maddamsetti, J. (2025). Using metaphors as tools for critical reflection: Elementary preservice teachers' accounts in the context of social justice teacher education. *Reflective Practice*, 26(4), 582-594. <https://doi.org/10.1080/14623943.2025.2494309>
- Magnusson, J., & Zackariasson, M. (2019). Student independence in undergraduate projects: Different understandings in different academic contexts. *Journal of Further and Higher Education*, 43(10), 1404-1419. <https://doi.org/10.1080/0309877X.2018.1490949>
- Maney, J. K. (1999). The role of technology in education: Reality, pitfalls, and potential. In G. J. Cizek (Ed.), *Handbook of educational policy* (pp. 387-415). Academic Press. <https://doi.org/10.1016/B978-012174698-8/50043-6>
- Marissa, K. (2021). Using an online social annotation tool in a content-based instruction (CBI) classroom. *International Journal of TESOL Studies*, 3(2), 5-22. <https://doi.org/10.46451/ijts.2021.06.02>
- Mayo, J. A. (2010). *Constructing undergraduate psychology curricula: Promoting authentic learning and assessment in the teaching of psychology*. American Psychological Association.

- McArthur, J. (2023). Rethinking authentic assessment: Work, well-being, and society. *Higher Education*, 85(1), 85-101. <https://doi.org/10.1007/s10734-022-00822-y>
- McMillan, J., Rodriguez, M. C., & Haladyna, T. M. (2013). Writing selected-response items for classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 293-311). SAGE Publications. <https://doi.org/10.4135/9781452218649.n17>
- Meijer, H., Hoekstra, R., Brouwer, J., & Strijbos, J.-W. (2020). Unfolding collaborative learning assessment literacy: A reflection on current assessment methods in higher education. *Assessment & Evaluation in Higher Education*, 45(8), 1222-1240. <https://doi.org/10.1080/02602938.2020.1729696>
- Mensink, P. J., & King, K. (2020). Student access of online feedback is modified by the availability of assessment marks, gender and academic performance. *British Journal of Educational Technology*, 51(1), 10-22. <https://doi.org/10.1111/bjet.12752>
- Mertler, C. (2016). *Classroom assessment: A practical guide for educators*. Taylor & Francis.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13-103). American Council on Education and Macmillan Publishing Company.
- Meuffels, H. L. M. (2004a). Cijfergeven over de grens [Grading across borders]. *Examens*, 2(2), 15-17.
- Meuffels, H. L. M. (2004b). De verguisde cijferaar: Cijfers en cijferschalen [The maligned grader: Grades and grading scales]. *Examens*, 3(3), 9-11.
- Michaelsen, L. K., & Sweet, M. (2008). The essential elements of team-based learning. *New Directions for Teaching and Learning*, 2008(116), 7-27. <https://doi.org/10.1002/tl.330>
- Milano, S., McGrane, J. A., & Leonelli, S. (2023). Large language models challenge the future of higher education. *Nature Machine Intelligence*, 5(4), 333-334. <https://doi.org/10.1038/s42256-023-00644-2>
- Milius, J. (2007). *Schriftelijk tentamineren: Een draaiboek voor docenten in het hoger onderwijs* [Written examinations: A guide for teachers in higher education]. IVLOS, Universiteit Utrecht.
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine*, 65(9), S63-67. <https://doi.org/10.1097/00001888-199009000-00045>
- Miller, M. D., Linn, R. L., & Gronlund, N. E. (2009). *Measurement and assessment in teaching*. Merrill/Pearson.
- Ministry of Education, Culture and Science. (1992). *Higher Education and Research Act [Wet op het hoger onderwijs en wetenschappelijk onderzoek, WHW]*. The Hague, Netherlands
- Ministry of Education, Culture and Science. (2021). *Ruimte in regels: Suggesties voor een betere balans tussen kwaliteit en regeldruk in het hoger onderwijs* [Room within the rules: Suggestions for a better balance between quality and regulatory burden in higher education].
- Moore, A. A., & Kaplan, J. J. (2015). Program assessment for an undergraduate statistics major. *The American Statistician*, 69(4), 417-424. <https://doi.org/10.1080/00031305.2015.1087331>
- Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open*, 5, 100151. <https://doi.org/10.1016/j.caeo.2023.100151>
- Möwes, A. D. (2015). Planning for your oral defense. *International Journal for Innovation Education and Research*, 3(12), 51-54. <https://doi.org/10.31686/ijier.vol3.iss12.485>
- Mullen, K., & Schultz, M. (2012). Short answer versus multiple choice examination questions for first year chemistry. *International Journal of Innovation in Science and Mathematics Education*, 20, 1-18.
- Mullins, G., & Kiley, M. (2002). 'It's a PhD, not a Nobel Prize': How experienced examiners assess research theses. *Studies in Higher Education*, 27(4), 369-386. <https://doi.org/10.1080/0307507022000011507>
- Murray, R. (2015). *How to survive your viva: Defending a thesis in an oral examination*. McGraw-Hill Education.
- Naismith, L. M., Cheung, J. J., Sibbald, M., Tavares, W., Cavalcanti, R. B., Haji, F. A., & Fraser, K. L. (2019). Using cognitive load theory to optimize simulation design. In G. Chiniara (Ed.), *Clinical simulation* (pp. 129-141). Academic Press. <https://doi.org/10.1016/B978-0-12-815657-5.00010-3>
- National Council on Measurement in Education. (2026). *Assessment glossary*. NCME Assessment Glossary.

- Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T. (2026). Who gives feedback matters: Student biases towards human and AI-generated formative feedback. *Journal of Computer Assisted Learning*, 42(1), e70153. <https://doi.org/10.1111/jcal.70153>
- Neely, P., & Tucker, J. (2012). Using business simulations as authentic assessment tools. *American Journal of Business Education*, 5(4), 449-456. <https://doi.org/10.19030/ajbe.v5i4.7122>
- Nelson, J. S., Range, L. M., & Ross, M. B. (2012). A checklist to guide graduate students' writing. *International Journal of Teaching and Learning in Higher Education*, 24(3), 376-382.
- Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education*, 46(5), 756-778. <https://doi.org/10.1080/02602938.2020.1823314>
- Nicol, D., & McCallum, S. (2021). Making internal feedback explicit: Exploiting the multiple comparisons that occur during peer review. *Assessment & Evaluation in Higher Education*, 47, 1-19. <https://doi.org/10.1080/02602938.2021.1924620>
- Nicol, D. J. (2011). *Developing the students' ability to construct feedback*. The Quality Assurance Agency for Higher Education (QAA).
- Nieminen, J. H., Dollinger, M., & Zou, T. X. P. (2025). Student partnership in assessment: What works, for whom, why, and under what circumstances? *Review of Education*, 13(1), e70059. <https://doi.org/10.1002/rev3.70059>
- Nitko, A. J., & Brookhart, S. M. (2013). *Educational assessment of students*. Pearson Education.
- NLQF. (2025). *The Dutch Qualifications Framework (NLQF)*. <https://nlqf.nl/english-information/>
- Noddings, N. (1984). *Caring, a feminine approach to ethics & moral education*. University of California Press.
- Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking. The practitioners' guide to teaching thinking series*. ERIC.
- NVAO. (2018). *Assessment framework for the higher education accreditation system of the Netherlands*. The Accreditation Organisation of the Netherlands and Flanders (NVAO).
- Nylén, A., & Isomöttönen, V. (2017, November 16–19). Exploring the critical incident technique to encourage reflection during project-based learning. Proceedings of the 17th Koli Calling International Conference on Computing Education Research, Koli, Finland.
- O'Connor, E., Obst, P., Furlong, M., & Hansen, J. (2015). Using the TARL model in psychology: Supporting first and final year students to compose reflections. In M. E. Ryan (Ed.), *Teaching reflective learning in higher education: A systematic approach using pedagogic patterns* (pp. 77-91). Springer International Publishing. https://doi.org/10.1007/978-3-319-09271-3_6
- Oakley, B., Felder, R. M., Brent, R., & Elhajj, I. (2004). Turning student groups into effective teams. *Journal of Student Centered Learning*, 2(1), 9-34.
- Olsen, T., & Hunnes, J. (2024). Improving students' learning—the role of formative feedback: Experiences from a crash course for business students in academic writing. *Assessment & Evaluation in Higher Education*, 49(2), 129-141. <https://doi.org/10.1080/02602938.2023.2187744>
- Packard, T., & Austin, M. J. (2009). Using a comprehensive case-based examination to evaluate and integrate student learning in social work administration. *Journal of Teaching in Social Work*, 29(2), 204-215. <https://doi.org/10.1080/08841230802240886>
- Palm, T. (2008). Performance assessment and authentic assessment: A conceptual analysis of the literature. *Practical Assessment, Research, and Evaluation*, 13(1).
- Panadero, E., & Broadbent, J. (2018). Developing evaluative judgement: A self-regulated learning perspective. In D. Boud, R. Ajjawi, P. Dawson, & J. Tai (Eds.), *Developing evaluative judgement in higher education* (pp. 81-89). Routledge.
- Parija, S. C., & Kate, V. (2018). Thesis, dissertation and project. In S. C. Parija & V. Kate (Eds.), *Thesis writing for master's and Ph.D. program* (pp. 3-10). Springer Singapore. https://doi.org/10.1007/978-981-13-0890-1_1
- Parkes, J. (2007). Reliability as argument. *Educational Measurement: Issues and Practice*, 26(4), 2-10. <https://doi.org/10.1111/j.1745-3992.2007.00103.x>

- Parkes, J., & Zimmaro, D. (2017). *The college classroom assessment compendium: A practical guide to the college instructor's daily assessment life*. Taylor & Francis.
- Parmenter, D. A. (2009). Essay versus multiple-choice: Student preferences and the underlying rationale with implications for test construction. *Academy of Educational Leadership Journal*, 13(2).
- Paul, R., & Elder, L. (2006). *Critical thinking: Tools for taking charge of your learning and your life*. Pearson/Prentice Hall.
- Paulson, F. L. (1991). What makes a portfolio a portfolio? *Educational Leadership*, 48(5), 60-63.
- Pavlovich, K., Collins, E., & Jones, G. (2009). Developing students' skills in reflective practice: Design and assessment. *Journal of Management Education*, 33(1), 37-58. <https://doi.org/10.1177/1052562907307640>
- Pearce, J., & Chiavaroli, N. (2020). Prompting candidates in oral assessment contexts: A taxonomy and guiding principles. *Journal of Medical Education and Curricular Development*, 7, 2382120520948881. <https://doi.org/10.1177/2382120520948881>
- Pearce, L. (2004). *How to examine a thesis*. McGraw-Hill Education.
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of Generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21. <https://doi.org/10.53761/q3azde36>
- Persons, J. B., Davidson, J., & Tompkins, M. A. (2001). Using the thought record. In J. Davidson, J. B. Persons, & M. A. Tompkins (Eds.), *Essential components of cognitive-behavior therapy for depression*. (pp. 129-169). American Psychological Association. <https://doi.org/10.1037/10389-005>
- Popham, W. J. (2008). *Transformative assessment*. Association for Supervision and Curriculum Development.
- Price, M., Carroll, J., O'Donovan, B., & Rust, C. (2011). If I was going there I wouldn't start from here: A critical commentary on current assessment practice. *Assessment & Evaluation in Higher Education*, 36(4), 479-492. <https://doi.org/10.1080/02602930903512883>
- Qudrat-Ullah, H. (2014). Yes we can: Improving performance in dynamic tasks. *Decision Support Systems*, 61, 23-33. <https://doi.org/10.1016/j.dss.2014.01.009>
- Rajagopal, K., Vrieling-Teunter, E., Hsiao, Y. P., Van Seggelen-Damen, I., & Verjans, S. (2021). Guiding thesis circles in higher education: Towards a typology. *Professional Development in Education*, 50(2), 315-332. <https://doi.org/10.1080/19415257.2021.1973072>
- Ramsden, P. (2003). *Learning to teach in higher education*. Taylor & Francis.
- Rawls, J., Wilsker, A., & Rawls, R. S. (2015). Are you talking to me? On the use of oral examinations in undergraduate business courses. *Journal of the Academy of Business Education*, 16.
- Reeve, J., & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36(4), 257-267. <https://doi.org/10.1016/j.cedpsych.2011.05.002>
- Reiner, C. M., Bothell, T. W., Sudweeks, R. R., & Wood, B. (2002). *Preparing effective essay questions*. New Forums Press.
- Resnick, L., Asterhan, C., & Clarke, S. (2018). *Accountable talk: Instructional dialogue that builds the mind*. Educational Practices Series.
- Richardson, J. T. E. (2015). Coursework versus examinations in end-of-module assessment: A literature review. *Assessment & Evaluation in Higher Education*, 40(3), 439-455. <https://doi.org/10.1080/02602938.2014.919628>
- Ringeisen, T., Lichtenfeld, S., Becker, S., & Minkley, N. (2019). Stress experience and performance during an oral exam: The role of self-efficacy, threat appraisals, anxiety, and cortisol. *Anxiety, Stress, & Coping*, 32(1), 50-66. <https://doi.org/10.1080/10615806.2018.1528528>
- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, 24, 3-13. <https://doi.org/10.1111/j.1745-3992.2005.00006.x>
- Roecker, L. (2007). Using oral examination as a technique to assess student understanding and teaching effectiveness. *Journal of Chemical Education*, 84(10), 1663. <https://doi.org/10.1021/ed084p1663>
- Roediger III, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. *Psychology of Learning and Motivation*, 55, 1-36. <https://doi.org/10.1016/B978-0-12-387691-1.00001-6>

- Rogers, J., Peecksen, S., Douglas, M., & Simmons, M. (2019). Validation of a reflection rubric for higher education. *Reflective Practice*, 20(6), 761-776. <https://doi.org/10.1080/14623943.2019.1676712>
- Romme, G., & Nijhuis, J. (2002). *Collaborative learning in thesis rings*. https://www.researchgate.net/profile/Georges-Romme/publication/259145056_Collaborative_Learning_in_Thesis_Rings/links/00b4952a06e173cb9c000000/Collaborative-Learning-in-Thesis-Rings.pdf
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1), 342-363. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Russell, M. (2010). *Assessment patterns: A review of the possible consequences (Effecting sustainable change in assessment practice and experience)*. University of Hertfordshire.
- Ryan, M. (2012). Conceptualising and teaching discursive and performative reflection in higher education. *Studies in Continuing Education*, 34(2), 207-223. <https://doi.org/10.1080/0158037X.2011.611799>
- Ryan, M., & Ryan, M. (2013). Theorising a model for teaching and assessing reflective learning in higher education. *Higher Education Research & Development*, 32(2), 244-257. <https://doi.org/10.1080/07294360.2012.661704>
- Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.
- Rayner, K., Schotter, E. R., Masson, M. E. J., Potter, M. C., & Treiman, R. (2016). So much to read, so little time: How do we read, and can speed reading help? *Psychological Science in the Public Interest*, 17(1), 4-34. <https://doi.org/10.1177/1529100615623267>
- Ryle, G., & Tanney, J. (2009). *The concept of mind*. Routledge.
- Sabbaghan, S., & Eaton, S. E. (2025). Navigating the ethical frontier: Graduate students' experiences with generative AI-mediated scholarship. *International Journal of Artificial Intelligence in Education*, 35(4), 1860-1886. <https://doi.org/10.1007/s40593-024-00454-6>
- Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30(2), 175-194. <https://doi.org/10.1080/0260293042000264262>
- Sadler, D. R. (2007). Perils in the meticulous specification of goals and assessment criteria. *Assessment in Education: Principles, Policy & Practice*, 14(3), 387-392. <https://doi.org/10.1080/09695940701592097>
- Sadler, D. R. (2009). Are we short-changing our students? The use of preset criteria in assessment. *TLA Interchange*(3), 1-8.
- Sadler, D. R. (2013). Assuring academic achievement standards: From moderation to calibration. *Assessment in Education: Principles, Policy & Practice*, 20(1), 5-19.
- Saif, N., Khan, S. U., Shaheen, I., Alotaibi, F. A., Alnfai, M. M., & Arif, M. (2024). Chat-GPT; validating Technology Acceptance Model (TAM) in education sector via ubiquitous learning mechanism. *Computers in Human Behavior*, 154, Article 108097. <https://doi.org/10.1016/j.chb.2023.108097>
- Salido, A., Syarif, I., Sitepu, M. S., Suparjan, Wana, P. R., Taufika, R., & Melisa, R. (2025). Integrating critical thinking and artificial intelligence in higher education: A bibliometric and systematic review of skills and strategies. *Social Sciences & Humanities Open*, 12, 101924. <https://doi.org/10.1016/j.ssaho.2025.101924>
- Sanders, P., & Verstralen, H. (2017). Het beoordelen van toetsscores [Evaluating test scores]. In P. Sanders (Ed.), *Toetsen op school [Assessment in schools]* (pp. 145-158). Cito.
- Schellekens, L. H., Bok, H. G. J., de Jong, L. H., van der Schaaf, M. F., Kremer, W. D. J., & van der Vleuten, C. P. M. (2021). A scoping review on the notions of Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL). *Studies in Educational Evaluation*, 71, 101094. <https://doi.org/10.1016/j.stueduc.2021.101094>
- Schultz, M. (2011). Sustainable assessment for large science classes: Non-multiple choice, randomised assignments through a learning management system. *Journal of Learning Design*, 4. <https://doi.org/10.5204/jld.v4i3.80>

- Schürmann, L., Kärner, T., & Ringeisen, T. (2022). Need strength, perceived need support, stress symptomatology, and performance in the context of oral exams: A typological approach. *Frontiers in Psychology*, 13, 992314. <https://doi.org/10.3389/fpsyg.2022.992314>
- Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, 35(4), 453-472. <https://doi.org/10.1023/A:1003196224280>
- Shakoury, K., & Boers, F. (2025). Reflective practice through a metaphor evaluation task: Should multicultural society be a melting pot, a mosaic or something else? *The Curriculum Journal*, 36(2), 309-319. <https://doi.org/10.1002/curj.298>
- Shaw, S. D., & Nave, G. (2026). *Thinking—fast, slow, and artificial: How AI is reshaping human reasoning and the rise of cognitive surrender*. https://doi.org/10.31234/osf.io/yk25n_v1
- Shen, J., & Tamkin, A. (2026). *How AI impacts skill formation*. <https://doi.org/10.48550/arXiv.2601.20245>
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120. <https://doi.org/10.1007/s11336-008-9101-0>
- Sluijsmans, D. M. A., Peeters, A., Jakobs, L., & Weijzen, S. (2012). Kwaliteit van toetsing onder de loep [The quality of assessment under scrutiny]. *OnderwijsInnovatie*, 4, 17-25.
- Sluijter, C., & Straetmans, G. J. J. M. (2024). Wat betekent die score? [On the meaning of test scores]. *Examens*, 21(4), 6-16.
- Sluijter, C., & Van Berkel, H. (2023). Normeren en cijfergeven [Standard setting and grading]. In H. Van Berkel, A. Bax, D. Joosten-ten Brinke, K. Beekman, & T. Van Schilt-Mol (Eds.), *Toetsen in het hoger onderwijs [Assessment in higher education]* (5th ed., pp. 95-107). Boom.
- Smith, J. K. (2003). Reconsidering reliability in classroom assessment and grading. *Educational Measurement: Issues and Practice*, 22(4), 26-33. <https://doi.org/10.1111/j.1745-3992.2003.tb00141.x>
- Smits, N., & Klinkenberg, S. (2026). Gangbare vuistregels voor indicatoren van toetskwaliteit zijn te strikt [Common rules of thumb for indicators of assessment quality are too strict]. *Tijdschrift voor Hoger Onderwijs*, 44. <https://doi.org/10.59532/hsnnhe76>
- Song, Y., & Thuente, D. (2015). *A quantitative case study in engineering of the efficacy of quality cheat-sheets*. <https://doi.org/10.1109/FIE.2015.7344082>
- Sotiriadou, P., Logan, D., Daly, A., & Guest, R. (2020). The role of authentic assessment to preserve academic integrity and promote skill development and employability. *Studies in Higher Education*, 45(11), 2132-2148. <https://doi.org/10.1080/03075079.2019.1582015>
- Sperl, L., Breier, C. M., Griebßbach, E., & Schweinberger, S. R. (2024). Do typing skills matter? Investigating university students' typing speed and performance in online exams. *Higher Education Research & Development*, 43(4), 981-995. <https://doi.org/10.1080/07294360.2023.2287724>
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72-101.
- Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, 30(4), 325-341. <https://doi.org/10.1080/02602930500099102>
- Summers, J., & Catarro, F. (2003). Assessment of handwriting speed and factors influencing written output of university students in examinations. *Australian Occupational Therapy Journal*, 50(3), 148-157. <https://doi.org/10.1046/j.1440-1630.2003.00310.x>
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. <https://doi.org/10.1007/978-1-4419-8126-4>
- Teaching and Learning Center. (2025). *TiU didactic framework*. Tilburg University.
- Teng, S.-Y. (1943). Chinese influence on the Western examination system: I. Introduction. *Harvard Journal of Asiatic Studies*, 7(4), 267-312.

- Teodorczuk, A., Fraser, J., & Rogers, G. D. (2018). Open book exams: A potential solution to the “full curriculum”? *Medical Teacher*, 40(5), 529-530. <https://doi.org/10.1080/0142159X.2017.1412412>
- Theobald, A. S. (2021). Oral exams: A more meaningful assessment of students' understanding. *Journal of Statistics and Data Science Education*, 29(2), 156-159. <https://doi.org/10.1080/26939169.2021.1914527>
- Thomas, J. (2016). Using ‘critical incident analysis’ to promote critical reflection and holistic assessment. In N. Gould (Ed.), *Social work, critical reflection and the learning organization* (pp. 101-116). Routledge.
- Thompson, N., & Pascal, J. (2012). Developing critically reflective practice. *Reflective Practice*, 13(2), 311-325. <https://doi.org/10.1080/14623943.2012.657795>
- Tierney, R. D. (2014). Fairness as a multifaceted quality in classroom assessment. *Studies in Educational Evaluation*, 43, 55-69. <https://doi.org/10.1016/j.stueduc.2013.12.003>
- Tilburg School of Economics and Management. (2024). *Assessment policy*. Tilburg University.
- Tilburg University. (2017a). *Knowledge, skills, character: Tilburg University educational vision statement*. <https://www.tilburguniversity.edu/sites/default/files/download/Educational%20Vision%202017.pdf>
- Tilburg University. (2017b). *Tilburg Educational Profile*. <https://www.tilburguniversity.edu/about/education/tep>
- Todd, M., Bannister, P., & Clegg, S. (2004). Independent inquiry and the undergraduate dissertation: Perceptions and experiences of final-year social science students. *Assessment & Evaluation in Higher Education*, 29(3), 335-355. <https://doi.org/10.1080/0260293042000188285>
- Tomas, C., Whitt, E., Lavelle-Hill, R., & Severn, K. (2019). Modeling holistic marks with analytic rubrics. *Frontiers in Education*, 4, 89. <https://doi.org/10.3389/feduc.2019.00089>
- Trafford, V. (2003). Questions in doctoral vivas: Views from the inside. *Quality Assurance in Education*, 11(2), 114-122. <https://doi.org/10.1108/09684880310471542>
- Traub, R. E., & Rowley, G. L. (1991). Understanding reliability. *Educational Measurement: Issues and Practice*, 10(1), 37-45. <https://doi.org/10.1111/j.1745-3992.1991.tb00183.x>
- TU/Delft. (n.d.). *How to write open-ended exam questions?* <https://www.tudelft.nl/teaching-support/educational-advice/assess/written-exams/open-questions>
- University of Edinburgh. (2024). *Definition of reflection*. <https://reflection.ed.ac.uk/>
- University of Twente. (2026). *The bachelor's thesis*. University of Twente. https://www.utwente.nl/en/bee/Bachelor_Curriculum/program_3_year/thesis/#follow-up-getting-your-diploma
- University of Waterloo. (n.d.). *Responding to writing assignments: Managing the paper load*. <https://uwaterloo.ca/centre-for-teaching-excellence/catalogs/tip-sheets/responding-writing-assignments-managing-paper-load>
- Usherwood, S. (2018). Assessment strategies in simulation games. In P. Bursens, V. Donche, D. Gijbels, & P. Spooren (Eds.), *Simulations of decision-making as active learning tools: Design and effects of political science simulations* (pp. 109-118). Springer International Publishing. https://doi.org/10.1007/978-3-319-74147-5_9
- Van Berkel, H., Bax, A., Joosten-ten Brinke, D., Beekman, K., & Van Schilt-Mol, T. (2023). *Toetsen in het hoger onderwijs [Assessment in higher education]* (5th ed.). Bohn Stafleu van Loghum.
- Van Blerkom, M. L. (2017). *Measurement and statistics for teachers*. Taylor & Francis.
- van de Veen, E., Pol, M., & Moerkerke, G. (2018). De toetscommissie: Borgen of zorgen? De rol en positie van de toetscommissie in het bekostigd hoger onderwijs [The examination committee: Assurance or concern? The role and position of the examination committee in funded higher education]. *Hoger Onderwijs Magazine*, 39(2), 19-21.
- Van Wyk, M. L., & Carl, A. E. (2010). The portfolio as an authentic assessment tool for learning: Is it serving its purpose? *Yesterday and Today*, (5), 141-161.
- VanderWeele, T. J. (2022). The importance, opportunities, and challenges of empirically assessing character for the promotion of flourishing. *Journal of Education*, 202(2), 170-180. <https://doi.org/10.1177/0022057421102690>
- Venkatesh, V., Bures, E., Davidson, A.-L., Wade, C. A., Lysenko, L., & Abrami, P. C. (2013). Electronic portfolio encouraging active and reflective learning: A case study in improving academic self-regulation through innovative

- use of educational technologies. In A. D. Ritzhaupt & S. Kumar (Eds.), *Cases on educational technology implementation for facilitating learning* (pp. 341-376). IGI Global Scientific Publishing.
<https://doi.org/10.4018/978-1-4666-3676-7.ch019>
- Verhavert, S., Bouwer, R., Donche, V., & De Maeyer, S. (2019). A meta-analysis on the reliability of comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 26(5), 541-562.
<https://doi.org/10.1080/0969594X.2019.1602027>
- WAC Clearinghouse. (2026a). *An introduction to writing across the curriculum*.
<https://wacclearinghouse.org/repository/teaching/intro/>
- WAC Clearinghouse. (2026b). *What is writing to learn?* <https://wacclearinghouse.org/repository/teaching/intro/wtl/>
- WAC Clearinghouse. (2026c). *Writing-to-learn activities*.
https://wacclearinghouse.org/repository/teaching/intro/wtl/wtlactivities/#wtl_p
- Wagner-Menghin, M., Preusche, I., & Schmidts, M. (2013). The effects of reusing written test items: A study using the Rasch model. *International Scholarly Research Notices*, 2013(1), 585420. <https://doi.org/10.1155/2013/585420>
- Wagner, M., Porto, M., & Tsvetkova, N. (2022). The final oral examination. In M. Byram & M. Stoicheva (Eds.), *The experience of examining the PhD* (pp. 45-54). Routledge. <https://doi.org/10.4324/9781003248569-5>
- Wainer, H. (1987). The first four millennia of mental testing: From Ancient China to the computer age. *ETS Research Report Series*, 1987(2), i-6.
- Wainer, H. (2000). Introduction and history. In H. Wainer, N. J. Dorans, R. Flaugher, B. F. Green, & R. J. Mislevy (Eds.), *Computerized adaptive testing: A primer* (pp. 1-21). Routledge.
- Wald, H. S., Borkan, J. M., Taylor, J. S., Anthony, D., & Reis, S. P. (2012). Fostering and evaluating reflective capacity in medical education: Developing the REFLECT rubric for assessing reflective writing. *Academic Medicine*, 87(1), 41-50. <https://doi.org/10.1097/ACM.0b013e31823b55fa>
- Walker, C., & Gleaves, A. (2016). Constructing the caring higher education teacher: A theoretical framework. *Teaching and Teacher Education*, 54, 65-76. <https://doi.org/10.1016/j.tate.2015.11.013>
- Walsh, A. (2007). An exploration of Biggs' constructive alignment in the context of work-based learning. *Assessment & Evaluation in Higher Education*, 32(1), 79-87. <https://doi.org/10.1080/02602930600848309>
- Walvoord, B. E., & Anderson, V. J. (2011). *Effective grading: A tool for learning and assessment in college*. John Wiley & Sons.
- Watts, J. H. (2012). Preparing doctoral candidates for the viva: Issues for students and supervisors. *Journal of Further and Higher Education*, 36(3), 371-381. <https://doi.org/10.1080/0309877X.2011.632819>
- Wefald, A. J., Benavides, M. T., & Seay, A. K. M. (2022). Learning from leadership capstone projects. *Journal of Leadership Education*, 21(3), 1-12. <https://doi.org/10.12806/V21/I3/R3>
- Wellington, J. (2010a). *Making supervision work for you: A student's guide*. SAGE Publications.
- Wellington, J. (2010b). Supporting students' preparation for the viva: Their pre-conceptions and implications for practice. *Teaching in Higher Education*, 15(1), 71-84. <https://doi.org/10.1080/13562510903487867>
- Wendt, A., & Kenny, L. E. (2009). Alternate item types: Continuing the quest for authentic testing. *Journal of Nursing Education*, 48(3), 150-156. <https://doi.org/10.3928/01484834-20090301-11>
- Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, 22(5), 537-548.
<https://doi.org/10.1080/0141192960220502>
- Williams, J. (2016). Quality assurance and quality enhancement: Is there a relationship? *Quality in Higher Education*, 22(2), 97-102. <https://doi.org/10.1080/13538322.2016.1227207>
- Wilson, O., Hargarter, A., & Van Rensburg, B. (2024). Using simulation-based assessment for event management students. *Journal of Ethics in Higher Education*, (4), 91-122. <https://doi.org/10.26034/fr.jehe.2024.5971>
- Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, 52(1), 17-37.
<https://doi.org/10.1080/00461520.2016.1207538>

- Wisker, G. (2012). *The good supervisor: Supervising postgraduate and undergraduate research for doctoral theses and dissertations*. Bloomsbury Academic.
- Wolf, K. (2015). Leniency and halo bias in industry-based assessments of student competencies: a critical, sector-based analysis. *Higher Education Research & Development*, 34(5), 1045-1059. <https://doi.org/10.1080/07294360.2015.1011096>
- Wolff, C. E., Pogg, C., & Hsiao, Y. P. (2024). *More thinking less writing: Exploring student emotions using a reflective chatbot to support critical reflection*. The 2024 2nd International Conference on Affective Education, Taipei, Taiwan.
- Wyatt-Smith, C., Adie, L., & Harris, L. (2024). Supporting teacher judgement and decision-making: Using focused analysis to help teachers see students, learning, and quality in assessment data. *British Educational Research Journal*, 50(3), 1420-1448. <https://doi.org/10.1002/berj.3984>
- Yan, Z., & Boud, D. (2022). Conceptualising assessment-as-learning. In Z. Yan & L. Yang (Eds.), *Assessment as learning: Maximising opportunities for student learning and achievement* (pp. 11-24). Routledge.

This practical guide provides unified and evidence-informed advice for designing high-quality course assessment. Designed for all stakeholders involved in assessment, it offers accessible explanations of theoretical rationales, attention to contextual considerations, and rich, workable examples of both course-level assessment design and commonly used assessment types.

- Section 1 introduces the foundational framework, key constructs, and criteria for effective assessment.
- Section 2 presents the theoretical background of commonly used assessment types.
- Section 3 focuses on application through real course assessment designs. Each chapter includes learning goals, guidance on transparency, reliability, and validity, and concludes with real-world examples, including detailed design cases and examples.

By consolidating existing school-level handbooks, this resource refers to general assessment policies in Dutch higher education and helps lecturers and examiners apply evidence-informed principles in practice. It strengthens professional judgment, enhances student-centered learning, and addresses emerging challenges, including workload pressures and the integration of generative artificial intelligence into assessment.