A PRACTICAL GUIDE TO QUALITY ASSESSMENT IN DUTCH HIGHER EDUCATION

A Practical Guide to Quality Assessment in Dutch Higher Education

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

OPEN PRESS TiU

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 Liebregts, Wynand Bodewes

Tilburg School of Humanities and Digital Sciences (TSHD)
Constanze Hopfner, Frea Dooremaal, Jos Swanenberg, Hanne Jacobs, Marijn van Wingerden, Minne Bakker, Pytrik Schafraad, Sinsemilla Blokvoort, Yvette Drissen

Tilburg School of Social and Behavioral Sciences (TSB)
Antonio Massella, Brenda Den Oudsten, Djurre Holtrop, Kerry Forrester, Leanne Prevc-Hekman, Lonneke Schellekens, Paula Mommersteeg, Roel Rutten

Tilburg Law School (TLS)
Bart van de Moosdijk

Teaching & Learning Center
Ya-Ping (Amy) Hsiao[1], Vincent Kalis

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

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 0).

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 0.
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	• 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 Berkel[2] and Bas Hemker[3], 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 Sluijter

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 ≈ 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 **rubric**s, 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 *On successful completion of this course, students will be able to:*	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 *On successful completion of this course, students will be able to:*	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 *On successful completion of this course, students will be able to:*	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) (Gehringer, 2004; Joncas et al., 2018).

When reusing exam questions in general, keep the following suggestions in mind (Gehringer, 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.

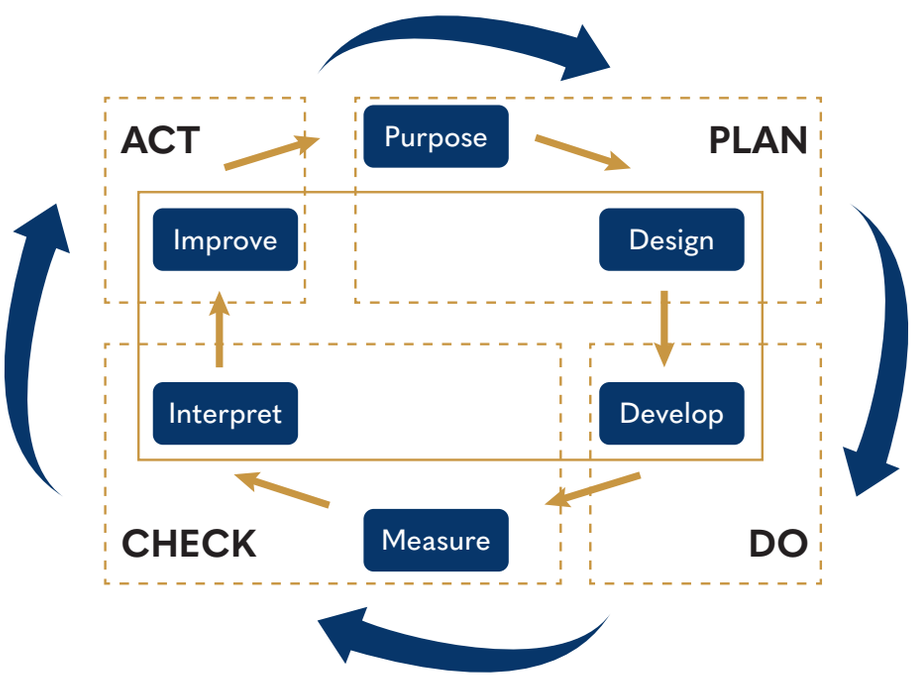


Figure 1.2.1
The Assessment Cycle Based on the PDCA

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, **rubric**s, 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/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** mandatory[5] 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 exam[6].

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.