



Universiteit
Leiden
The Netherlands

Intermediate assessment in higher education

Day, I.N.Z.

Citation

Day, I. N. Z. (2018, June 28). *Intermediate assessment in higher education*. Retrieved from <https://hdl.handle.net/1887/63156>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/63156>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The following handle holds various files of this Leiden University dissertation:
<http://hdl.handle.net/1887/63156>

Author: Day, I.N.Z.

Title: Intermediate assessment in higher education

Issue Date: 2018-06-28

Intermediate Assessment in Higher Education

Characteristics, perceptions, and effects

Indira Day

Intermediate Assessment in Higher Education
Characteristics, perceptions, and effects

Indira Day



Universiteit
Leiden
ICLON

ICLON, Leiden University Graduate School of Teaching

ico

The research was carried out in the context of the *Dutch Interuniversity Center for Educational Sciences*.

Title: Intermediate Assessment in Higher Education: Characteristics, perceptions, and effects.

Titel: Tussentijds Toetsen in het Hoger Onderwijs: Kenmerken, percepties, en effecten.

ICLON PhD Dissertation Series

Print: Mostert & van Onderen! Leiden

Cover design: Shanti Day

Lay-out: Mostert & van Onderen!

ISBN/EAN: 978-94-90383-22-0

© 2018, Indira Day

All rights reserved. No part of this thesis may be reproduced, stored in retrieval systems, or transmitted in any form by any means, electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the author.

Intermediate Assessment in Higher Education
Characteristics, perceptions, and effects

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 28 juni
klokke 13:45 uur

door

Indira Nadia Zinzi Day

geboren te Leiden
in 1988

Promotores

Prof. dr. W.F. Admiraal

Prof. dr. P.M. van Westenberg

Copromotor

Dr. F.M. van Blankenstein

Promotiecommissie

Prof. dr. S.F. Akkerman

Prof. dr. ir. F.J.J.M. Janssen

Prof. dr. P.W. van den Broek

Prof. dr. P.J. den Brok, Wageningen University

Dr. N. Saab

Table of Contents

Chapter 1. General introduction	11
1.1 Introduction	12
1.2 Intermediate assessment in higher education	13
1.2.1 Mechanisms that underlie the effects of intermediate assessment	13
1.3 Implementing intermediate assessment	14
1.3.1 Implementation at Leiden University	15
1.4 Central aim of this dissertation	16
1.5 Overview of this dissertation	17
 Chapter 2. A review of the characteristics of intermediate assessment and their relationship with student grades	19
2.1 Introduction	21
2.2 Methods	22
2.2.1 Search and Selection	22
2.2.2 Reliability of Selection	25
2.2.3 Analysis	25
2.3 Results & Discussion	29
2.3.1 Article Background Information	29
2.3.2 Assessment Type	29
2.3.3 Assessment Type and Discipline	30
2.3.4 Assessment Type and Frequency of Assessment	30
2.3.5 Assessment Type and Feedback	30
2.3.6 Assessment Type and Mandatory Assessments	32
2.3.7 Assessment Type and Assessor	33
2.3.8 Assessment Type and Reward	35
2.4 Intermediate Assessment Scenarios	36
2.4.1 Scenario 1: Quiz	36
2.4.2 Scenario 2: Writing assignment	38
2.4.3 Scenario 3: Authentic assessment	39
2.5 Concluding remarks	40

Chapter 3. Teacher and Student Perceptions of Intermediate Assessment in Higher Education	43
3.1 Introduction	45
3.1.1 Intermediate Assessment in Higher Education	46
3.1.2 Perceptions of Intermediate Assessment	48
3.1.3 Current study	48
3.2 Method	49
3.2.1 Context	49
3.2.2 Participants	50
3.2.3 Data	50
3.2.4 Analysis	52
3.3 Results	53
3.3.1 Use of Intermediate Assessment	53
3.3.2 Teacher Perceptions	57
3.3.3 Student Perceptions	60
3.4 Discussion and Conclusion	64
3.4.1 Limitations	66
3.4.2 Implications for practice	67
3.4.3 Concluding remarks	67

Chapter 4. University teachers' conceptions of their current and ideal intermediate assessment: An A+ is good, but speaking your mind is better	69
4.1 Introduction	71
4.1.1 The current study	73
4.2 Method	74
4.2.1 Recruitment	74
4.2.2 Participants	74
4.2.3 Interviews	75
4.2.4 Analysis	75
4.3 Results	78
4.3.1 Teachers who display no difference between the current and ideal assessment	79
4.3.2 Teachers who show differences between current and ideal assessment	85
4.3.3 Reasons for discrepancy between current and ideal assessment	91

4.4 Discussion	91
4.4.1 Directions for future research	93
4.4.2 Concluding remarks	94

Chapter 5. Explaining Individual Student Success Using Intermediate Assessment Types and Student Characteristics 95

5.1 Introduction	97
5.1.1 The University Environment and Student Success	97
5.1.2 Student Characteristics and Student Success	99
5.1.3 Intermediate Assessment and Student Characteristics	100
5.2 Methods	101
5.2.1 Context	101
5.2.2 Participants	103
5.2.3 Materials	103
5.2.4 Procedure	104
5.2.5 Ethics	104
5.3 Results	105
5.3.1 Descriptive Statistics	105
5.3.2 Preliminary Regression Analyses	106
5.3.3 Student Characteristics, Assessment Type, and Student Achievement	106
5.4 Discussion	109
5.4.1 Limitations and Directions for Future Research	112
5.4.2 Concluding Remarks	113

Chapter 6. General discussion 115

6.1 Introduction	116
6.2 Main findings per chapter	116
Chapter 2. A Review of the Characteristics of Intermediate Assessment and their Relationship with Student Grades	116
Chapter 3. Teacher and Student Perceptions of Intermediate Assessment in Higher Education	117
Chapter 4. University Teachers' Conceptions of their Current and Ideal Intermediate Assessment	118
Chapter 5. Explaining Individual Student Success Using Intermediate Assessment Types and Student Characteristics	118

6.3 General findings	119
6.4 Reflections on the design of intermediate assessment	120
6.4.1 Assessment Type	120
6.4.2 Use of Feedback	121
6.5 Limitations	122
6.6 Implications for Practice	123
6.6.1 Encourage use of intermediate assessment	123
6.6.2 Minimise institutional assessment guidelines	123
6.6.3 Support teachers' assessment practice	124
6.7 Directions for future research	125
6.8 Concluding Remarks	126
References	127
Summary	143
Introduction	144
Chapter 2	145
Chapter 3	146
Chapter 4	148
Chapter 5	149
Discussion	151
Nederlandstalige Samenvatting	153
Hoofdstuk 1: Introductie	154
Hoofdstuk 2: Een overzicht van de literatuur over tussentijds toetsen in het Hoger Onderwijs	155
Hoofdstuk 3: Percepties over tussentijds toetsen van docenten en studenten	157
Hoofdstuk 4: Opvattingen van docenten over hun huidige en ideale tussentijdse toets	158
Hoofdstuk 5: De invloed van tussentijds toetsen en studentkenmerken op studiesucces	160
Hoofdstuk 6: Discussie	162

List of publications and presentations	165
Scientific Publications	166
Professional Publications	166
Paper Presentations	166
Poster Presentations and Roundtables	167
Curriculum Vitae	169
Dankwoord	171
ICLON PhD dissertation series	173

Chapter **1**

General introduction

1.1 Introduction

In education, teachers use assessment to measure student's learning outcomes. However, familiar questions like 'will this be on the exam?' also exemplify that assessment actually drives student learning. Therefore, assessment should be an integral part of education. Research has indicated that assessment can be used to improve learning (assessment *for* learning) in addition to simply measuring learning (assessment *of* learning). Assessment for learning is a method of formative assessment (van der Kleij, Vermeulen, Schildkamp & Eggen, 2015; Wiliam, 2011). Formative assessment has been widely researched and results from several studies and reviews indicate that it is a potent measure to improve learning, better than summative assessment (assessment of learning) does (Black & Wiliam 1998; Kluger & DeNisi 1996).

In educational practice, there is often no explicit division between formative and summative assessment. Summative assessments mainly measure what a student knows, but teachers can provide feedback to students based on assessment results. Conversely, formative assessments are primarily used to provide feedback, but also show students' knowledge level. Therefore, Taras (2009) argues that education should not focus on whether an assessment has a formative or a summative function, but on the formative and summative processes of assessment.

In higher education, students are often continuously assessed throughout the semester, a process referred to as intermediate assessment. In this dissertation, intermediate assessment is defined as all forms of assessment that students face during the semester. Some intermediate assessments may count towards students' final course grade, others may be conditional, or even voluntary. Intermediate assessment may also be referred to as frequent, continuous or in-course assessment. Intermediate assessment lends itself for quite a few formative processes, because it takes place during the course period. Providing feedback to students and amending teacher instruction, for example, may be more pertinent reactions to intermediate than final to assessment.

Previous research has indicated that using intermediate assessment could improve student performance (e.g., Larsen, Butler, & Roediger, 2008; Palmer, 1974). At Leiden University, intermediate assessment resurfaced after signing performance agreements with the government in 2012 (Zijlstra, 2012). In these agreements, universities in the Netherlands pledged to improve student success,

mainly by increasing the number of students that graduate from a three-year bachelor program within four years. Intermediate assessment was introduced in several departments as a measure to ensure this higher number of graduates. Of course, several university teachers had already been using intermediate assessment before the performance agreements.

1.2 Intermediate assessment in higher education

In general, intermediate assessment is claimed to influence three types of student outcomes. First, the already mentioned cognitive outcomes, often expressed as student grades. Second, there are affective outcomes, like student engagement with the course. Third are the behavioural outcomes, where students will increase their time spent studying for courses that use intermediate assessment. These three outcomes are often intertwined. For example, students who are more engaged with their course will probably also spend more time studying, which in turn may lead to higher grades.

There are several mechanisms underlying the positive outcomes of intermediate assessment. These will be discussed in the following section.

1.2.1 Mechanisms that underlie the effects of intermediate assessment

Assessing students regularly during the course has several benefits. The first of these benefits is that repeated testing of information improves retention of this information more strongly than repeated studying. This is referred to as the testing effect, which is extensively reviewed by Roediger and Karpicke (2006). Studying by testing is often referred to as retrieval practice, since the increased memory retention by repeated testing is caused by the effort expended retrieving the information from memory. Evidence for the testing effect has been found in laboratory studies, as well as in authentic classroom studies in secondary and tertiary education (Dirkx, Kester, & Kirschner, 2014; McDaniel, Anderson, Derbish, & Morrisette, 2007; Roediger & Karpicke, 2006). Many studies evaluating the testing effect use the same questions in the practice test as in the final test, but Butler (2010) has focused on transfer of knowledge to new final test questions and shows that the testing effect still works when new questions are asked.

A second benefit of intermediate assessment is that when assessments are dispersed throughout the semester, students will spread their study time accordingly. Distributed practice, or spacing, is a more effective mechanism for retention than massed studying (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Additionally, distributed practice interacts with retrieval practice. According to Karpicke and Roediger (2007) delaying the initial moment of retrieval increases the retrieval effort which subsequently enhances long-term retention.

Furthermore, distributed studying plays a role in the relationship between time spent studying and student grades. Several authors indicate that increased time spent studying, so called time-on-task, will lead to better study results (e.g. Admiraal, Wubbels & Pilot, 1999; Doumen, Broeckmans & Masui, 2014; Wang, Haertel, & Walberg, 1990). However, Michaels and Miethe (1989) found that this relationship is only true for students who spread that increased study time across the course period. Students who spent more time cramming did not score better than their peers who spent less time.

Additionally, introducing intermediate assessment may allow for more time for reflection for students. For example, because students get intermediate results and feedback that they can use to gauge their own understanding and learning. Reflection is an important aspect of learning, and Moon (1999) poses that when we reflect we get access to deeper knowledge.

1.3 Implementing intermediate assessment

The studies cited in the previous sections show that there are various ways to implement intermediate assessment in higher education. For example, McDaniel et al. (2007) used multiple choice or short answer quizzes, Dirkx et al. (2014) used factual and application short answer questions, and Admiraal et al. (1999) used homework assignments. All facets of intermediate assessment that can vary in implementation will be referred to as intermediate assessment characteristics in the current dissertation. An overview of possible assessment characteristics can be found in Table 1.1. Some assessment characteristics may be more influential for student performance than others, but there is a dearth of studies investigating this.

Table 1.1. Possible assessment characteristics

Assessment characteristic	Examples of variations
Type of assessment	Quiz, presentation, essay, cases
Frequency of assessment	Weekly, single assessment, 8 assessments
Duration	30 minutes, 5 minutes, 2 hours
Number and type of questions	Multiple choice, short answer
Scoring	Percentage of grade, performance score, bonus points
Assessor	Teacher, peer, self
Feedback	Corrective, elaborate, feedback form
Medium	Computer, pen and paper
Location	In class, at home, anywhere
Mandatory	Yes, no
Other characteristics	Open book, reattempts

1.3.1 Implementation at Leiden University

The empirical studies for this dissertation were conducted at Leiden University. Leiden University is the oldest research university in the Netherlands and prides itself on using evidence-based educational practices. For several course programs at Leiden University, intermediate assessment became an obligation in the 2013 – 2014 academic year, starting in the first-year curriculum and gradually expanding across educational years in the following academic years. This dissertation focuses on two first-year curricula of the Leiden Law School (Law and Criminology) and two at the department of Social Science (Psychology and the International Bachelor of Psychology, IBP).

Several changes to the curriculum were introduced in addition to intermediate assessment, at both departments. These changes included intensifying of class time, internationalisation of the curriculum, and reshuffling the course order in the curriculum. All decisions on the adaptation of the curriculum were made by institutional or department boards, but teachers were given sufficient freedom to redesign their own courses and were only presented with minimal constraints. At the two departments the guidelines were very different.

Leiden Law School took a ‘minimalist’ approach to obligating intermediate assessment. Only teachers teaching full semester courses, worth 10 European Credits, were obligated to offer a partial exam, often in the form of a midterm. For the two programs under investigation, this amounted to three courses with obligatory intermediate assessment. All other teachers did not need to offer intermediate assessment but were given the freedom to introduce intermediate

assessments into their courses if they wanted to. In the Law curriculum, all final exams were a combination of multiple-choice questions and an open-ended case question. Criminology final exams more often consisted of only open-ended questions.

The psychology programs (that have the exact same curriculum, but the IBP is geared towards international students and taught in English) needed to offer what the department calls *plural assessment*, which focuses on offering more than one assessment type. Under the plural assessment regime, the multiple-choice final exam of a course was not allowed to make up more than 70% of the final grade. Teachers could decide for themselves what type of assessment they wanted to offer to students for the remaining 30% of the grade, and the majority of teachers chose some sort of intermediate assessment.

1.4 Central aim of this dissertation

Several studies have demonstrated the effectiveness of intermediate assessment in improving student performance. However, there is still uncertainty about whether different characteristics of intermediate assessment are more optimal for student improvement than others. In this dissertation, I try to answer the following overarching research question:

‘In what ways can intermediate assessment in higher education be designed to improve study behaviour and study results?’

When answering this research question, I explore the influence of intermediate assessment characteristics with the following three focal points in mind. The first is looking at how intermediate assessment is used in higher education, with a specific focus on differing characteristics (implementation of intermediate assessment). The second focal point investigates the ideas of higher education teachers and students about intermediate assessment (perceptions of intermediate assessment). The third and final focal point looks how intermediate assessments with different characteristics influence student achievement (effects of intermediate assessment). These three focal points each present a facet of the ways in which intermediate assessment can be designed.

1.5 Overview of this dissertation

To answer the overarching research question, in Chapters 2 to 5, several facets of intermediate assessment in higher education will be investigated. At least one of the three focal points discussed in the central aim of this dissertation is central to each chapter.

Chapter 2 is a literature review. The research question for this review is ‘what characteristics of intermediate assessment are related to student grades?’ emphasising the implementation of different intermediate assessment characteristics and effects in the form of student grades. The literature review synthesises outcomes from 88 articles and focuses on four specific assessment characteristics, in addition to providing an overview of different assessment types employed in higher education. For each characteristic, i.e., the use of feedback, whether an assessment is mandatory, who is the assessor, and what is the reward for the assessment, a general overview and exemplar articles are discussed. This chapter ends with three scenarios for intermediate assessment to further exemplify how these assessment characteristics could be applied in higher education.

After this exploration of the ways in which intermediate assessment is employed in higher education across countries, disciplines, and educational years, this dissertation focuses on the local context of Leiden University. Chapters 3 and 4 explore intermediate assessment in first-year psychology, law and criminology courses, whereas Chapter 5 specifically focuses on first-year law and criminology courses. For these three chapters, the definition of intermediate assessment is slightly amended, only focussing on assessments that are mandatory and for which completion is checked by the teacher.

Chapters 3 and 4 are both interview studies, where perceptions and implementation of intermediate assessment are the focal point. In Chapter 3, eighteen teachers were interviewed on their perceptions of intermediate assessment following the implementation of an educational reform obligating teachers to use intermediate assessment in their courses. Furthermore, students’ perceptions were gauged using the outcomes on student evaluations of teaching, and seven students were interviewed to get a more detailed account. The two research questions of this chapter are ‘what types of intermediate assessment are used in the programmes under investigation?’ and ‘how are these types of intermediate assessment perceived by teachers and students?’

Chapter 4 is a deeper investigation into teachers' ideas about intermediate assessment. In this chapter, thirteen teachers are asked to reflect on the ideas behind their current intermediate assessment, and on how they would shape intermediate assessment in their ideal situation. The research question is 'What differences in conceptions of intermediate assessment do university teachers display when discussing their current and ideal intermediate assessment?'

Chapter 5, like Chapter 2, focuses on assessment characteristics and effects in the form of student grades. In this chapter, 94 first-year law and criminology students completed a questionnaire measuring their motivation, self-regulation and several demographic variables. The following two research questions are answered 'to what extent does the type of intermediate assessment relate to academic achievement?' and 'what role do gender, high school achievement, motivation and self-regulation play in this relationship?'

Finally, in Chapter 6 the results of the four previous chapters will be summarised and synthesised. This chapter will also reflect on the central aim of this dissertation and put forth practical implications of and directions for future research into intermediate assessment in higher education.

Chapter 2

A review of the characteristics of intermediate assessment and their relationship with student grades

In-course assessment, such as midterms, quizzes or presentations, is often an integral part of higher education courses. These so-called intermediate assessments, influence students' final grades. The current review investigates which characteristics of intermediate assessment relate to these grades. In total, 88 articles were reviewed that examined the relationship between intermediate assessment and student grades. Four main characteristics were identified: the use of feedback, whether the assessment is mandatory, who is the assessor, and what reward do students get for participating. Results indicate that corrective feedback leads to the most positive results, but elaborate feedback may benefit lower achieving groups. Moreover, no difference in results was found for mandatory versus voluntary intermediate assessments. Additionally, peer assessment seemed to be beneficial. Finally, rewarding students with course credit improves students' grades more than other rewards. Three scenarios are presented on how teachers can combine the different characteristics to optimise their intermediate assessment.

This chapter has been published in adapted form as: Day, I. N. Z., van Blankenstein, F. M., Westenberg, M., & Admiraal, W. (2017). A review of the characteristics of intermediate assessment and their relationship with student grades. *Assessment & Evaluation in Higher Education*, Advance Online Publication. doi:10.1080/02602938.2017.1417974.

2.1 Introduction

Assessment is a powerful force in education. Dunlosky, Rawson, Marsh, Nathan, and Willingham (2013) posit that practice testing is an effective studying technique, next to distributed practice (the process of studying constantly throughout a study period instead of cramming). Since these two processes enhance student learning in self-study, it stands to reason that distributed assessments should be an integral part of a course, encompassing different assessment moments during the course instead of just a final examination. In higher education, these assessments during the course often take various forms.

In the current paper, all forms of assessment during the course will be referred to as ‘intermediate assessment’ (synonyms are ‘frequent’ or ‘continuous’ assessment, see e.g., Isaksson 2008; Rezaei 2015). Intermediate assessment may influence three different types of student outcomes, namely affective, cognitive, or behavioural outcomes. Some examples of these outcomes are student perceptions of their learning (e.g. Bälter, Enström, & Klingenberg 2013), student engagement (e.g. Holmes 2014), student studying behaviour (e.g. Admiraal, Wubbels, & Pilot 1999), and student grades (e.g. De Paola & Scoppa 2011).

Higher education teachers employ intermediate assessment with several goals in mind. Results from the interview study described in Chapter 3, indicate that teachers mainly employ intermediate assessment to motivate students to start studying, and to be able to measure a variety of learning goals, knowledge and skills. However, they often expressed uncertainty about the influences of different assessment characteristics on their goals. Therefore, in the current review we further investigate different assessment characteristics.

Intermediate assessment in higher education is a subject that is widely researched (e.g. Admiraal et al. 1999; De Paola & Scoppa 2011; Bälter et al 2013; Holmes 2015). Assessments can vary on a multitude of characteristics, some of which are more thoroughly researched than others. One characteristic is the type of the assessment. Different results may be gained when using quizzes, essays or presentations. Often the type of an assessment is decided by the learning goals a teacher wants to assess, so called constructive alignment (Biggs, 1996).

Other characteristics include the use of feedback, or the use of formative assessments (see Bennett 2011, for a review on formative assessment). Studies show that formative assessments are potent learning tools (Black & Wiliam 2004), and that instructive feedback enhances learning. Other examples of

characteristics are the frequency of the assessment, who is the assessor and whether students are graded on the assessment or not. The latter might be important to distinguish as students appear to be focused on grades (Harland, McLean, Wass, Miller, and Sim 2015).

There is a lot of evidence from studies that intermediate assessment has a general positive effect on students' final grades. However, these studies (e.g. Admiraal et al. 1999; Leeming 2002; De Paola & Scoppa 2011) often focus on one specific type of assessment, not contrasting them with other types.

The reality of higher education often prevents teachers from trying out different types of intermediate assessment within the same instalment of their course. The goal of the current paper is to provide an overview of research that has been conducted into intermediate assessment in higher education, across disciplines, and its relation to student grades. We have decided on grades as a student outcome because in educational research, grades are often used as a proxy for cognitive learning (see, for example, Richardson, Abraham, & Bond 2012). Assessment drives learning (Clouder, Broughan, Jewell, and Steventon 2012, p. 2), making grades a suitable outcome variable.

With this review, the authors aim to provide a more general overview of what intermediate assessments have led to positive outcomes and to inform higher education professionals who are designing their curriculum. We try to answer the following research question: What characteristics of intermediate assessment are related to student grades?

2.2 Methods

2.2.1 Search and Selection

Papers for the review were searched using the 'All databases' search in Web of Science. In each of the searches one of the following keywords was combined with 'assessment': 'intermediate', 'frequent', 'continuous', 'programmatic', 'in-between', 'formative' and 'summative'.

The different steps of the search process, including which parts of the articles were examined, the inclusion criteria and resulting selected articles are presented in Figure 2.1. Steps 1 and 2 (search and first selection) were conducted in March 2016, which was also the cut-off date for published articles. The year 2000 was chosen as a start date. The number of published articles sharply decreased the

closer the studies were published to 2000, confirming our decision. In each subsequent step of the search process, the inclusion criteria were expanded to select more focused studies.

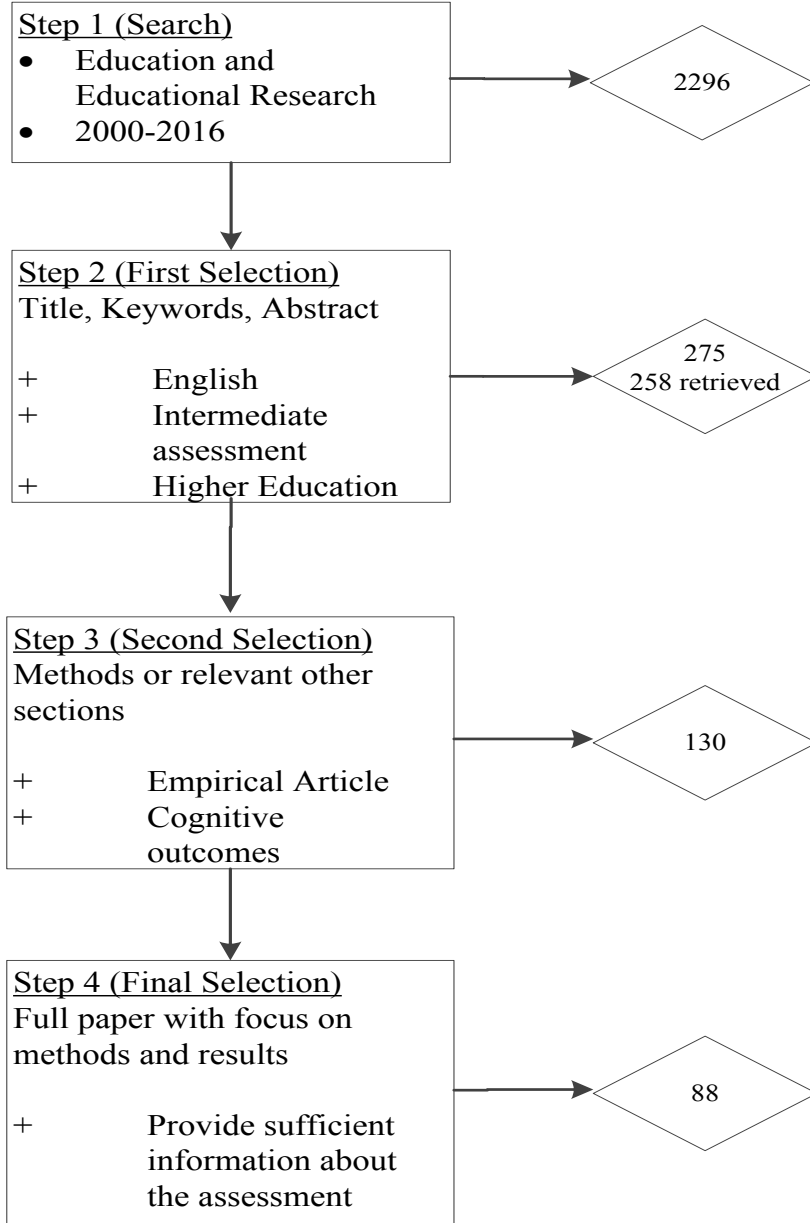


Figure 2.1. Schematic Overview of the Search and Selection Process.

Ultimately, 88 articles were selected for inclusion in the review. These articles included (quasi) experiments as well as descriptive studies. While reading the methods and results sections all relevant information concerning the intermediate assessment and its results was extracted from the articles. All information extracted from the articles fits in the categories presented in Table 2.1. Context variables and assessment characteristics were defined while reading the first few articles and discussed amongst authors. Furthermore, at the annual meeting of the Netherlands Educational Research Association, educational researchers and educationalists attending a discussion session (Day et al. 2016) were asked to mark which variables they deemed interesting and useful.

Table 2.1. Overview of Information Extracted from the Articles Included in the Review

Context variables	Assessment Characteristics
Country	Type of assessment e.g. quiz, presentation, essay, cases
Discipline	Frequency of assessment e.g. weekly, single assessment, 8 assessments
N of students	Duration
Study Phase e.g. first year, undergraduate	Number and type of questions e.g. multiple choice, short answer
Use of control group	Scoring e.g. percentage of grade, performance score, bonus points
Cognitive Outcome Variable e.g. final examination score, GPA	Feedback e.g. elaborate, feedback form
Other variables in the research e.g. gender, comparing low and high achieving students	Assessor teacher, peers, self
Findings of the research	Medium e.g. computer, pen and paper
	Location in class, at home, anywhere
	Mandatory
	Other characteristics e.g. open book, reattempts

All relevant information from the articles was added to a Microsoft Access database. When an article does not provide information on one of the variables, ‘no info’ is entered. After entering all information several iterations were performed to make sure similar terms were used for similar concepts across articles.

2.2.2 Reliability of Selection

The search and selection process was carried out by the first author. During this process, quality was ensured in two ways. During the search (step 1), the second author selected articles from the ‘continuous’ assessment search in parallel to the first author. Each author made a short overview of why they selected these articles. Subsequently, they discussed and argued their choices till an agreement on which articles to include in step 2 was reached.

Additionally, in step 4, the second author judged a subset of 14 articles from step 3 (130 total, a little over 10%) on whether they should be included in the full review. The authors agreed on 12 of the articles, but both included one article that the other did not include. This led to an acceptable inter-rater-reliability of Cohen’s $\kappa = 0.65$. After careful reconsideration of the two articles, one was not relating intermediate assessment to final examination outcomes and was therefore not included. The other article was at first discarded by one author, because the information about intermediate assessment was limited, but since it did relate intermediate assessment to final examination results it was included in the final selection.

2.2.3 Analysis

Several papers describe research on multiple cohorts, multiple assessment types, or multiple studies. During the analyses, these will be treated separately, raising the total number of outcomes over 88.

Relevant characteristics for analysis were decided upon discussion with educational experts (Day et al. 2016) and inspired by the typology of peer assessment by Topping (1998) as, to our best knowledge, a similar typology of intermediate assessment is not available. Furthermore, characteristics that seemed salient while reading the articles were added as well. All characteristics were investigated while articles were grouped on the assessment type level. Table 2.2 shows an overview of all articles and their assessment characteristics. All studies in this table are identified by their first author and publication year. The table is organised on the level of assessment type, and each condition (or assessment type) is presented on a separate row, to be able to clarify the differences in characteristics within the same articles.

To further exemplify the clusters of characteristics of intermediate assessment that are present in the literature, illustrative articles for the clusters will be discussed in more detail. These articles are selected because they present relevant results pertaining to the cluster, but they are not supposed to be representative of all studies included in the cluster.

Table 2.2. Overview of Most Important Assessment Characteristics of the Articles Included in the Review

Assessment Type	Article	Country	Discipline	Frequency	Feedback	Mandatory	Assessor	Scoring
Bedside Assessment	Burch (2006) Deane (2015)	ZA/US IE	Medical Medical	1 per week 5	Structured Feedback Form Verbal feedback and feedback form	Yes Yes	Clinician Educator Facilitator	Rating, not contribute to grade Rating, not contribute to grade
Case	Maley (2008) Palmer (2014) (1) Palmer (2008)	AU AU AU	Medical Medical Medical	No info 12-38 12	Model answers Model answers, explanations Detailed	No No No	No info No info No info	No info None None
Clickers	Cruz (2011) (1) Gauci (2009) Hancock (2010) Knight (2005) Nicol (2009) (2) Pascal (2002) (1)	BR AU US US UK US	Physics Medical Business Biology French Engineering	No info No info No info Each lecture No info No info	No info Correct answers and discussion No info Group discussion Group discussion Discussion	No No No info No info No info	No info No info Teacher No info No info Teacher	No info None Course credit Points, contribute to grade No info Course credit
Clinical Logbook	Anziniani (2008)	UK	Dentistry	17	No info	Yes	Supervisor	Formative, not contribute to grade
Computer Aided Assessment	Gill (2008)	UK	Mechanics	No info	Elaborate, worked examples	No	No info	No info
Conferencing Assessment	Baleghirzadeh (2012)	IR	English	8	Content	Yes	Instructor	Formative, not contribute to grade
Data interpretation	English (2006)	UK	Medical	1	Model answers	Yes	Peers / Teacher	~20% of grade
Dynamic Assessment (tweets)	Cacchione (2015)	ES	Italian	Two moments during class	Content comments and correction	No	Peers / Teacher	No info
Exercises	Buchanan (2000) (1) Buchanan (2000) (2) Carnegie (2015) Kennelly (2011) Krontritz-Litowitz (2008) (2) Snowhall (2013) (3)	UK UK CA IE US ZA	Psychology Psychology Medical Economics Economics	3, 2 revisions 5 4 10 1 12	No info No info Content, worked example No info No info No info	No info Yes No No info Yes No	No info No info No info Teacher No info No info	No info No info Points, contribute to grade 25% of grade No info None
Formulate questions	Hardy (2014) Zhang (2015)	UK US	Science Medical	At least 1 8	No No	Yes Yes	Teacher No info	2-5% of grade None
Game based assessment	Hooshyar (2016) (1) Hooshyar (2016) (2) Hooshyar (2016) (3) Schlegel (2014)	MY/TW/KR MY/TW/KR MY/TW/KR DM	Computer Science Computer Science Computer Science Medical	1 1 1 No info	Immediate knowledge of results Delayed knowledge of results Immediate elaborate Formative	Yes Yes Yes No info	No info No info No info Teacher	Score, not contribute to grade Score, not contribute to grade Score, not contribute to grade Points, no info about grade
Homework	Cruz (2011) (1) Cruz (2011) (2)	BR BR	Physics Physics	No info No info	No info No info	No No	No info No info	None 2 bonus points on examination
Instructional Design Task	Kim (2013)	KR	Teacher Education	1	Rubric and feedback messages	No info	Peers / Self	Rubric
Learning Modules	Hernick (2015)	US	Pharmacology	6	Content	No	No info	Formative, not contribute to grade

Table 2.2. Overview of Most Important Assessment Characteristics of the Articles Included in the Review (continued)

Assessment Type	Article	Country	Discipline	Frequency	Feedback	Mandatory	Assessor	Scoring
Midterm / examination / test / assessment	Azzi (2015) (2)	CA	Medical	2	Immediate feedback	Yes	No info	10% of grade
	Brown (2015) (2)	ZA/US	Medical	13 / 1	No info	Yes	Teacher	~21.7% of grade / 2 pts per q
	Carillo-de-la-pena (2009)	ES	Medical/Psychology/Biology	1	Correct answers	No	Teacher	Up to 5% bonus on grade
	Carillo-de-la-pena (2012)	ES	Psychology	3	Feedback on performance	No	No info	30% of grade
	Ernst (2008) (1,2)	AU	Medical	No info	Model answers and tutorial discussion	No	No info	Formative, not contribute to grade
	Fakharoemphol (2011)	US	Physics	No info	Correct answer or full solution	No	No info	No info
	Harwood (2006)	UK	No info	1	No info	No	No info	No info
	Henley (2003)	AU	Dentistry	After each educational meeting	Correct answers, model answers	No info	No info	No info
	Henley (2001)	AU	Science	After each educational meeting	Correct answers, model answers	Yes	No info	No info
	Jain (2012)	IN	Biochemistry	4	Correct answers, frequent errors, tips	No info	Teacher	No info
	Kerdijk (2013)	NL	Medical	3	Correct answers	Yes	Teacher	100% of grade
	Krasne (2006)	US	Medical	8	Correct / Incorrect, explanations	7 required	No info	Score, not contribute to grade
	Lawrence (2013) (1)	US	Psychology	3	No info	Yes	No info	No info
	Lawrence (2013) (2)	US	Psychology	1	No info	No	No info	Extra credit
	MacDermott (2013)	US	Economics	3	None	Yes	Teacher	60% of grade
	Micklewright (2010)	US	Sport Science	7	Specific and generic class based	Yes	Teacher	100% of grade
	Nelson (2009)	US	Biology	Weekly	Peer critique, teacher modelling	No info	No info	Points, not contribute to grade
	Nicol (2009) (2)	UK	French	Fortnightly / 1	No info	Some	Self and Teacher	~30% of grade
	Olson (2004)	US	Dentistry	1	Correct answers	No	No info	Score, not contribute to grade
	Palmer (2014) (2)	AU	Medical	Weekly	None	Yes	No info	10% of grade
	Pedrigones (2009) (2)	ES	Engineering	Weekly (8) / monthly	Correct answers	Yes	Teacher	Bonus points on final examination
	Ramirez (2010)	CL	Medical	3 / 1	No info	Yes / No	Teacher/Self	Graded / not contribute to grade
	Raupach (2013) (1)	DE	Medical	1	No info	Yes	Teacher	1% of credit of the year
	Raupach (2013) (2)	DE	Medical	1	No info	Yes	Teacher	Score, not contribute to grade
	Rezaei (2015) (1,2,3)	US	Research Methods	No info	No info	Yes	Teacher	Contributes to grade
	Sample (2014)	US	Veterinary Medicine	Weekly (6)	Model answers	Yes	Teacher	Scale, no info about grade
	Sargent (2012) (2)	US	Law	1	No info	Yes	Teacher	Graded
	Stweya (2014)	ZA	Chemistry	1 / 3-4 / 1	No info	Yes	Teacher	100% of grade (50/40/10)
	Snowball (2013) (2)	ZA	Economics	2	No info	No	No info	30%
	Sullivan (2013)	UK	Medical	1	None	No	Four examiners	No info
	Taradi (2005)	HR	Medical	10	Performance feedback	No	Self	Not contribute to grade
	Tuunila (2015) (1)	FI	Chemical Engineering	4-5	General feedback, suitable answers	No	Teacher	Contributes to grade
	Tuunila (2015) (2)	FI	Chemical Engineering	1 / 6	General feedback, suitable answers	No	Teacher	Contributes to grade
	Van der Kleij (2012) (1)	NL	Economics	1	Immediate, elaborate	No	Computer	Score, not contribute to grade
	Van der Kleij (2012) (2)	NL	Economics	1	Delayed, elaborate	No	Computer	Score, not contribute to grade
	Van der Kleij (2012) (3)	NL	Economics	1	Delayed knowledge of results	No	Computer	Score, not contribute to grade
	Velan (2008)	AU	Biomedical Science	1	No info	No	Self	No info
	Winniger (2005)	US	Psychology	1	Class discussion	Yes	Teacher	Extra credit
No info	Pedrigones (1) (2009)	ES	Engineering	No info	No info	No info	No info	No info
Oral Examination	Priya (2) (2012)	IN	Dentistry	6	No info	Yes	Teacher	Score, not contribute to grade
Portfolio	Behrouzi (2015)	IR	English	12 entries	Conference with instructor	Yes	Instructor/Self	No info
	Thome (2006)	SE	Biomedical Science	3	Portfolio discussions	Yes	Teacher	Scale, not contribute to grade
Practice Attachment	Hafling (2012)	SE	Medical	2	Feedback form and narrative comments	Yes	Tutor/Self	Score, not contribute to grade
Pre-test/post-test	Lemus-Zuniga (2015)	ES	Computer engineering	No info	Yes	No info	Teacher	No info
	Munson (2015)	US	Pharmacology	6	Correct answers	No	No info	Score, not contribute to grade
Presentation	Pedrigones (2009) (3)	ES	Engineering	1	No info	Yes	Teacher/Peers	~20% of grade
Project	Rezaei (2015) (1,2,3)	US	Research Methods	No info	No info	Yes	Teacher	No info

Table 2.2. Overview of Most important Assessment Characteristics of the Articles Included in the Review (Continued)

Assessment Type	Article	Country	Discipline	Frequency	Feedback	Mandatory	Assessor	Scoring
Quiz	Azzi (2015) (1)	CA	Medical	17	Immediate feedback	Yes	No info	Formative, not contribute to grade
	Bijol (2015)	US	Medical	5	Correct answers and explanations	No	Self	No info
	Bookhan (2014)	ZA	Dentistry	2	Correct answers and group discussion	No	No info	% score, not contribute to grade
	Brown (2015) (1)	ZA / US	Medical	13	No info	Yes	Teacher	~11.6% of grade
	Cukusic (2014)	HR	Economics	8	Correct / incorrect	≥60% on 6	Self	Score, not contribute to grade
	Dobson (2008)	US	Exercise Physiology	10	Yes	Yes	No info	10% of grade
	Domenech (2015)	ES	Economics	10	Suggested answers	Yes	Teacher	85% of grade
	Fauth (2015)	US	Chemistry	Weekly	No info	No info	No info	No info
	Glass (2009)	US	Psychology	28	Correct answers	No info	No info	Bonus points
	Ibabe (2010)	ES	Psychology	No info	Correct / incorrect, hints	No	No info	Score, not contribute to grade
	Jacoby (2014)	UK	Biology	Weekly	Yes	No info	Self	Score, not contribute to grade
	Kibble (2007) (1)	GD	Medical	2	No info	No	No info	Score, not contribute to grade
	Kibble (2007) (2)	GD	Medical	2	No info	No	No info	1% of grade, on login
	Kibble (2007) (3)	GD	Medical	2	No info	No	No info	2% of grade, score >30%
	Kibble (2007) (4)	GD	Medical	2	Incorrect answers shown	No	No info	2% of grade, based on actual score
	Kibble (2007) (5)	GD	Medical	2	Incorrect answers shown	No	No info	4% of grade, based on actual score
	Kibble (2011)	US	Medical	Weekly (12)	Correct answers, faculty available for questions	No	No info	None
	Krontiris-Litowitz (2008) (1)	US	Neuroanatomy	2	No info	Yes	No info	No info
	Lawrence (2013) (3)	US	Psychology	After each chapter	No info	No	No info	No info
	Marden (2013) (1,2)	AU	Medical	3	Individualised	Yes	Teacher	5% of grade
	Marden (2013) (3)	AU	Medical	2	Individualised	Yes	Teacher	7.5% of grade
	Marden (2013) (4)	AU	Medical	5	Individualised	Yes	Teacher	2% of grade, score >90%
	McNulty (2015) (1)	US	Medical	3	Review session in class	No	Self	No info
	McNulty (2015) (2)	US	Medical	After each lecture (14)	Instructive online feedback	No	Self	No info
	Mitra (2015)	MY	Medical	3	Content feedback	No info	No info	% score, not contribute to grade
	Paschal (2002) (2)	US	Biomedical Engineering	4	Correct answers	No info	Teacher	Course credit
	Rezaei (2015) (2-3)	US	Research Methods	No info	No info	Yes	Teacher	30% of grade
	Sargent (2012) (1)	US	Law	5	Model answers and rubric	Yes	Teacher	Ungraded
	Williams (2008) (1)	UK	Chemistry	20	Online feedback	No	Self	Formative, not contribute to grade
	Williams (2008) (2)	UK	Chemistry	20	No info	No	Teacher	5% of grade, for total of quizzes
	Wilson (2011)	CA	Geography	After each class	Correct / incorrect	No	Self	None
	Zhang (2015) (1,2)	US	Medical	10	Correct answers, open feedback	Yes	Teacher	None
Simulated Patient	Gums (2014)	US	Pharmacology	1	Rubric and talking points for individual	Yes	Facilitator	Rubric, not contribute to grade
Writing Assignment	Birjandi (2012) (1)	IR	Teaching English	Each session (12)	Oral or written teacher feedback, self-reflection	Yes	Teacher/Self	No info
	Birjandi (2012) (2)	IR	Teaching English	No info	Rubric	Yes	Self/ each	Rubric, not contribute to grade
	Birjandi (2012) (3)	IR	Teaching English	No info	Rubric	Yes	Peers/Teacher	Rubric, not contribute to grade
	Birjandi (2012) (4)	IR	Teaching English	No info	Rubric	Yes	Self/Peers	Rubric, not contribute to grade
	Birjandi (2012) (5)	IR	Teaching English	No info	No info	Yes	Teacher	No info
	Covic (2008)	AU	Psychology	1 (resubmission)	Comments and suggestions, Form, Group feedback	No	Teacher	Score, 20% of grade
	Fisher (2011)	AU	Management	1 (draft)	Feedback sheet and discussion	No	Teacher	None
	Gielen (2015)	BE	Educational Science	3 (drafts)	Feedback form	Yes	Peers	Rubric, not contribute to grade
	Issakson (2008)	SE	Archaeology	10	Short written feedback	Yes	Lecturer	100% of grade
	Mulder (2014)	AU	Animal Behaviour	1 (draft)	Structured review sheet	Yes	Peers	Score, not contribute to grade
	Nicol (2009) (1)	UK	Psychology	6 light, 6 deep	Peer discussion, model answers afterwards	Yes	Peers/Self	No info
	Perdigones (2009) (3)	ES	Engineering	1	No info	Yes	Journal/Teacher	30% of grade
	Priya (2012) (1)	IN	Dentistry	6	No info	Yes	Teacher	Score, not contribute to grade
	Snowball (2013) (1)	ZA	Economics	2	Rubric	No	Peers/Teacher	30% of grade
	Stellmack (2012)	US	Psychology	1 (resubmission)	Rubric	No info	Peers/Self/Teacher	Half points for draft
	Wakefield (2014)	UK	Sport Studies	1	Feedback checklist or marking sheet	Yes	Teacher	60% of grade

Note. Bolded references have found a positive relation between the assessment and student grades.

2.3 Results & Discussion

2.3.1 Article Background Information

Research into intermediate assessment in higher education is conducted all over the world, evidenced by papers from South Africa (3), Iran (3), and Brazil (1). However, the majority of the research discussed in the current review is conducted in the United States of America (25), Australia (12) and the United Kingdom (12).

Similarly, research is being conducted in different disciplines, with Medicine (38), STEM fields (19), the Social Sciences (12), and Law, Business & Economics (12) being main contenders. The largest group of studies is conducted with first year students (with 23 articles exclusively using first year students).

Generally, research on intermediate assessment falls into the following two categories. The first category is correlational studies addressing the relationship of the intermediate assessment and final assessment grades ($N=16$). The second category includes studies that compare final examination scores for groups with and without intermediate assessment, or scores on different courses ($N=62$). Furthermore, fourteen studies investigated a correlation as well as a comparison. In the 88 selected articles, 77 studies found positive results of intermediate assessment, and 23 found no results. To reiterate, several articles discussed multiple outcomes, which is why the total number of results exceeds 88.

2.3.2 Assessment Type

In total, 24 different types of assessment are used in the 88 articles. Table 2.2 shows that the most prevalent assessment types are midterms/examinations, quizzes and writing assignments. These three are all well-known, 'generic', conventional assessment forms, but the range of assessments also includes game-based assessments, clinical bedside assessments, and the use of twitter as a dynamic assessment tool.

In the following sections, assessment type will be used as the backbone in the exploration of the different assessment characteristics and their relation to student grades. Because of the alignment between the assessment type and learning goals, it stands to reason when adapting an assessment, teachers will vary other characteristics first. The first two sections focus on the prevalence of assessment types per discipline and frequency of assessment, and will be discussed briefly. In the final four sections we will discuss characteristics that can

be fine-tuned by teachers, and a short description of the characteristic will be followed by a more in-depth discussion of articles that portray this characteristic.

2.3.3 Assessment Type and Discipline

The disciplines in which the research took place can be found in the ‘Discipline’ column of Table 2.2. When disciplines are aggregated to a more general level it becomes evident that midterms (or related generic forms of examinations) and writing assignments are the only types of assessment used across all disciplines; Quizzes are also represented across the majority of disciplines. These three generic assessment types can be implemented in all disciplines. There are also some discipline specific assessment types, like clinical bedside assessment for medical students. Interestingly, the oral presentation, which can be considered a generic assessment type as well, is only used in one study in one discipline.

2.3.4 Assessment Type and Frequency of Assessment

Students are confronted with a vastly differing amount of assessments, as can be seen in the frequency column of Table 2.2. The majority of authors describe courses that consist of one to ten intermediate assessments per course. Palmer and Devitt (2014) mention the highest number of intermediate assessments in one course: students had access to 38 cases for intermediate assessment.

Since not all courses run for the same amount of time, and eight assessments in an eight-week course will be perceived differently than eight assessments in a twenty-week course, direct comparison of assessment frequency in different studies may be biased. Some authors specify that their assessments occur weekly (e.g. Nelson et al. 2009), fortnightly (e.g. Nicol 2009), or after each educational meeting (e.g. Henly & Reid 2001).

Overall, it can be argued that not the frequency of assessments is the most important characteristic, but rather, the alignment between the frequency of assessment and the course goals. For example, the 38 cases of Palmer and Devitt (2014) may seem like a lot, but they reported completion rates of 90%, indicating that the number of assessments was suitable for the course.

2.3.5 Assessment Type and Feedback

The feedback column in Table 2.2 indicates that the majority of authors report some form of feedback. The elaborateness of the feedback ranges from simply stating whether an answer was correct, to giving elaborate, content-related,

qualitative feedback. It seems that some types of feedback are closely related to specific types of assessment, which is most clearly the case for the rubric or feedback form, which is mainly employed in more open tasks like writing assignments and patient encounters. Interestingly, it seems that the category of feedback which one expects to have the largest influence (elaborated content feedback) does not produce more positive effects than supplying students with their results or the correct answers. This may have to do with different feedback effects for different student groups. For example, Shute (2008) poses that while elaborated feedback is the optimal feedback type for lower achieving students, high achieving students often get enough information from corrective feedback. This is also in line with work by Brookhart (2001), who discovered that successful students often use summative test information for learning purposes.

Two papers that illustrate how the level of feedback can relate to student results are Hooshyar et al. (2016) and van der Kleij et al. (2012). Both of these papers compare conditions with elaborate feedback to feedback that only consist of knowledge of results.

Hooshyar et al. (2016) investigated an online game based formative assessment (a tic-tac-toe variation, where a correct answer on a question allows the student to place a token on the grid) and its influence on problem solving skills. The population was 52 first year computer science students, 22 in a control class working with what the authors call 'traditional approaches' (Hooshyar et al. 2016, p. 27), and 30 in the experimental class working with the game-based assessment. There were three feedback conditions in the game-based assessment: delayed knowledge of results, immediate knowledge of results, and immediate elaborate feedback. Results indicate that students in the experimental condition performed better than their peers in the control group, and that immediate elaborate feedback proved to be the most useful.

The research by van der Kleij et al. (2012) focused on a computer based assessment for learning task for 152 first year economics students. The study compared summative assessment results of three groups. Group one consisted of 52 students who received immediate knowledge of correct response and elaborate feedback on their performance. Group two (N=48) received delayed knowledge of correct response and elaborate feedback, and group three (N=52) only received delayed knowledge of their results. This intervention had no effect on the final summative assessment results, indicating immediate or delayed elaborate feedback did not improve final performance.

The studies by Hooshyar et al. (2016) and van der Kleij et al. (2012) exhibit a different influence of feedback, where the former does find an influence of elaborate feedback and the latter does not. There are several similarities in the design of the two studies. Both focus on a single assessment for first year students, both use multiple-choice questions and both assessments are performed on a computer. The main difference is that for Hooshyar et al. the assessment was a mandatory part of the course, following a specific learning module on problem solving. In the case of van der Kleij et al., the assessment for learning was mainly part of the experiment, and participation in the assessment was not a prerequisite of participating in the course.

Because of this clear difference, it is of interest to investigate the papers under review on their use of mandatory assessments. It could be reasoned that whether intermediate assessment is mandatory can influence students' final grade.

2.3.6 Assessment Type and Mandatory Assessments

There is a somewhat even divide in the use of mandatory versus voluntary intermediate assessment, which can be seen in the mandatory column of Table 2.2. Interestingly, whether or not an assessment is mandatory does not immediately seem to influence students' results. For mandatory and voluntary assessments, the amount of studies finding no positive influence of intermediate assessment is approximately 36% and 31% respectively.

Carrillo-de-la-Peña and Pérez (2012) used mandatory intermediate assessment in an interesting way. Their research focused on three classes of second year psychology students for three consecutive years, with a total population of 903 students. These students were given the choice to partake in Continuous Assessment (CA) in their course, or to stick to Traditional Assessment (TA). Although the choice for CA was voluntary, the 30% of students who chose CA were then required to participate in mandatory formative assessments and 'complementary individual and group tasks' (p. 46). At the end of the educational term all students took the same final examination, which counted for 70% for CA students and 100% for TA students.

Comparison of the groups shows that fewer CA students dropped the course during the semester, more CA students passed the examination, and CA students got higher marks for the examination. These results were consistently found for all three academic years under investigation. Surprisingly, TA students remarked in a questionnaire that they believed the CA system would have helped their

learning and 33.3 to 60.4% of them would have chosen it, if they were starting the course now.

Snowball and Mostert (2013) described voluntary participation in intermediate assessment for a writing assignment. During their course, Snowball and Mostert (2013) were faced with technical difficulties during the period of online submission of the essay assignment for peer assessment, which led them to also offer an alternative exercise, making the peer assessment voluntary. This alternative led to only about half of the 800 students enrolled in the macroeconomics class participating in peer assessment. Regression analyses showed that participating in the peer assessment did not predict students' final essay grades. Even though 58% of the students agreed that peer feedback helped them to improve the quality of their essay, only 10% of the students made significant changes to their essays. Furthermore, when comparing peer assessment on the first version with tutor assessment of the final version, it seemed that peers were marking in a much smaller range (between 60-68%) than the tutors were, which may not have motivated students to implement significant changes.

In sum, research does not show differences between mandatory and voluntary assessments with respect to students' grades. However, in the case of voluntary assessments, selection effects should be taken into account. For example, Covic and Jones (2008) gave third year psychology students the opportunity to resubmit their 1200-word essays, however, high performing students chose not to participate in the resubmission. In other cases, it may be that high performers do participate, whereas lower performing students do not. This begs the question what other variables influence the role of intermediate assessment. One that clearly comes forward from Snowball and Mostert's research is 'who is the assessor?', and another is 'what can students gain with participating in the assessment?'. We will address these questions in the following sections.

2.3.7 Assessment Type and Assessor

As is shown in the assessor column of Table 2.2, the majority of authors describe assessments that were assessed by the teacher (or a lecturer, facilitator, supervisor, etcetera), or did not explicitly name who the assessor was. Research that uses peer-assessment or peer feedback is mainly limited to writing assignments, whereas the majority of authors that employ self-assessment do so using quizzes.

One study that compares the differences between self-, peer-, and teacher-assessment, is Birjandi & Hadidi Tamjid (2012). In this paper, the researchers investigate the writing performance of 157 Teaching English as a Foreign Language students, divided into four experimental groups and one control group. Students in all five groups wrote essays and subsequently were assessed on those essays in different ways. In the first group, teachers gave oral feedback when students read their written essays aloud in class and students wrote reflection journals detailing what they learned and what problems they had during the class. In the second group, students self-assessed their essays using a rubric and were able to make revisions based on this assessment. Afterwards, teachers assessed the essays and students were able to compare their self-assessment to that of the teacher. The third group was set up very similarly to the second group, but instead of self-assessment, this group was divided into smaller peer groups where students rated each other's work using the rubric. The fourth group was the only group that did not use teacher assessment. In this group students first assessed themselves and subsequently went into peer groups to assess the work of their peers. The fifth group was the control group, where teachers assessed the essays.

Before and after writing their essays, students took an English proficiency test and a writing test, to see if their performance had improved. For all five groups, scores on the writing test increased after the classes on essay writing. Analysis of variance with post-hoc tests indicated that the third group (peer- and teacher assessment) scored significantly better than the first, second and fifth groups.

These findings, together with several other articles using peer assessment, suggest a positive influence of peer assessment. However, the influence of teacher assessment should not be excluded. When comparing the different groups in Birjandi and Hadidi Tamjid's (2012) research, it shows that the fourth group (using self- and peer assessment) did not differ from the other groups, indicating that the results of the third group are not just because of a favourable role of peer assessment, but of the combination of teacher and peer assessment. When combining two or more assessors, the teacher is often the assessor who decides on the final evaluation of the assessment. This sequence indicates that in most assessments, the teacher is still the decider.

2.3.8 Assessment Type and Reward

Another important factor of the assessment is the type of reward students get for participating. In the scoring column of Table 2.2 the two largest categories contain assessments that either make up a percentage of the final grade, or assessments on which students simply receive a (formative) score, that does not count towards the final grade. In some cases teachers gave students bonus points for completing the assessments. Out of the two biggest categories the most positive results are found in studies awarding a percentage of course grade, maybe indicating that students who feel that they have something to gain with the assessment will exert more effort.

Kibble (2007) provides an interesting example of how rewards seem to influence the effectiveness of an intermediate assessment. Five cohorts of medical physiology students could participate in two voluntary quizzes as a means of formative assessment. In the first cohort, students got no rewards for participating in the quizzes. In the second cohort, students could earn 0.5% course credit per quiz simply for taking the quiz, irrespective of their grade. Students in cohort three received 1% credit per quiz if they scored >30% and they were free to retake the quiz as often as they wanted in a one-week period. In cohorts four and five students were rewarded with 1 or 2 percent course credit, 'allotted as a function of the actual quiz score attained from the better of two attempts' (p. 254) respectively.

As Table 2.2 indicates, in cohorts four and five there was no longer a positive effect of participating in the quizzes. However, participation in the quizzes increased drastically, from 52% in the first to 98% in the fifth cohort, when the rewards were greater. In all cohorts the summative assessment scores of quiz takers were compared to non-takers, which may explain the lack of difference in cohorts four and five, since the group of non-takers was small. Furthermore, results seem to indicate that students in cohorts four and five did not engage with the quizzes with the formative purpose in mind, but mainly to gain the course credit. Kibble's main evidence for this is that these students got a very high score on their first attempt of the quiz, subsequently did not try to retake the quiz for improvement and were unable to sustain their high performance on the summative examination.

This may be evidence to be cautious with rewards for intermediate assessments with a formative purpose. Gibbs and Simpson (2004, p. 23) mention that assessment only supports learning when the provided feedback is 'received

and attended to', subsequently showing several research outcomes that indicate students often discard feedback and only tend to grades when they are provided together.

However, grades or bonus points can work as an incentive for students to participate in the intermediate assessment. Harland et al. (2015) stated that students are focused on working for graded assessments more than for ungraded educational activities.

2.4 Intermediate Assessment Scenarios

The review shows how various characteristics of intermediate assessments are related to student grades. In this section, three scenarios for intermediate assessment are presented, which indicate combinations of the characteristics discussed in the results section. These scenarios will illustrate how the different characteristics discussed in the results section could be implemented in a higher education setting. Each scenario will focus on a different assessment type: the quiz, the written assignment, and the authentic professional assignment. These three types reflect different aspects of the intermediate assessment process. In the case of all these scenarios it is important to keep in mind that different assessment goals may ask for different assessment characteristics.

2.4.1 Scenario 1: Quiz

As is evidenced in the overview of the assessment types, quizzes are an often used assessment type. In the current scenario, the quiz is implemented in a first-year social science course. This course has weekly lectures and lasts for ten weeks. Teachers have felt that students who are attending the lectures often come unprepared, so they decided to implement quizzes to keep the students on track in their learning. These quizzes are not mandatory, because the teachers also want to facilitate students' ownership over their learning, but participation in the quizzes is rewarded with a 0.5 (out of 10) bonus when students pass all quizzes. The student population of this course is fairly large, with about 400 students enrolled.

Quizzes take place after each lecture, to encourage students to keep up with their studying. A quiz after each lecture clearly aligns the assessment with the materials studied during the lecture. For students taking a 10-20 item

quiz, consisting mainly of multiple choice questions and some short answer questions, is also a good instrument for self-assessment and to check whether they understand all the class materials.

To alleviate teacher workload, the quizzes are implemented using an online platform, with the additional bonus that students can access the quizzes at their own convenience. In the online quiz, the assessment process is automated and students are supplied with feedback. This feedback is tailored to the student's level of performance, since Shute (2008) suggests that low achieving students benefit from elaborated feedback, where for high achieving students simple corrective feedback is often enough. Elaborated feedback on the multiple-choice questions consists of explanations on each answer option, including why the correct choice is correct. For the short answers questions, a model answer with the most important parts of the correct answer will be supplied after answering the question. Because the computer cannot adequately assess the correctness of short answer questions, only the multiple-choice questions are used when deciding whether or not a student passed the quiz.

2.4.1.1 Reflections on the quiz scenario

Several advantages of the quiz are: They are often easy to design and use, administering them is time-efficient, they are fit for use in large groups, they help keep students on track, they promote active learning, and can reduce test anxiety. In the scenario the main two advantages were the ease of large group assessment and the ability to keep students on track.

An advantage of hosting the quizzes on an online platform is that questions for each week can be randomly chosen from a larger bank of quiz items. Quizzes that are composed by selecting items from a larger item bank allow meaningful repetition of the quiz more easily, indicating that students learn the associated learning materials instead of simply memorising the correct answers on a quiz. However, building a large item bank is a labour-intensive process. Furthermore, if the item bank is large enough, there may be no new questions available for the final examination anymore.

One disadvantage of intermediate quizzes is strategic quiz taking, as was indicated when discussing the research by Kibble (2007). Simply memorising quiz answers, or other surface learning approaches for answering multiple choice questions, often do not promote long term learning effects (Gibbs & Simpson 2004). This can be overcome by integrating a discussion of quiz answers in the following lecture.

We believe that the majority of the characteristics described in Scenario 1 can be applied to scenarios containing a midterm or formative examination as intermediate assessment as well. The main difference would be that examinations usually occur less often, are larger in size than quizzes are, and usually have higher stakes than quizzes do. When midterms are used to complete certain topics that will not return on the final examination, it may be of less importance to offer elaborate feedback.

2.4.2 Scenario 2: Writing assignment

Writing assignments were also among the most often used assessment types. This scenario concerns a twelve week third-year business course of 80 students, where one of the final assessments is a 3000-word business plan. The intermediate assessment that is part of this scenario is a peer assessment. Students are invited to voluntarily submit a draft of their business plan as an opportunity to receive formative peer feedback.

In the second week of the course, before students start writing their plans, they are invited to participate in a peer feedback workshop. During this workshop the teacher introduces them to the procedure of peer feedback, using an online peer feedback platform. Furthermore, during this workshop, the students familiarise themselves with the assessment criteria. All students are asked to provide feedback on a business plan written by a student in a previous cohort, using a proposed feedback form. After this exercise the teacher and students discuss the criteria and amend them where necessary.

After the peer feedback workshop, students have two weeks to finish their draft for peer feedback. Following submission of the draft, the online platform automatically matches two students to give feedback on each other's drafts. The teacher decided against predetermined couples matched on students' ability, since the voluntary nature of the draft submission meant that there was a risk that several couples would remain incomplete. Students have two weeks to complete the peer feedback assignment. When students have received their drafts with feedback, they subsequently have four weeks to finish their business plan and submit the final version to their teachers.

2.4.2.1 Reflections on the writing assignment scenario

This type of assessment is versatile and is often used to assess writing skill and content knowledge or discipline specific skills simultaneously. Writing

assignments often allow students to display their in-depth knowledge and pose a good opportunity to implement peer assessment into the curriculum. Peer assessment is a valuable learning tool, where students learn from the feedback they receive, as well as from providing feedback to their peer. Furthermore, adopting a peer feedback moment into the writing assignment provides the students with an opportunity to receive feedback on their work, without significantly increasing teacher workload.

It is instrumental that students receive adequate peer feedback training (van Zundert, Sluijsmans, & van Merriënboer 2010). Van Zundert et al. found that peer feedback training positively influenced the quality of peer feedback as well as students' attitudes towards peer assessment. Furthermore, during this training students can familiarise themselves with the assessment criteria and propose to amend them where they feel that they do not fit the assignment.

Results from Snowball and Mostert (2013) indicate two possible problems with peer assessment on writing. First, students did not make significant improvements based on the peer feedback. Second, peers marked on a much smaller range than teachers did, indicating they were less critical, but also less positive. The latter problem could be overcome by training and gaining experience with providing peer feedback.

2.4.3 Scenario 3: Authentic assessment

Even though authentic assessments were not highly prevalent in the current sample, they do make up an important facet of the assessment domain. In courses in, for example, the medical field more traditional written assessments are often unable to assess the professional skills that are needed in the workplace, necessitating assessment types that simulate these practical situations, or sometimes even take place in practice.

This scenario concerns an intermediate assessment in a semester long fourth year medical course, attended by 35 students. This course focuses on clinical skills development and the final examination in this course is a 10 station Objective Structured Clinical Examination (OSCE). Since clinical skills cannot be taught by purely lecturing, this course features a weekly practical tutorial in addition to the lectures focusing on theory. These tutorials are small scale, with each tutorial only consisting of seven students. To fully prepare students for the examination, all stations that are included in the OSCE are administered formatively during the tutorials, at a rate of one station every other week.

The formative OSCE focuses on the stations that were discussed during lectures and tutorials that week. All students perform the tasks associated with the station on a simulated patient, while being observed by their peers and a tutor. The tutor subsequently judges students' performance on the station by completing a rating form. On this form there is also ample space for the tutor to provide narrative feedback on the students' performance. Peers have no formal assessment role, but they are invited to share their observations and tips.

2.4.3.1 Reflections on the authentic assessment scenario

Authentic assessments help students experience real life (or close to real life) situations where they need to apply their skills. However, these assessments are often time consuming.

An important facet of the current scenario is that students are learning from each observing each other. This does mean that the first student who completes the station may be at a disadvantage, because they have not observed a peer. It is up to the tutor to facilitate that different students are the first to participate in the station each assessment. Furthermore, teachers can model the correct performance or show video examples of experts.

2.5 Concluding remarks

This review set out to identify which characteristics of intermediate assessment may play a role in improving students' final examination results. The results of the review show that intermediate assessment is employed widely in higher education, and that the variety in characteristics is very large. The instructional designs vary per study, depending on the educational context and the specific objectives the instructors had in mind.

The main outcomes indicate the following:

- (1) Corrective feedback is often enough, but lower achieving students may need more elaborate feedback.
- (2) Both mandatory and voluntary assessments can influence students' grades. Therefore, intermediate assessment is not only effective when it is mandatory, but can also be effective when it is voluntary.
- (3) Peer assessment is beneficial to students' grades, especially since providing feedback is a learning opportunity for students as well.

- (4) Rewarding students increases their participation, but may negatively influence the learning effects of the assessment when students are strictly focused on the reward and do not engage with feedback.

In a way, intermediate assessment characteristics are different pieces of a puzzle that, combined together, influence students' grades. Teachers can optimise their courses by checking which characteristics align with their goals, and combining these characteristics in such a way that students can achieve the educational learning goals more easily and perform better.

Chapter 3

**Teacher and Student Perceptions of Intermediate
Assessment in Higher Education**

Universities introduce intermediate assessment because it is understood to have positive effects on student behaviour and achievement. Yet, how intermediate assessment is perceived might be conditional for its success. The current study investigates both teachers' and students' perceptions of intermediate assessment. Teachers and students were interviewed and Student Evaluations of Teaching were examined. Results indicate that both teachers and students had generally positive perceptions of intermediate assessment. However, the two groups provided different reasons for their positive perceptions. Teachers and students agreed that intermediate assessment helps students to keep up with their study work. Moreover, teachers mentioned that they could assess various knowledge and skills with intermediate assessment, whereas students preferred intermediate assessments to test the same knowledge and skills as the final exam. This finding suggests that teachers and students in our study had different goals for intermediate assessment.

This chapter has been published in adapted form as: Day, I. N. Z., van Blankenstein, F. M., Westenberg, P. M., & Admiraal, W. F. (2017). Teacher and student perceptions of intermediate assessment in higher education. *Educational Studies*, Advance Online Publication. doi:10.1080/03055698.2017.1382324.

3.1 Introduction

Universities often take measures to entice students to start studying earlier in the semester, to make completing study programmes more feasible, and to increase study success in general. One of these measures is intermediate assessment (Van Berkel, Jansen, & Bax, 2012). Intermediate assessment, often also referred to as continuous or frequent assessment, refers to assessment or assignments that take place during the course period. Measures to increase study success, like intermediate assessment, are often only regarded as successful when they improve academic achievement or pass percentages. The perceptions of stakeholders, like teachers and students, are often overlooked in this process. It can be argued that successful implementation of intermediate assessment hinges on its perceived usefulness. Several authors (e.g. Entwistle, 1991; Prosser, Walker and Millar, 1996), have indicated that student perceptions of the learning context play an important role in student learning. Van Dinther, Dochy, Segers, & Braeken (2014) found that perceptions predict self-efficacy, which in turn predicts student outcomes. Since intermediate assessment can take various forms, it can be hypothesized that teacher and student perceptions are influenced by the type of assessment that is used. This study investigates both teachers' and students' perceptions of different forms of intermediate assessment.

3.1.1 Intermediate Assessment in Higher Education

In the current study, intermediate assessment is defined as an assessment which takes place during the course period, which can take many different forms, and is handed in to the teacher. As was mentioned before, intermediate assessment can have many different forms, like quizzes, essays, presentations, or projects for example. Other characteristics that can differ between intermediate assessments are for example the frequency, amount of questions, time to finish, modality (written/oral/computer/etc.), grading and weight of the assessment. Furthermore, the quality of the assessment is of importance as well. Lizzio, Wilson and Simons (2002) reiterated the importance of perceived appropriateness of the assessment for student learning approaches and student results. Intermediate assessments can be used for summative purposes, for formative purposes, or both. Voluntary homework assignments and class readings do not count as intermediate assessment. Findings from research into the effects of assessment in general and intermediate assessment in particular indicate both positive and negative effects. These will be discussed in more detail below.

3.1.1.1 *Positive effects of intermediate assessment*

Intermediate assessment can lead to increased time-on-task as evidenced in a study by Admiraal, Wubbels and Pilot (1999), where law students in a problem method course with several assessments spent more time studying than their peers in the traditional case-based course. Doumen, Broeckmans and Masui (2014) demonstrated a relationship between increased study time and higher exam grades, but Michaels and Miethe (1989) discovered that only students who spread their study time during the semester, as opposed to students who resort to cramming before the exam, profit from increased study time. Since students often focus their learning activities around examinations, intermediate assessments are an incentive for students to spread their study time (Cohen-Schotanus, 1999).

Higher results for students who spend more time studying during the semester may be explained by the cognitive advantages of spreading. Laboratory studies have shown that spreading content while studying leads to better retention in long-term memory than cramming does (Kornell, 2009). The advantages of spreading, or distributed practice, have also been extensively reviewed by Dunlosky, Rawson, Marsh, Nathan, and Willingham (2013) who conclude that it is one of two learning techniques they rate highest.

The second high-rated learning technique is practice testing (Dunlosky et al., 2013). Whereas practice testing is usually undertaken as a homework activity by students, intermediate assessment works following a similar mechanism of regular testing. The effectiveness is caused by the “testing effect”, which indicates that information is retained in memory better by testing it (Roediger & Karpicke, 2006). The testing effect is explained by the advantages of effortful retrieval, that is to say, by trying to recall knowledge from memory, the retention of that memory is strengthened. Roediger and Karpicke (2006) reviewed a large body of testing effect studies and found support for the testing effect in laboratory studies as well as in authentic learning environments.

Intermediate assessment creates opportunities for feedback, especially when teachers implement formative intermediate assessments (assessments with a main goal of improving learning by feedback instead of judging achievement, Bennet 2011). Gibbs and Simpson (2004) propose that feedback improves learning especially when it is provided ‘just in time’ (i.e. at a moment when it is still useful for students). Providing feedback to an intermediate assessment can improve exam results, as indicated by Bouwmeester and colleagues (2013). They

found that students who participated in voluntary online formative assessments scored higher on midterms and final exams, compared to their peers who did not participate.

3.1.1.2 Negative effects of intermediate assessment

Intermediate assessment can also have adverse effects. Harland, McLean, Wass, Miller and Sim (2015) concluded that assessing students frequently for grades led to strategic behaviour, because students were mainly focused on getting grades and not on gaining a deeper understanding of the material. Furthermore, competition arose, with non-graded study activities losing out. Harland et al. (2015) called this an “assessment arms race”. The notion of an assessment arms race is backed up by findings from Vos (1992) who found that students in the Netherlands spent a maximum of seven hours on studying. When the workload exceeded these seven hours, competition between study activities would occur.

Harland et al.’s (2015) conclusions with regard to the lack of deep understanding of the material may especially be problematic when there is a large overlap in content between the intermediate and final assessment. Research by Kling et al. (2005) indicates that frequent testing was only beneficial with regard to final exam scores when there was overlap between test items in the intermediate and final assessments. However, using several intermediate assessments with the same items as in the final exam is more in line with teaching to the test (for a definition see Popham, 2001), and probably does not lead to deep learning.

Within the context of universities, perceptions of intermediate assessment are not always positive. An example of this is a report from our university’s newspaper where the following point is raised: Using a lot of intermediate assessments in higher education can make students feel patronised, and treated like they are back in high school (Bongers, 2015).

Finally, intermediate assessment requires teachers to invest time in designing more assessments, and to provide more feedback and grades, than with final exams only (Leeming, 2002). Additionally, when intermediate assessments are administered during class time, this reduces the time available for teaching. However, Leeming (2002) argued that this reduced teaching time mainly led to less redundancy of materials and not to omission of teaching core materials.

3.1.2 Perceptions of Intermediate Assessment

The aforementioned literature reflects both positive and negative effects of intermediate assessment in higher education. However, when discussing the potential and actual effects of intermediate assessment, these studies were silent about perceptions. We believe that teacher perceptions of assessment are an important factor in the implementation process. Several researchers have investigated assessment perceptions of teachers and students, however, these studies often focus on assessment in general and not on intermediate assessment in particular.

Goos, Gannaway and Hughes (2011) used surveys, focus groups and semi-structured interviews to investigate the assessment perceptions of associate deans, course coordinators, and students, showing that these three groups identify different challenges when discussing assessment. Course coordinators' comments focused a lot on what can be described as assessment conditions: workload, bureaucracy, and lack of support. Conversely, students' main point of dissatisfaction seemed to be insufficient feedback, even though first year students were unable to qualify what sufficient feedback meant to them. Associate deans also identified providing feedback as the most pressing assessment issue.

Myers and Myers (2015) also found indication that teaching workload influenced teachers' choices when designing assessments. Based on a national survey among postsecondary teachers they found that teachers who taught more classes were less likely to use learner centred assessment, a term Myers and Myers use to describe "the assessment component of effective pedagogy" (p. 1905). However, Myers and Myers suggest that teachers see workload and learner centred assessment as a trade-off, which may explain why some teachers with higher teaching loads do employ learning centred assessments and others do not. This individual trade-off can be influenced by teachers' assessment perceptions.

3.1.3 Current study

Previous studies investigated assessment perceptions with assessment as a general construct. The current study aims to get insights into the perceptions of *intermediate* assessment. The main focus of this paper is teacher perceptions of intermediate assessment as these perceptions are crucial for implementing assessments in the first place. Student perceptions are also investigated, as intermediate assessment is implemented with student goals in mind. The

following two research questions were formulated:

- (1) What types of intermediate assessment are used in the programmes under investigation?
- (2) How are these types of intermediate assessment perceived by teachers and students?

3.2 Method

3.2.1 Context

The research was conducted at the Law School (Bachelor programmes of Law and Criminology) and the Faculty of Social Sciences (Bachelor programmes Psychology and the International Bachelor of Psychology, IBP) of a research university in the Netherlands. Every university programme in the Netherlands consists of 60 European Credits (EC; 1680 study hours) per year, in the current programmes divided into 5 and 10 EC courses (running for a half or a full semester). Psychology and Law programmes traditionally used to have a limited number of contact hours, mainly in the form of large-scale lectures, and a lot of time for independent study. Over the past few years, the focus has shifted towards more contact hours and intermediate assessments to keep students on track. These programs usually offer three concurrent courses, with one weekly lecture and one weekly or biweekly tutorial.

The programmes have obligatory intermediate assessments in the first-year courses, but the preconditions are very different. At the Law School, in courses with a study load of 10 EC, teachers are obliged to offer a partial exam, which counts toward the students' final grade. In 5 EC courses, teachers are free to choose whether they want to offer intermediate assessments. For both Psychology programmes, all students are graded on two different types of output: a final multiple-choice exam which makes up 70% of the course grade, and an intermediate assessment, usually taking place in the workgroup sessions, for the other 30%. Apart from these preconditions, set by the institutional boards, all teachers are free to find a type of intermediate assessment that fits their course. Students are obliged to complete the assessments that are offered in their courses.

3.2.2 Participants

Twenty-two first-year courses (eleven at each department) taught in the 2013-2014 academic year were selected for the study. Teachers were invited for interviews. To ensure that the interviews would have information about the lectures as well as the workgroups, for some courses pairs of teachers consisting of a lecturer and a workgroup teacher were invited. In total, 28 teachers were invited for interviews.

Eighteen teachers (10 males) representing fifteen courses accepted the invitation, ten teachers did not reply to the invitation or declined participation, because of time constraints, or because they felt they could not offer valuable insights. The average teaching experience of the participating teachers was 5.3 years (*SD* 3.6 years) of teaching the specific course they were interviewed on.

The number of full-time first-year students enrolled in the programmes was 830 (Law, including Fiscal and Notarial Law), 145 (Criminology), 480 (Psychology) and 120 (IBP). Information about student participation is included in the Data section.

3.2.3 Data

The university's electronic study guide (e-prospectus) provides an overview of all courses offered. Courses that used intermediate assessment were selected. For the selected courses, detailed information from the study guide was analysed to gain insight into course details and intermediate assessments.

3.2.3.1 Teachers' Perceptions

Interviews. Semi-structured interviews using a pre-determined topic list were conducted by the first author. Teachers were asked about their general opinion on intermediate assessment, the plans they had for intermediate assessment and their reasons for deciding on these plans. Furthermore, the interview focused on the teachers' experiences using intermediate assessment in their course and their evaluations of it. Teachers were also asked about student opinions on their course and about their ideas on the results from Student Evaluations of Teaching (SET, for more information see below).

Teacher interviews were conducted individually or in pairs, took place at their workplace and lasted between 30 minutes and one hour. All interviews were audiotaped and the interviewer took notes during the interview.

3.2.3.2 Students' Perceptions

Student Evaluations of Teaching (SET). To gain insight into the student opinions on the courses, relevant questions from Student Evaluations of Teaching (SET) were selected. The SETs were familiar to students and used the same terminology that students were used to hear in class. For Psychology SETs, three questions were chosen: 1) students rating their opinion on the course on a scale of 1 to 10 (with 1= very poor and 10= excellent); 2) students rating the workload of the course given the number of ECs on a 5-point scale (with 1= much too light, 3= in proportion and 5= much too heavy); and 3) students rating on a 5-point scale whether they agreed that the assignments, including the intermediate assessment, were instructive (with 1=disagree and 5= agree). The latter question was not obligatory and was therefore not answered in all SETs. For the Law School, two questions were selected: 1) students' general opinion of the course on a 5-point scale (with 1= inadequate and 5= very good), and 2) an open-ended question where students could give a more detailed account of their opinions.

All SETs were administered anonymously. SETs of the Psychology and Criminology programmes were administered as hard copy evaluations after the final exam, SETs of the Law programme were online evaluations. No courses were evaluated both online and in hard copy. The response rate on the online evaluations was much lower than on the hard copy equivalents. The response rates to SETs are included in Tables 3.2 and 3.3.

Interviews. To get a more detailed account of the SET scores, a small sample of Psychology students was recruited for an interview using the department's online research participation system (SONA System) and flyers in the department building. One female Psychology student was interviewed in June 2014 and six students (1 male) from the IBP were interviewed during the 2014-2015 academic year.

Semi-structured student interviews focused on how students perceived intermediate assessment in their courses. Students first gave their general opinion on intermediate assessment and discussed general study behaviour. Subsequently, they reported what the intermediate assessment in a course was, what they felt about the assessment and what they felt about the workload. Whereas teachers were interviewed about the one course they taught, students provided information about all courses they had taken up until that point. Student interviews were conducted individually on campus and took about half an hour.

3.2.3.3 Ethics

The current study was approved by the ethics committee of the Psychology Department. All students signed informed consent forms before they were interviewed.

3.2.4 Analysis

3.2.4.1 Interviews

All interviews were transcribed and structured using Atlas.ti 7 by the first author. Themes were used as a structuring tool to group quotes in general themes. We decided against a predetermined coding scheme to be able to stay close to the actual data, because of our small sample size. A preliminary set of five teacher interviews was structured inductively to get a feeling for the information in the interviews.

Subsequently, the first, second and fourth author decided on an open scheme consisting of general factors. After the general scheme with factors and preliminary themes had been set, all interviews were structured. New themes were created inductively during this process but fitting within the original factors. When five interviews had been structured using the new scheme, all the themes were reviewed critically by the first author and themes that overlapped were merged or redefined. The preliminary set of structured interviews was restructured as well.

The scheme consisted of four factors. The first factor was Content, where themes have a direct relation to the content of the studies. Example themes from the Content factor were “course” (all information directly related to the course, such as design, planning, exams, etc.) and “student” (information related to student opinions, behaviour, results, etc.) The second factor was Nature. The Nature factor applied to the type of information being discussed, such as “opinion” or “attitude”. The third factor was Judgement, which denoted whether the information discussed was “positive” or “negative”. The final factor was Not Relevant, where all information not directly related to the interview was grouped. Examples for this factor were “other programmes” and “general conversation”.

After structuring the teacher interviews, the scheme consisted of 56 themes, with 37 themes in the Content factor, seven in the Nature factor, six in the Judgment factor and six in the Not Relevant Factor as well. These factors and their specific themes were not mutually exclusive, that is, different factors could be combined in one quote, to indicate a students’ positive opinion, for example.

Student interviews were structured using the same scheme. To accommodate for the student experiences, three themes specific to the student situation were added to the Content factor. After structuring all interviews, therefore, the scheme consisted of a total of 59 themes, with 40 themes in the Content factor.

Quote length was determined by the content of the quote, as long as the same topic was being discussed by the participants and the interviewer it counted as one quote. As a result, quotes differed in length. All information in the interviews was structured.

To assess the replicability of the analyses the second author, who had been involved in designing the scheme but not in structuring the interviews, investigated two main themes from the results. First, he investigated one interview to see whether all problems with the assessment that were discussed by the teacher were in fact categorised as problems by the first author. Differences between the first and second author were discussed to further the quality of the scheme. Second, he explored all quotes that were discussing workload to see whether the results that were drawn from these quotes were an accurate representation of the opinions expressed by the teachers. He agreed with the conclusions drawn by the first author, and substantiated the conclusions with information from the quotes.

3.2.4.2 SET

To investigate differences in SET scores between the different assessment types, paired t-tests were run. These paired t-tests were run separately for students in the regular Psychology bachelor and the International Bachelor of Psychology, since students from both programs have different backgrounds. The Law and Criminology programs overlap in some courses, therefore it was not possible to run separate paired t-tests for these programs.

3.3 Results

3.3.1 Use of Intermediate Assessment

Table 3.1 indicates that there is a wide variety in types of intermediate assessment across the different courses. Furthermore, several teachers employed different types of assessment within their course. In the teacher interviews, two main reasons for this diversity in assessment methods were mentioned.

Table 3.1. Overview of First Year Courses that Use Intermediate Assessment

Course	EC	Final Exam	Intermediate Assessment	Details of Assessment	N	Feedback	Grade	Group
Law 1	10	MC questions	Partial exam, open ended ^a	Essay question based on a case study	1		20%	3
			Digital MC trial test	Course content	12	automated		
			Workgroup assignment ^a	Case study	12	personalised once		
Law 2	5	Essay questions	Experiment & Report ^a		1		pass/fail	2
			Proposition ^a		2		pass/fail	
			Duo presentation ^a		1		pass/fail	
Law 3	5	MC questions Essay question	Workgroup assignment ^a		1	peer feedback		1
Law 4	10	Open questions	Portfolio ^a	Nine written workgroup assignments e.g. APA writing, observation, essay	1		30%	2
		MC questions	Presentations	Portfolio assignments				
Law 5	5	MC questions Essay question(s)	Workgroup assignment ^a	Includes Case, Plea, Debate	6			2
Law 6	10	Essay questions	Partial exam, open ended ^a	Essay questions on first half of the course	1		50%	3
			Group presentation ^a		1		pass/fail	
Law 7	5	MC questions Essay question	Workgroup assignment ^a	Includes case, Essay,	7			2
Law 8	5	MC questions	Workgroup assignment ^a	Essay	1	peer feedback		2
		Essay question		Practice exam questions	4	plenary feedback		

Table 3.1. Overview of First Year Courses that Use Intermediate Assessment (continued)

Course	EC	Final Exam	Intermediate Assessment	Details of Assessment	N	Feedback	Grade	Group
Law 9	5	Open questions	Paper ^a	Conduct and report on an interview in pairs	1	peer feedback	30%	2
Law 10	5	Written exam	Workgroup assignment ^a		3	plenary feedback		1
Law 11	5	Final Paper	Written assignments ^a		5	peer feedback		1
Psychology 1	5	Self-reflection MC questions	Short essay test ^a		8	personalised once	30%	1
			Digital MC trial test		1			
Psychology 2	5	MC questions SPSS skills test	Digital MC trial test		1			4
Psychology 3	5	MC questions	Workgroup assignment ^a	Short essay, debate	4		30%	2
			Digital MC trial test		1			
Psychology 4	5	MC questions	Essay ^a		3	peer feedback		1
			Essay ^a		1		30%	
			Digital MC trial test		1			
Psychology 5	5	MC questions	Workgroup assignment ^a	Cases, MC questions	3		30%	2
				Poster	1	peer feedback		
			Digital MC trial test		1			
Psychology 6	5	MC questions	Workgroup assignment ^a	Presentation in Pairs, paper	2		30%	2
			Digital MC trial test		1			

Table 3.1. Overview of First Year Courses that Use Intermediate Assessment (continued)

Course	EC	Final Exam	Intermediate Assessment	Details of Assessment	N	Feed-back	Grade	Group
Psychology 7	5	Two choice questions	Behavioural change paper ^a Digital MC trial test				30%	2
Psychology 8	5	MC questions SPSS skills test	Digital MC trial test					4
Psychology 9	5	MC questions SPSS skills test	Digital MC trial test					4
Psychology 10	5	MC questions Essay questions	Practical assignments ^a	Analyse and interpret experiments Write research report on one experiment	4 1		pass/ fail pass/ fail	2
Psychology 11	5	MC questions	Test questions ^a		4		30%	1

Note. Law denotes the Law School; the Law 1 and 7 courses are part of the Bachelor of Law. The Law 2, 4,6, 9, 10 and 11 courses are part of the Bachelor of Criminology. The Law 3, 5 and 8 courses are part of both programmes. Psychology courses are the same for the IBP and the regular Bachelor Psychology. Courses in bold indicate teachers participated in an interview. N = frequency of an assessment.

^a Mandatory Intermediate Assessment

First, teachers mentioned that students need to learn a diverse set of skills in their course, and choosing diverse assessments is connected with that. Second, teachers reported that they wanted to experiment with different methods. Almost all teachers mentioned the relationship between the content of their course and the type of assessment as one of their main rationales for deciding on a specific type. The two teachers of Law course 5 described their decision for assessment methods as follows:

Teacher 1: Since we had no experience with more activating methods, we also thought, well let's try a few different things to see what works.

Teacher 2: Yes, and of course it is connected with the subject matter you're working on that week (Law 5)¹

In about half of the courses the intermediate assessment was new. Based on the assessment information as presented in the study guide, the courses were clustered into four types: 1) only written assignments as intermediate assessment; 2) a mix of different assessment types; 3) a partial exam, in combination with another form of intermediate assessment; 4) a voluntary trial exam.

3.3.2 Teacher Perceptions

The results can be grouped in three categories: 1) general opinion on intermediate assessment; 2) conceptual remarks to improve the intermediate assessment; and 3) practical difficulties with intermediate assessment. Teachers who taught courses with different assessment types did not display large differences in experience. For this reason, the results will be discussed for all types together.

3.3.2.1 General Opinion on Intermediate Assessment

All teachers reported positive perceptions of intermediate assessment. Several reasons were mentioned. Ten teachers reported that intermediate assessment encouraged students to actively engage with the course content during the course period, which they in turn hoped would lead to more active participation in class. For example:

¹ Quotes have been translated from Dutch and edited for length and legibility where applicable.

‘[...] the intermediate assessment is very good for different reasons, namely to force or encourage them to really start working on the subject-matter because the deadline of the assessment is earlier than the exam [...]’ (Law 9).

Furthermore, teachers in six courses mentioned the added value of being able to test at multiple occasions and to measure a variety of knowledge and skills.

One of the teachers in a course that had a partial exam was critical of the preconditions surrounding the partial exam in her course, but felt very positive about the written assignments that students handed in, which she felt led to more intelligent questions and participation in class. Another teacher mentioned that intermediate assessments can make a course more manageable for students and introduce them to the demands of the course.

Their positive perceptions notwithstanding, teachers also had quite a few criticisms. A few of these considerations were conceptual, others were of a more practical kind.

3.3.2.2 Conceptual Criticisms

The majority of the teachers did not mention that the intermediate assessment could lead to exam training or loss of student responsibility. Only three teachers argued that they believed the goal of intermediate assessment should not be to endlessly practise and make the courses too easy for students. Conversely, one teacher mentioned actually wanting to practise the skills for students even more.

Some teachers were critical of the preconditions that were set for the intermediate assessment. One of the teachers of a course with an obligatory partial exam reported: ‘That it’s obligatory, yeah I don’t think that’s strictly necessary’ (Law 6). Her opinion was shared by another teacher who was not obliged to use an intermediate assessment but was opposed to compulsory assessments. Other criticisms of the preconditions were related to the fact that students got two partial grades that they could use to compensate. The possibility for compensation may be especially problematic for courses where two very different sets of skills or knowledge were tested. One teacher said ‘I think that it actually should not be possible, because we want them to be able to do both and not one or the other by chance’ (Psychology 2). This teacher and one of his co-workers also felt that the distribution of 70% final exam and 30% intermediate assessment was giving too much weight to the latter.

3.3.2.3 Practical Considerations

Teachers mentioned keeping their own workload in mind while deciding on an assessment type. Both the Psychology and the Law programme have a large student body, whereas the student population of the Criminology programme is moderate in size.

The workload encountered by the teachers was different for each course. Teachers in courses with personalised feedback and grades reported a heavier burden of workload than the teachers of other courses. As Table 3.1 shows, three courses did not have graded assignments or individualised feedback, so the only time investment for the teachers was to check if students had completed the assessment. However, in one of the Law courses one teacher was responsible for grading 150 portfolios, a massive workload. Two teachers mentioned the burden of the workload but also reported that it was worth it.

In one course with graded assignments, the teacher mainly reflected upon the workload involved in setting up the whole system of assessment and grading. However, he did comment that '[...] of course that's done now, it was a one-time investment' (Psychology 1). Five courses used peer feedback to alleviate teacher workload.

In addition to the workload directly related to the assessment, teachers acknowledged that there was an increased workload in administration and communication. One example of this was dealing with students who missed an assessment, especially since missing too many assessments often led to exclusion from the final exam.

Five teachers discussed plagiarism and their solutions to prevent this. Because of the large student body, the work group meetings of the courses were dispersed over the week. This meant that students in a Monday work group could let their peers in a Friday work group know what the assessment questions would be. Teachers tried to overcome this problem in several ways. In three courses the deadline for completing the assessment was the same regardless of the timing of the work group, which sometimes led to criticism from students, for example because the deadline was before the lecture discussing the topic of the assignment. Another teacher invested a lot of time in designing different questions for each work group. Yet another teacher made all assignment questions available beforehand but found that students posted the correct answers on social media, which he hoped to prevent next year by increasing the number of possible questions.

Another practical consideration that came up during three of the interviews was re-sits of the intermediate assessment. According to institution guidelines, students had the right to re-sit a failed exam, but in one course the intermediate assessment did not count separately. If a student failed the final exam they had to retake the intermediate and final assessment as one set, even if they had passed the intermediate assessment. Additionally, in two courses the re-sit had a different form from the intermediate exam, which teachers feared might lead to strategic thinking on the part of students:

‘But then they can decide [...] instead of going to a work group eight times, I’ll just prepare the book, take two exams, one multiple choice, one essay exam, because that way I can pass as well’ (Psychology 1).

When asked if they wanted to change things in the assessment, teachers mainly addressed fine-tuning the current assessment practice and not reshaping it entirely. Some teachers wanted to eliminate specific aspects they felt did not work as planned. One teacher specifically mentioned that it would be inadvisable to change the intermediate assessment right away: ‘Look, we need to consider this. You shouldn’t immediately overturn things. So we need to continue this structure’ (Psychology 6).

3.3.3 Student Perceptions

3.3.3.1 Student Evaluations of Teaching

Tables 3.2 (Psychology) and 3.3 (Law and Criminology) show the SET results of the two different departments. Results will be discussed per department.

Table 3.2. Student Evaluation of Teaching Results for Psychology Courses Divided in Groups

Course (N ^R ;N ^{IBP})	Assign-ments		Work-load		Course	
	Regular	IBP	Regular	IBP	Regular	IBP
Group 1						
Psychology 1 (523; 91)			3.6 (0.8)	4.0 (0.8)	7.6 (1.0)	8.2 (1.1)
Psychology 4 (449; 71)	3.2 (1.1)	3.1 (1.1)	3.7 (0.7)	3.8 (0.8)	6.8 (1.2)	7.2 (1.2)
Psychology 11 (458; 23)	4.0 (1.0)	4.0 (0.9)	3.3 (0.6)	3.4 (0.6)	6.9 (1.4)	6.8 (0.8)
Group 2						
Psychology 3 (495; 95)	2.5 (1.1)	3.3 (1.1)	3.5 (0.8)	3.5 (0.7)	4.8 (1.8)	7.4 (1.4)
Psychology 5 (451; 72)	3.5 (1.2)	3.7 (1.2)	3.6 (0.7)	3.4 (0.7)	7.4 (1.2)	8.1 (1.2)
Psychology 6 (443; 79)	3.1 (1.2)	3.8 (1.0)	3.3 (0.7)	3.5 (0.7)	7.3 (1.0)	7.9 (1.2)
Psychology 7 (485; 72)			3.4 (0.6)	3.4 (0.8)	6.4 (1.4)	5.7 (1.9)
Psychology 10 (427; 63)	3.2 (1.0)	3.8 (0.8)	3.5 (0.7)	3.6 (0.8)	6.5 (1.3)	6.6 (1.3)
Group 4						
Psychology 2 (506; 91)			3.1 (0.6)	3.2 (0.7)	7.7 (0.9)	7.9 (1.1)
Psychology 8 (506; 84)			3.6 (0.7)	3.6 (0.7)	7.1 (1.4)	7.6 (1.4)
Psychology 9 (449; 72)			3.4 (0.6)	3.5 (0.7)	7.4 (1.0)	7.3 (1.6)

Note. N^R = response rate in the regular bachelor, N^{IBP} = response rate in IBP.

Results from the paired samples t-tests comparing assessment types, indicate that students did not evaluate courses differently depending on the type of intermediate assessments administered (all p 's > .05). This was apparent in the regular Bachelor of Psychology as well as the IBP. Independent of the assessment types, students rated most courses as 6 (out of 10) or higher, indicating a moderately positive evaluation.

At the Law School, paired t-tests also showed no differences between the three types of assessment (all p 's > .05), again implying that courses were not evaluated differently solely based on the type of intermediate assessment that was used. As can be seen in Table 3.3, all courses except for one were rated 3.5 (out of 5) or higher, indicating a generally positive evaluation.

Table 3.3. Student Evaluation of Teaching Results for Law and Criminology Courses Divided in Groups

Course (N)	Course	
	Law	Criminology
Group 1		
Law 3 (255)	3.5 (0.8) ^a	3.5 (0.8) ^a
Law 10 (97)		3.8 (0.6)
Law 11 (105)		3.9 (0.5)
Group 2		
Law 2 (124)		4.1 (0.5)
Law 4 (126)		3.7 (0.7)
Law 5 (150)	3.7 (0.8) ^a	3.7 (0.8) ^a
Law 7	n.a.	
Law 8 (147)	3.8 (0.7) ^a	3.8 (0.7) ^a
Law 9 (92)		2.9 (0.7)
Group 3		
Law 1 (200)	3.9 (0.7) ^a	
Law 6 (122;117)		3.8 (0.7); 3.8 (0.8)

Note. No student evaluations of teaching were administered for the Law 7 course in the 2013-2014 academic year. Law 6 was a two-part course and both parts were evaluated separately.

Courses Law 3, Law 5, and Law 8 are part of the Law program as well as the Criminology program, and therefore students from both bachelor programs answer the same SET within these courses.

^a denotes online SETs, all others are hard copy

3.3.3.2 *Students’ Detailed Accounts on SET*

The Psychology students provided three explanations in the interviews which flesh out their generally positive evaluations of intermediate assessment. First, just like teachers, students valued the fact that the intermediate assessment kept them on track. Second, they believed that completing the intermediate assessment improved their subject knowledge as shown in the following quotation:

‘It kind of made you want to do the material and you really learned it before and then you could read it again afterwards and then it was kind of stuck in your brain and that was good’ (IBP student 1).

Third, students mentioned that they liked the fact that the intermediate assessment lowered the stakes of the final exam, since their course grade was no longer totally dependent on their final exam results.

One conceptual criticism students mentioned was that they preferred intermediate assessments that had a clear connection with the final exam. The following quotation illustrates this:

‘Those essays were pretty annoying because they had something to do with the book, but not really. And it was nothing not really about the exam or about the lectures’ (IBP student 2).

They did not endorse the problems with exam training that some teachers reported, that is, no students mentioned that there could be such a thing as too much practice.

Like the teachers, students mentioned that a major practical problem with intermediate assessment was the workload. Most of them mentioned it being heavy, especially when they were just starting their studies. This finding aligns with the SET scores on workload as shown in Table 3.2 (with all mean scores above 3.0):

‘This was massive in the beginning [...] you try to really know them by heart, [...] and it’s insane in the beginning and I was like the walking dead after this’ (IBP student 3).

Some students did, however, also mention the workload seeming lighter for courses that they enjoyed and the work being ‘tough but useful’ (IBP student 4). One major problem that is related to the workload was competition with other study activities. A few students mentioned rushing through the course work because they did not have enough time, or deciding not to complete non-obligatory work, like some readings and homework, because they were spending all their time on the intermediate assessments.

Students’ detailed accounts on the SET indicated that Law students felt that the intermediate assessment was a good way to familiarise themselves with the demands of a course. One student commented that ‘The partial exam is a useful tool to see whether you understood the first part correctly’. This desire to use the assessment to practise came up in several detailed accounts in the SET.

Furthermore, several comments suggest that students wanted more personalised feedback, also indicating that they were looking for pointers on how to fulfil their assignments correctly.

Like the Psychology students, the Law students' main conceptual criticism was that intermediate assessment was often unrelated to the final exam. One comment was that 'the paper had little to do with the material'.

On the more practical side, the Law students rarely complained about the workload of the different assessments. However, in one course students protested that they had to hand in their assessment at the beginning of the week, before they had the opportunity to follow lectures and workgroups and enhance their understanding of the material.

3.4 Discussion and Conclusion

In this research, we investigated how intermediate assessment is perceived by teachers and students. Their perceptions were clustered into 1) general positive opinions, which mainly concerned the fact that students kept up with study work; 2) conceptual criticisms, where teachers and students voiced their critiques on the preconditions and content of the intermediate assessments, and 3) practical considerations, where the main issue discussed was workload.

One positive finding of the current study was that both teachers and students value intermediate assessment because it guided study behaviour and made students study more. Some students mentioned that they would not have kept up with the coursework had the intermediate assessment not been there. Students in the paper by Harland et al. (2015) also feared that they would not make effective use of independent study time and therefore appreciated frequent assessments to serve as a guide for studying.

The relation between student time-use and student results has been extensively studied (e.g. Admiraal et al., 1999; Doumen et al., 2014; Michaels & Miethe, 1989). However, the results have not been straightforward. Doumen et al. (2014) for example, found that the influence of time use on course grade was course dependent, whereas Michaels and Miethe (1989) observed that, for the undergraduate students in their sample, study time correlated positively with results for students who studied during the week, but not for students who crammed at the last minute. Admiraal et al. (1999) concluded that increased time on task led to better study results for their sample.

In the current situation, intermediate assessments could be seen as an incentive to study, but several teachers argued that university students are adults and should take ownership of their study process and their learning. Harland et al. (2015) raise the question of who is responsible for student learning and refer to the system of frequent assessment as a pedagogy of control. With regard to the question of responsibility, Van Berkel et al. (2012) note that universities should take measures to optimise the feasibility of their programmes and that student responsibility comes into play only within these optimised programmes. One measure they suggested to increase feasibility is encouraging students to study by assessing them regularly (Cohen-Schotanus, 1999), which is the measure taken by the programmes under investigation.

The favourable attitudes towards intermediate assessment are, however, characterised by one remarkable difference in opinion between teachers and students. Teachers felt that intermediate assessments allowed them to test a broad range of knowledge and skills, but students preferred assessments that measured similar knowledge and skills to those in the final exam. This student preference for overlap is partially supported by testing effect studies that indicate that an overlap in questions is necessary for favourable final exam results (Kling et al. 2005; McDaniel, Wildman, & Anderson, 2012). Interestingly, even though students reported that being tested on the same content twice improved their learning, survey studies with American college students (Karpicke, Butler, & Roediger, 2009) and Dutch secondary school students (Dirkx, 2014) have shown that students often do not use self-testing as a strategy while studying. Apparently, students need to be prompted to perform this kind of intermediate testing.

The difference in preference for content overlap between teachers and students can possibly be explained by the fact that teachers and students have different goals. As mentioned above, teachers' goals were to measure a broad set of knowledge and skills, whereas students' goals were to pass the exam. A few teachers voiced their concerns that the intermediate assessments could lead to exam training and teaching to the test. Teaching to the test is a type of teaching that focuses on passing the final exam, instead of on furthering the general knowledge of the curriculum content (Popham, 2001). To avoid teaching to the test our teachers' assessment goals were more focused on assessing different skills and knowledge, instead of simply improving the knowledge of one specific subdomain. Jessop, El Hakim, and Gibbs (2014) noted that assessment goals are

often not clear to students, which might also have been a factor in the current discrepancy in preference.

A more practical issue that both teachers and students faced was the workload that is associated with intermediate assessment. For students this workload seemed to lead to competition between the different course activities. Non-mandatory study activities got pushed aside for intermediate assessments, especially if they were graded. Vos (1992) had already suggested that competition may be an issue for students, if programmes do not keep student workload in mind while designing their curriculum. In their interview study, Harland et al. (2015) also encountered strong evidence for competition, where students walked out of a non-graded lecture because they would rather spend time on their assessment. One possible solution to this problem is to take a more programmatic approach towards assessment (van der Vleuten et al. 2012). Programmatic assessment involves a more comprehensive approach, where several low-stake formative assessments build up to a high-stake summative decision. When the assessment programme is designed as a whole instead of as different components, teachers are more aware of the possible workload faced by students.

The fact that teachers mentioned their workload is in concordance with the results of Goos et al. (2011), who found that teachers' main concerns were regarding assessment conditions. With regard to teacher workload, it is important to keep the assessment goals in mind. Harland et al. (2015) encountered the problem that teachers felt there was no time to provide feedback because of the large number of assessments. Leeming (2002) also noted that frequent assessment comes with a sizeable workload. However, he suggests adapting the type of assessment to reduce workload; for example, by reducing the number of essay questions instead of the number of assessments.

3.4.1 Limitations

The sample of teachers, courses, and assessment types is relatively small and selected from four programs in one university in the Netherlands. Furthermore, all teachers that were interviewed during this study were part of a department-wide educational innovation project. They were actively engaged in improving their courses during this project. The fact that the teachers were participating in an innovation project could have had an influence on their motivation and their perceptions of intermediate assessment.

Teachers in the context of Law, Criminology and Psychology programs undergoing educational innovation projects are only a small facet of the total population of university teachers and therefore generalising to the general population of university teachers is inadvisable. We do think our findings have generated plausible hypotheses to be tested in future research. Future research could expand the research population in two directions. First, they could include teachers from different departments and other universities. Second, teachers who are not a part of an educational innovation should be investigated about their perceptions.

3.4.2 Implications for practice

The limited sample size in the current papers precludes bold implications, but we believe there are three general indications for practice that can follow from our results and the literature discussed in this section. First, students need to be made more aware of the goals of intermediate assessment, as better awareness would probably add to their positive perception of it. Second, teachers and students need to be made explicitly aware of the cognitive benefits of intermediate assessment on top of behavioural effects, like keeping on track. The third implication is that measures should be taken to alleviate teacher and student workload. One of these measures could be to take a programmatic approach towards assessment, to spread out study activities and reduce competition between parallel courses.

3.4.3 Concluding remarks

University education is subject to change and curriculum innovations often follow each other in quick succession. These innovations are usually evaluated by looking at grades, pass percentages, or by using questionnaires. However, teachers and students need to support the innovation for any of these measures to be useful. The results from this study suggest that teachers and students at the four programs under investigation generally have positive perceptions of intermediate assessment, along with some conceptual criticisms and with concerns about practical considerations. The assessment type of a course does not seem to play a role in these perceptions, but student and teacher characteristics might. Future research should take these characteristics into account. Research by Baeten, Dochy and Struyven (2012) has already indicated that student perceptions of their learning environment are influenced by characteristics such as their motivation.

Chapter 4

**University teachers' conceptions of their current
and ideal intermediate assessment:**

An A+ is good, but speaking your mind is better

Assessment in higher education with a transformational instead of a reproductive purpose can be a powerful way of supporting student learning. Since university teachers usually design their own assessments, it is important to investigate their conceptions of assessment. The current study focuses on teachers' conceptions of their current and ideal assessment with a focus on intermediate assessment. Thirteen teachers teaching law, psychology and criminology, reflected on their current and ideal assessment in an attempt to eliminate the influence of practical constraints on assessment practice. Results indicate that the majority of teachers have transformational conceptions of their intermediate assessment practice, and, in general, their conceptions of the ideal assessment are even more transformational. This suggests that teachers' main focus for assessment is on student learning and that a lack of transformational assessments in practice may be mainly caused by external constraints.

This chapter has been submitted in adapted form as: Day, I. N. Z., van Blankenstein, F. M., Westenberg, P. M., & Admiraal, W. F. University teachers' conceptions of their current and ideal intermediate assessment: An A+ is good, but speaking your mind is better

4.1 Introduction

Assessment is used to measure, guide, focus and enhance student learning in higher education. Several researchers have investigated what assessments optimally support student learning. An oft cited example are Gibbs and Simpson (2004), who proposed ten conditions that assessment in higher education should meet to support student learning, including that the assessment tasks should encourage students to engage with learning materials and that feedback should be provided in such a time that students can still attend to it. Other researchers focus on formative, as opposed to summative, assessment as a way to support student learning.

The distinction between the summative and formative roles of assessment is critically reviewed by Bennett (2011). The author suggests a more nuanced view of the relationship between assessment purpose (assessment *of* learning and assessment *for* learning) and assessment type (summative or formative), distinguishing between primary and secondary purposes. Summative assessments primarily document what students know and can do, but secondarily can support students' learning directly or indirectly by modification of teacher instruction. Similarly, formative assessment's primary focus is to support students' learning and inform teacher instruction, but secondarily also exhibits what students know and can do. Formative assessment can be defined as follows: 'Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited' (Black & Wiliam 2009, p. 9). Literature reviews (e.g., Black & Wiliam 1998; Kluger & DeNisi 1996) and meta-analyses (e.g., Kingston & Nash 2011) suggest that well-designed formative assessment can cause achievement gains, dependent on student characteristics and how formative assessment has been carried out.

Assessment can measure a variety of knowledge and skills, which are often classified by using the two dimensions of Bloom's revised Taxonomy (Kratwohl 2002). The first dimension focuses on the kind of knowledge that is assessed, and the second dimension focuses on six cognitive processes associated with the assessment. These cognitive processes can be divided in reproductive processes (remember, understand, and apply) and transformational processes (analyse, evaluate, and create). In general, these processes are viewed in a hierarchical way

(Kratwohl 2002), and assessment using transformational processes is deemed 'better'.

In general, the content of higher education curricula is not very constrained. This allows teachers to decide which specifics they want to teach and how they want to assess their students. Subsequently, teachers' ideas about assessment play a large role in how assessment in higher education is designed. Several authors have investigated higher education teachers' ideas about assessment (Postareff et al. 2010; Reimann and Sadler 2017; Sadler and Reimann 2018; Samuelowicz and Bain 2002; Watkins, Dahlin, and Ekholm 2005). Three of these studies constructed categorisation schemes to classify these assessment ideas.

First, Samuelowicz and Bain (2002) conducted interviews with twenty Australian academics. Interview questions focused on specific exam questions brought in by the interviewees, as well as on assessment in general. The interviews resulted in six assessment belief dimensions. These dimensions focused on 1) the nature and structure of knowledge, 2) the degree of integration of knowledge, 3) the degree of transformation of knowledge, 4) the differences between good and poor answers, 5) the role of assessment in teaching and learning, and 6) the use of feedback gained from assessment. On each dimension, a belief could range from knowledge reproduction to knowledge construction or transformation. Subsequently, these belief dimensions were used to categorise assessment practice orientations. Examples of these orientations are 'reproducing bits of knowledge', where teachers' beliefs all were focused on reproduction, and 'transforming conceptions of the discipline/world', where all beliefs focused on knowledge construction or transformation. Teachers' orientations to assessment were found to be highly correlated with their orientation towards teaching and learning.

Second, Watkins et al. (2005) interviewed 26 university teachers from Hong Kong and twenty from Sweden, to investigate teachers' conceptions of the backwash effect (the way in which you assess influences the way students learn) and how teaching and assessment relate to each other. They constructed eight categories, which could then be grouped in three groups. The first three categories focused on the content that students learned, the second group consisted of three categories that focused on learning processes but still placed importance on basic knowledge, and the final two conceptions could be put in a group where basic knowledge played no role, but deeper learning strategies did. These groups partly overlap with dimension five of Samuelowicz and Bain (2002), focusing on the role of assessment in teaching and learning.

Third, Postareff et al. (2012) interviewed 28 Finnish university teachers. Interview questions focused on what teachers think is important in assessment, the purpose of assessment, and how teachers assessed student learning. Interview results were categorised in five groups representing assessment purposes. These purposes were 1) measure repetition and memorisation of facts, 2) measure how well students cover the contents of the study module, 3) measure application of knowledge, 4) measure deep understanding and students' own thinking, and 5) measure the process and development of students' own thinking. The first three purposes indicate reproductive conceptions and the last two transformational conceptions. The majority of teachers displayed reproductive assessment conceptions. Furthermore, teachers' assessment conceptions and assessment practices were consistently aligned, indicating that students often have to complete traditional assessments, focused on reproducing knowledge.

Myers and Myers (2015) found that teachers' assessment practices were constrained by their workload, evidenced by the fact that teachers who taught more classes often used less learner centred assessments. Additionally, Goos, Gannaway and Hughes (2011) found that teachers' challenges when using assessment focused on conditions like workload and bureaucracy as well. When assessment practices are constrained by outside factors, teachers are not given the opportunity to translate their assessment conceptions into practice.

4.1.1 The current study

The aim of this study is to increase insights in higher education teachers' conceptions of assessment. Whereas previous studies focused on teachers' assessment conceptions in general, our study specifically focuses on the conceptions of intermediate assessment.

Intermediate assessment, also known as continuous or frequent assessment, focuses on the assessments that take place during the course period, as opposed to end-of-term assessments. Gibbs and Simpson (2004) mention that frequent assessments with formative feedback were 'central to student learning' (p. 8). Intermediate assessments can have various types, and because of their placement during the course period they often address different goals than end of term assessments do. An end-of-term assessment, for example, may address whether students show sufficient knowledge of the course material, but an intermediate assessment can focus more on the development of this knowledge by, for example eliciting study time and providing feedback at a moment that students

can still use this feedback in their learning (two of the ten conditions by Gibbs & Simpson 2004) Because of this difference in assessment goals, it stands to reason that teachers' assessment conceptions may differ as well. As with other forms of assessment, teachers' ideas of their current practice might be different from their ideal, as external constraints might inhibit teachers to put their ideas into practice. Therefore, we ask teachers to reflect on their practice as well as their ideals. The following research question guided our research:

'What differences in conceptions of intermediate assessment do university teachers display when discussing their current and ideal intermediate assessment?'

4.2 Method

4.2.1 Recruitment

The university's digital study guide was used to gain an overview of the first-year law, criminology or psychology courses that used intermediate assessment. A total of seventeen courses qualified, and course coordinators of these courses were invited for an interview by email. This email provided a short overview of the interview topic, and the themes that would be discussed. If a teacher had not responded after two weeks, a reminder email was sent.

4.2.2 Participants

All seventeen invited course coordinators were involved in teaching the courses. Three course coordinators declined participation and two did not respond to the invitation. Of the participating teachers, five taught psychology courses (teachers 2,3,9,10, and 12), four taught criminology courses (teachers 1,4,6, and 7), one taught a law course (teacher 11) and two taught a course that was part of both the law and criminology curriculum (teachers 5a and 8). Teacher 5a asked to be interviewed with a co-worker because he had only recently taken over the coordinating duties. Ultimately, thirteen teachers (seven female) were interviewed during twelve interviews.

With regard to the teaching experience of the participants, eight teachers were involved in teaching or coordinating the courses under investigation since the curriculum of the programs was revised in the 2013-2014 academic year and two were first time teachers of this current course. The remaining three teachers had taught the course under investigation once or twice before. Most

teachers expressed experience with teaching courses beyond the course under investigation, or with teaching the course before the curricular revision.

4.2.3 Interviews

All semi-structured interviews were conducted by the first author during December 2016 and January 2017. All interviews followed the same topic list. The first questions were about teachers' current intermediate assessment practice, and their goals and experiences. Subsequently, teachers reported what their ideal intermediate assessment would be, focusing on a 'perfect world' without practical constraints. Finally, the interviewer presented a hand-out with a short overview of intermediate assessment in the literature (preliminary results from the literature review described in Chapter 2). Using the information on this hand-out, teachers further reflected on their ideas for ideal assessment.

Teachers were interviewed in their office, all interviews took between 30 and 60 minutes. One teacher was abroad as a visiting scholar and was interviewed using Skype™. All interviews were audiotaped and annotated, and the interview transcripts were presented to teachers for a member check. Nine teachers greenlighted the transcript, and two wanted to have further insight in how their quotes were going to be used in the final article, but had no objection to their portrayal. Two teachers were no longer working at our university during the member check period.

4.2.4 Analysis

4.2.4.1 Coding Scheme

Previous research investigating assessment conceptions using bottom up coding techniques has found very comparable categories of conceptions of general assessment (Samuelowicz and Bain 2002; Watkins et al. 2005; Postareff et al. 2012). To align our investigation of intermediate assessment conceptions with the existing literature, our coding scheme consisted of the five assessment conceptions devised by Postareff et al. (2012) and two of the assessment belief dimensions from Samuelowicz and Bain (2002), specifically dimensions five (regarding the role of assessment) and six (regarding the use of feedback).

Two codes were used to distinguish between conceptions of the current and ideal intermediate assessment. An additional code was added to indicate when teachers expressed negative ideas about some conceptions, for example to say that intermediate assessment should not just be exam training (negative about

purpose of assessment R1; reproduction of knowledge). A full overview of the codes used can be found in Table 4.1.

Table 4.1. Overview of the codes used in the analysis of the interviews.

	Purpose of assessment is to measure (Postareff et al. 2012 87)	Role of assessment in teaching and learning (Samuelowicz & Bain 2002 182-183)	Use of feedback gained from assessment (Samuelowicz & Bain 2002 182-183)	Other Codes
Reproduction	R1. Repetition and memorisation of facts	R1. Academic believes students have to be forced to study and assessment is believed to be the best tool to achieve this	R1. Academic believes that feedback from student performance should be used to alter his/her teaching	Current intermediate assessment
	R2. How well students cover the contents of the study module	R2. Academic believes that assessment forces students to study, and that marks give them an indication of the progress made and rewards their efforts	R2. Academic believes that feedback from student performance should be used to change the academic's or students' actions	Ideal intermediate assessment
	R3. Application of knowledge		R3. Academic believes that feedback from student performance should be used to monitor students' learning and help them improve	Negative modifier

Table 4.1. Overview of the codes used in the analysis of the interviews (continued).

	Purpose of assessment is to measure (Postareff et al. 2012 87)	Role of assessment in teaching and learning (Samuelowicz & Bain 2002 182-183)	Use of feedback gained from assessment (Samuelowicz & Bain 2002 182-183)	Other Codes
Transformation	T4.	T3.	T4.	
	Deep understanding and students' own thinking	Academic believes assessment to be an integral part of teaching and learning, a means of helping students learn.	Academic believes that feedback from student performance should be used to challenge students' existing ideas and understandings.	
	T5.			
	The process and development of students' own thinking			

4.2.4.2 Analysis procedure

The interview transcripts were analysed by the first author. All transcripts were read, and information pertaining to the current assessment or the ideal assessment was highlighted in two different colours. Subsequently the transcript files were uploaded in to ATLAS.ti 7 for analysis. Units of analysis were decided by speech turns, where uninterrupted speech by a teacher was taken as a singular unit.

The analysis procedure was conducted in several iterations. In the first iteration, the relevant interview sections were coded as either *current assessment* or *ideal assessment*. Sections of the text that were not dealing with either two assessments were not included in the final analysis. The second iteration consisted of applying Postareff et al.'s (2012) five conceptions for the purpose of assessment to the statements relating to either assessment. Multiple conceptions could apply to a single speech turn. In the third iteration of analysis, the two assessment belief dimensions of Samuelowicz and Bain (2002) were used to classify the statements.

In all interviews, there were sections, for example descriptions of the design of the assessment, that did pertain to the current or ideal assessment, but could not be classified under the assessment conceptions and beliefs (Postareff et al. 2012; Samuelowicz & Bain 2002, respectively). These sections are used as context in the case descriptions.

After analysis, frequencies of different codes were tallied per assessment mode, and bar charts for each teacher's conceptions for the current and the ideal assessment were constructed as a visual guide. Since, to our knowledge, our study is the first focusing specifically on intermediate assessment, we decided to portray the whole variety of teachers' conceptions and beliefs instead of classifying them in to one group according to their highest order or most uttered conception or belief.

4.2.4.3 Reliability of the analysis

To ensure the reliability of the coding, the second author conducted an audit on the coded transcripts. During this audit he critically assessed each transcript with the coding scheme. Afterwards the first and second authors discussed the outcomes and adapted the codes where necessary. Across all twelve interviews the second authors' comments led to thirteen added and eight deleted codes. Furthermore, in sixteen cases the authors agreed that the original coding was sufficient. Ultimately, across twelve interviews there were 653 codes, with 208 relating to teachers' conceptions and beliefs for the ideal and current assessment.

4.3 Results

The current assessments utilised by teachers can roughly be classified in six groups. Teachers used presentations (10), preparatory workgroup assignments (5ab, 6), short written assignments (3, 8, 9), research reports (1, 7, 12), partial exams (4, 11) or multiple-choice questions (2). When proposing ideal assessments, teachers often borrowed ideas from their own experiences as a student, or from other undergraduate or graduate courses they taught. These ideal assessments could also be clustered in six groups. Four teachers wanted to use one or more essays, two wanted to use preparatory workgroup assignments, two teachers wanted to integrate assessment into teaching by continuously assessing, and one wanted to use audience response systems to assess student

learning during lectures, one wanted students to be involved in the research process. Three teachers wanted to use what could be referred to as classroom assessment, inspired by imagery of Ivy League universities:

But compare it to Harvard. Look at how they do it, like you see in the movies. Students know, you bring your books and the class is not that big, not small but not that big. You bring your books and the teacher asks [a question]. When you're lost for words, you are [ridiculed], so no one says 'Um...' (Teacher 9)²

In general, teachers have strong transformational conceptions for the current and ideal assessment, as can be seen in Figure 4.1. However, there are large individual differences between teachers. Based on their conceptions about the purpose of their current and ideal assessments (Postareff et al. 2012) the teachers were divided in two main groups. The first group consisted of teachers that displayed no difference in conceptions between their current and ideal assessment. The second group represented teachers who do portray different conceptions between the current and ideal assessment. Within these two main groups, five subgroups can be discerned. These subgroups, with a short summary of conceptions are presented in Table 4.2 and further discussed in the following sections. To reiterate, reproductive conceptions focus on memorisation, covering content and application of knowledge, whereas transformational conceptions focus on deep understanding and the development of understanding, consult Table 4.1 for a full list of codes related to reproductive and transformational conceptions.

² All quotations have been translated from Dutch and edited for legibility where necessary.

Table 4.2. Overview different subgroups of teachers and a summary of conceptions and beliefs within each subgroup.

No difference in conception			Difference in conception	
No transformational conceptions (N = 1)	A mix of conceptions (N = 3)	Only transformational conceptions (N = 1)	Ideal is more transformational (N = 7)	Current is more transformational (N = 1)
Class participation is better when students know the material. Can students apply material that they can reproduce?	Assessment should help students practice their skills. Give feedback by discussing written work in detail. Students should know all the material.	Student knowledge should be developed in small scale classes with enough supervision.	Students should learn through direct feedback. Actively working with the material presents the students with a chance for knowledge development. Learning curves are important.	Application of knowledge is important. Give feedback by discussing written work in detail.

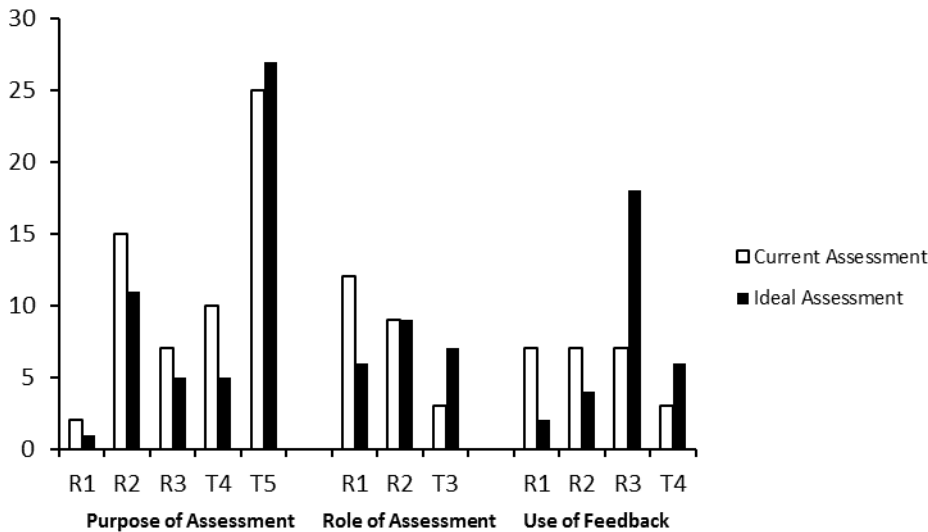


Figure 4.1. Overview of the total number of quotes discussing assessment conceptions and beliefs of teachers. See Table 4.1 for an extensive description of each code.

4.3.1 Teachers who display no difference between the current and ideal assessment

4.3.1.1 No transformational conceptions

Figure 4.2 shows the teacher (5a) that mentioned no transformational conceptions for the current or the ideal assessment. He did touch on several reproductive conceptions for his ideal assessment, which were workgroup assignments. His beliefs about the role of an assessment focused on the fact that assessment forces students to study and that students could be rewarded by extra exam questions, for example. However, simultaneously he underscored the fact that '[...] it should come from students themselves. It frustrates me that it is needed to do it this way, and that it only happens when we do' (Teacher 5a). This that he would prefer that intermediate assessment was not needed, but he conceded that it may be necessary to have students prepare for class. Regarding the use of feedback, he briefly touched upon the fact that feedback should mainly be used to monitor learning and adjust the learning process where necessary.

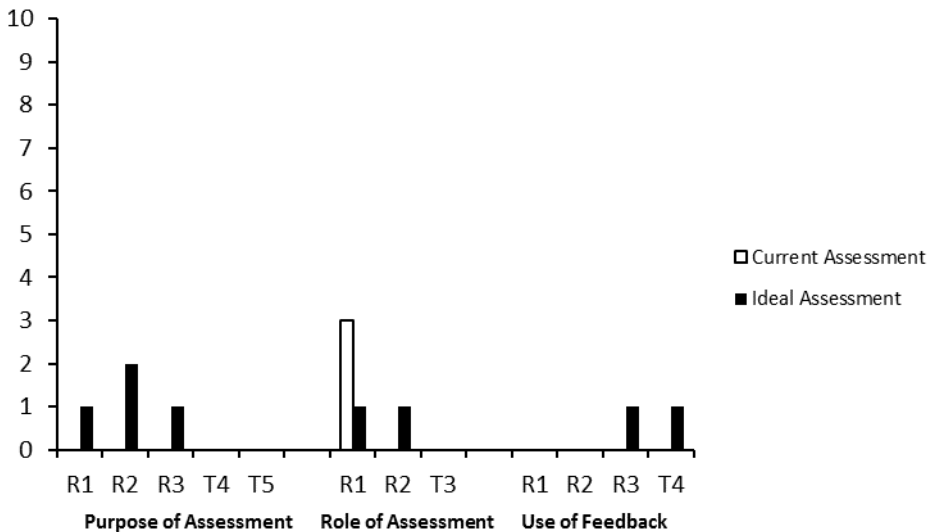


Figure 4.2. Number of quotes discussing assessment conceptions and beliefs of teacher 5a. See Table 4.1 for an extensive description of each code.

4.3.1.2 A mix of reproductive and transformational conceptions

Three teachers (4, 6, 9) displayed a mix of reproductive and transformational conceptions for both current and ideal assessment. Two of these teachers (4 and 9) mentioned the ‘Harvard’ type of classroom assessment as one of their ideal assessments, in addition to essays (both) and using different methods of assessment to conclude a portion of the subject matter every week (teacher 4). Both teachers display a range of conceptions for the ideal assessment. When looking at their beliefs, teacher 4 did not discuss the role of assessment, but he believed feedback for ideal assessments should be used to monitor learning and challenge misconceptions, as can be seen in Figure 4.3. ‘I try to work with assessment forms that have a reasonable amount of detail, but it is good to provide opportunities to ask questions outside of that [...] sometimes written feedback raises new questions’ (teacher 4). Conversely, the role of feedback did not come up in conversation with teacher 9, but Figure 4.4 shows that his belief about the role of his ideal assessment is that it is an integral part of education in addition to a potent measure to get students to study. This is also apparent in his quote on page 10, where he stresses the importance of preparing for class.

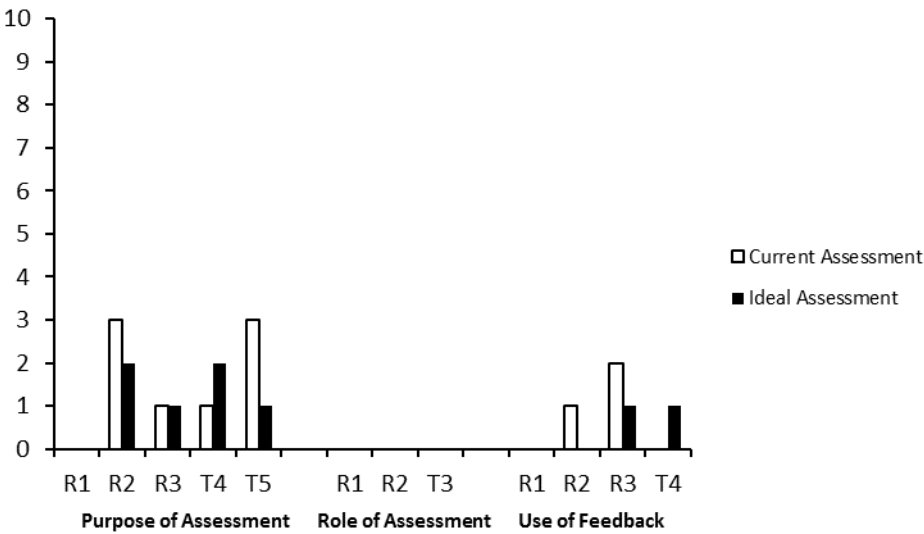


Figure 4.3. Number of quotes discussing assessment conceptions and beliefs of teacher 4. See Table 4.1 for an extensive description of each code.

The ideal assessment of teacher 6 is continuous assessment through practice material, since her course (a methodology and statistics course) lends itself very well for this. 'And in terms of a skill development track, what they learned with me they can practice before they start [the second-year methodology course], to refresh their knowledge' (Teacher 6). Additionally, she would like to make more use of digital tools to facilitate this practice material. This continuous assessment goal is also reflected in her assessment beliefs, where assessment is an integral part of education and learning and feedback should be used to improve student learning, see Figure 4.5.

4.3.1.3 Only transformational conceptions

One teacher (teacher 7) displayed only transformational conceptions for both assessment types. As can be seen in Figure 4.6, the main focus of both assessments is different though. For the current assessment this teacher is focused on students' deep learning and understanding, by having them complete an interview assignment and writing a paper about this. For her ideal assessment, she would focus more on the development of students' understanding by 'small

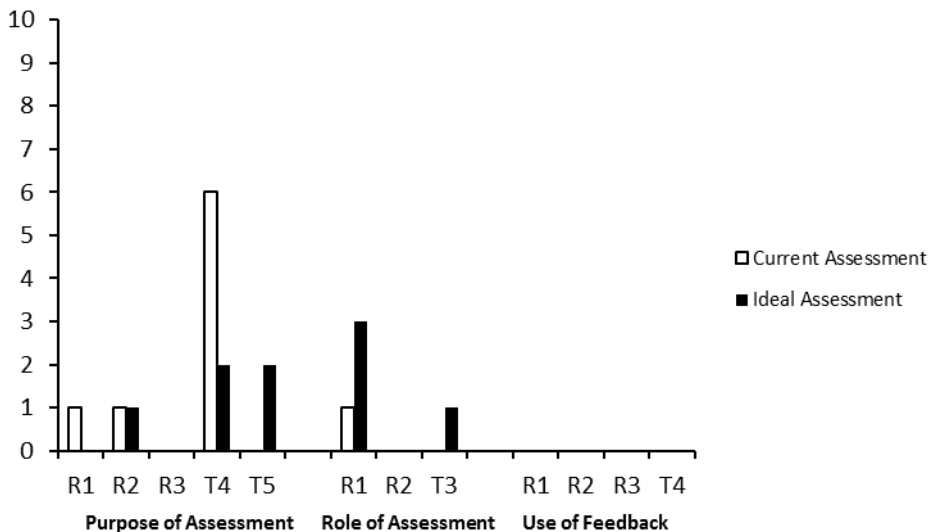


Figure 4.4. Number of quotes discussing assessment conceptions and beliefs of teacher 9. See Table 4.1 for an extensive description of each code.

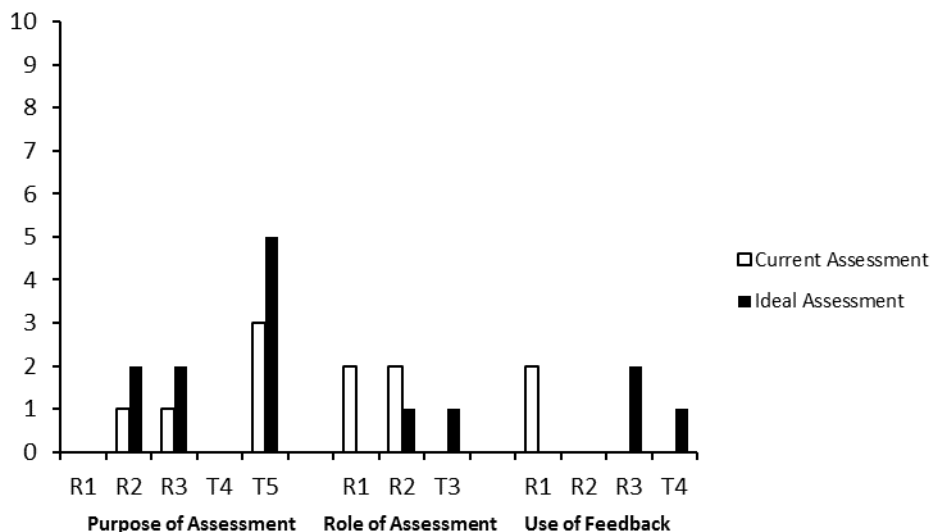


Figure 4.5. Number of quotes discussing assessment conceptions and beliefs of teacher 6. See Table 4.1 for an extensive description of each code.

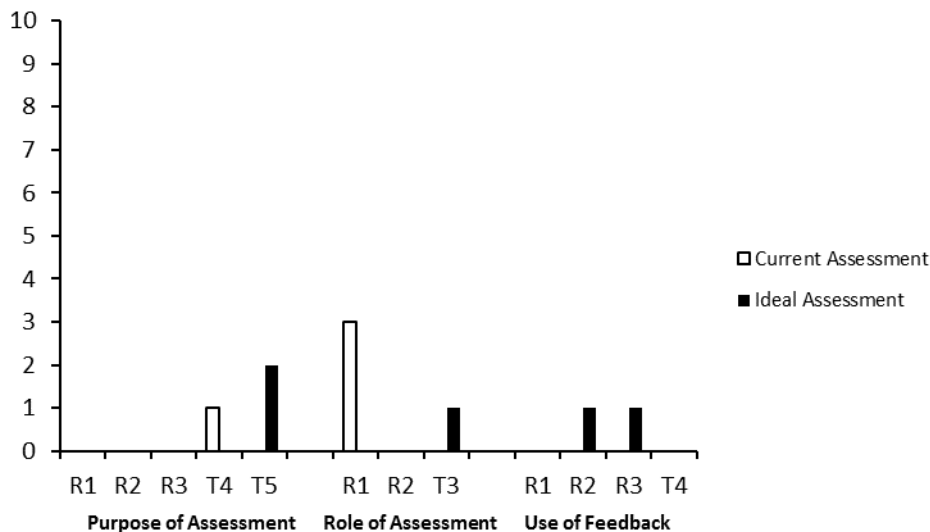


Figure 4.6. Number of quotes discussing assessment conceptions and beliefs of teacher 7. See Table 4.1 for an extensive description of each code.

scale education where they go through the full [research] cycle [...] that's something I would enjoy doing and [something] I think they would really benefit from' (Teacher 7).

When investigating her assessment beliefs, she displays more transformational beliefs for the ideal assessment. Whereas the current assessment was shaped to force students to start working before the final exam, she felt the ideal assessment should be integrated in a research skill development track across courses. This teacher only expressed feedback beliefs for the ideal assessment, again indicating that her focus is more transformational for this assessment than for the current.

4.3.2 Teachers who show differences between current and ideal assessment

4.3.2.1 Ideal assessment is more transformational

Seven teachers (2, 3, 5b, 8, 10, 11, 12) portrayed more transformational conceptions for their ideal assessment. They would focus on developing student understanding, by administering essays (3, 5b, 8, 10), workgroup assignments (teacher 11), continuously testing (teacher 2) or testing during lectures (teacher 12).

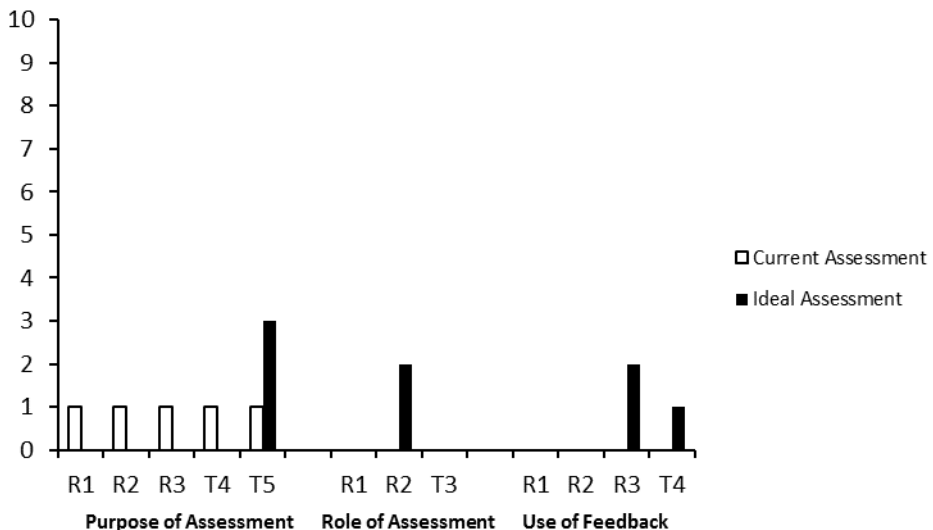


Figure 4.7. Number of quotes discussing assessment conceptions and beliefs of teacher 10. See Table 4.1 for an extensive description of each code.

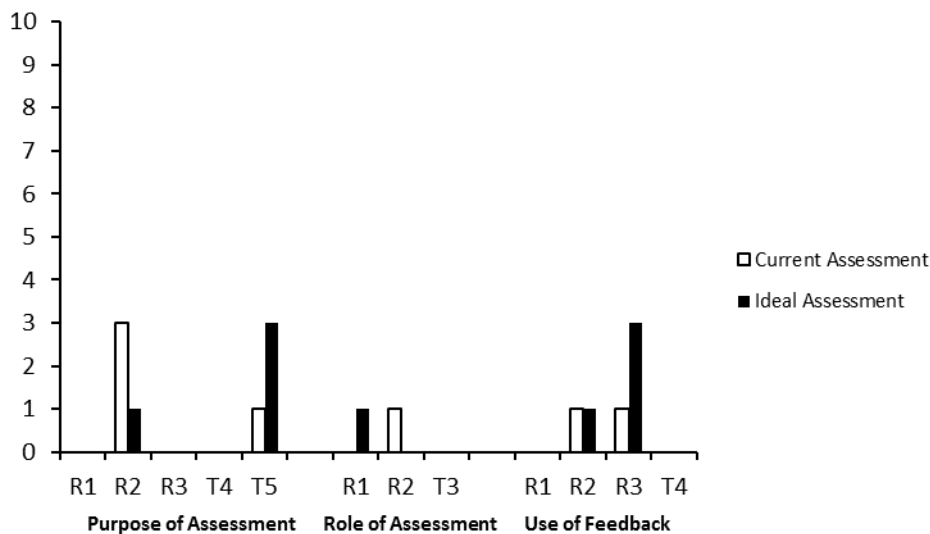


Figure 4.8. Number of quotes discussing assessment conceptions and beliefs of teacher 8. See Table 4.1 for an extensive description of each code.

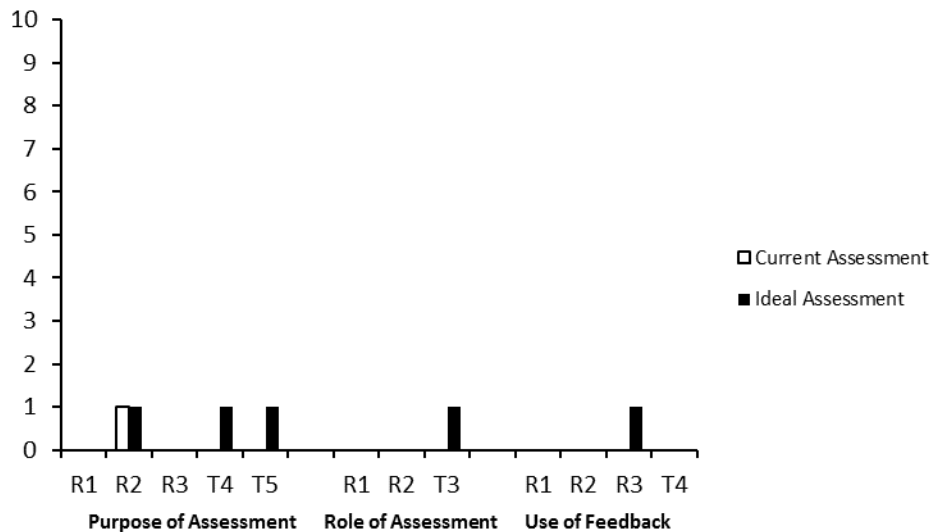


Figure 4.9. Number of quotes discussing assessment conceptions and beliefs of teacher 5b. See Table 4.1 for an extensive description of each code.

All four teachers that would prefer essays as their assessment, also portrayed beliefs that feedback should be used to monitor and improve student learning, portrayed in the 'use of feedback' columns of Figures 4.7-4.10. About students going through a learning curve by using multiple essays, teacher 10 said 'In the beginning [students] relatively have a lot of trouble with it, and in the end they really realise how they should do it, it is very rewarding for the student and the teacher'. Teacher 8 emphasised the importance of monitoring learning in the following way 'A drawback of just doing an exam is that all subsets of the material have the same weight. That way, you can fundamentally misunderstand one facet, but compensate that with knowledge of the other material.' He felt that by assessing students weekly teachers could gain more insight in the specific knowledge gaps of students.

As is apparent from Figures 4.7 – 4.10, the teachers have differing beliefs about the role of assessment in education. Like her colleague teacher 5a, teacher 5b laments the fact that students often need to be forced to study, because she would prefer students to be more independent and in charge of their learning. 'I

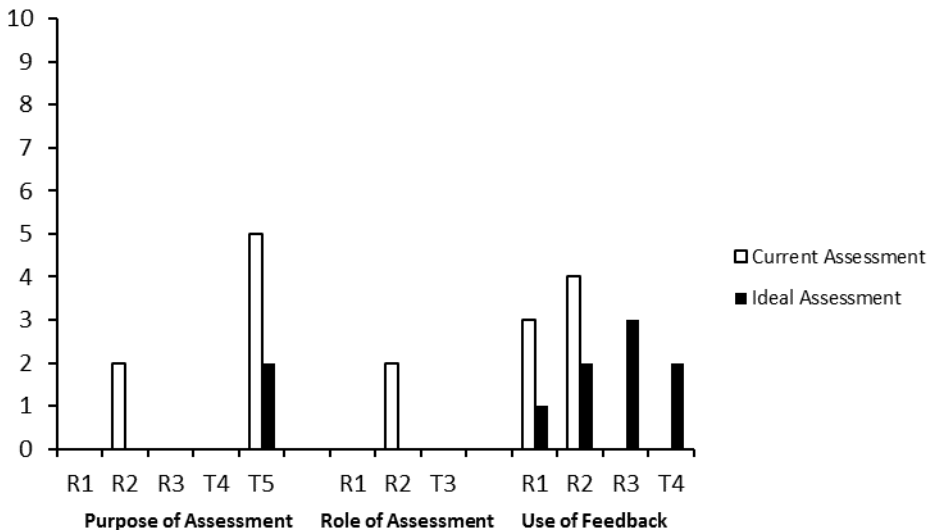


Figure 4.10. Number of quotes discussing assessment conceptions and beliefs of teacher 3. See Table 4.1 for an extensive description of each code.

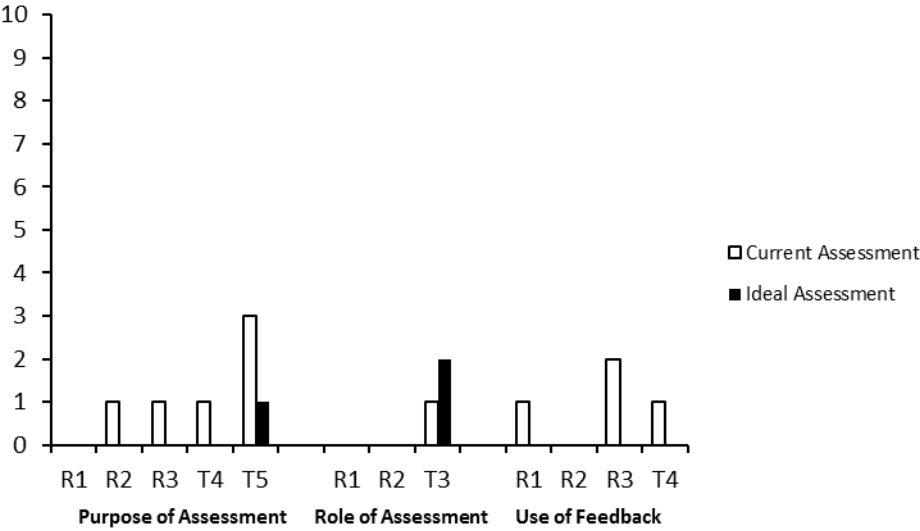


Figure 4.11. Number of quotes discussing assessment conceptions and beliefs of teacher 2. See Table 4.1 for an extensive description of each code.

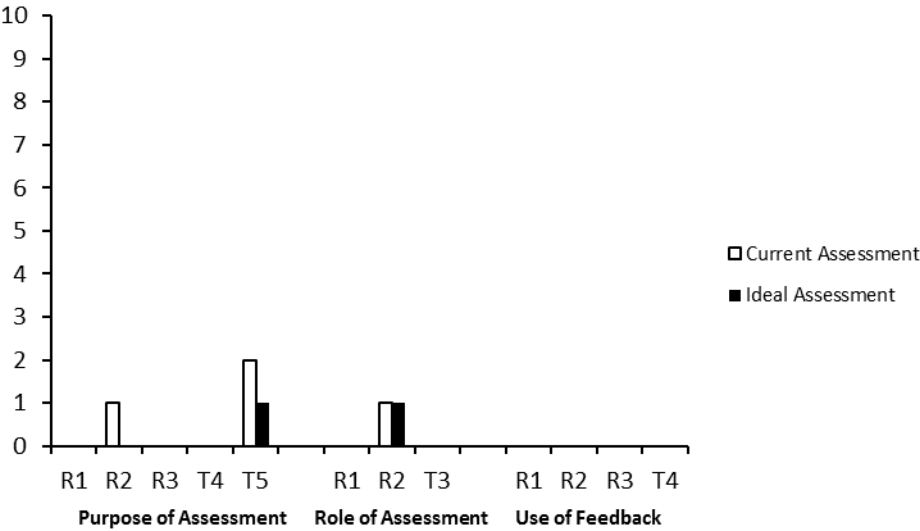


Figure 4.12. Number of quotes discussing assessment conceptions and beliefs of teacher 11. See Table 4.1 for an extensive description of each code.

believe intrinsic motivation is so important. You can devise so many intermediate assessments and everything, but a large part is dependent on [intrinsic motivation].’ Teacher 3 also touched on intrinsic motivation, wanting to spark enthusiasm by giving students free reign in choosing an essay topic ‘Getting an A+ is nice, but I believe speaking your mind is much nicer.’

The three teachers with other ideal assessments than essays did not discuss transformative feedback beliefs for the ideal assessment. However, teacher 2 wondered ‘whether [something] is an intermediate assessment if there is no feedback’ when discussing the current assessment. Furthermore, like Figure 4.11 exemplifies, teacher 2 believes assessment is an integral part of education in concordance with his ideal assessment. ‘I would test them continuously and I would advise them to continuously test themselves.’ Teacher 11 (Figure 4.12) is more focused on inciting students to work on learning, citing the advantage of intermediate assessment is that students ‘may learn to organise their study behaviour’ and get an indication whether they understand the material. For her ideal assessment, teacher 12’s goals were two-fold. She wanted to use testing during lectures to gauge the quality of her lectures and her teaching on the one

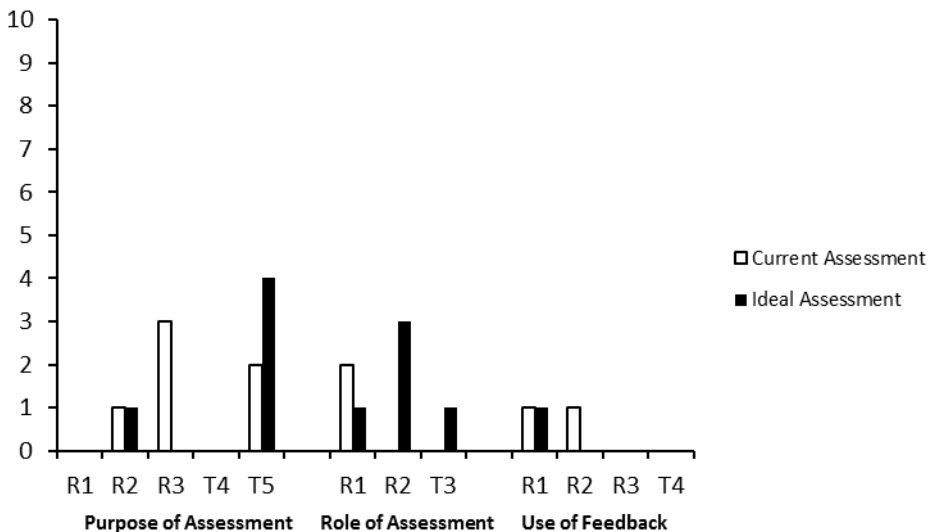


Figure 4.13. Number of quotes discussing assessment conceptions and beliefs of teacher 12. See Table 4.1 for an extensive description of each code.

hand, and to improve learning on the other hand, since ‘the effect of lectures is dependent on how actively you are processing the material. So how much of it is memory retrieval, independent thinking etcetera.’ Making the assessment again an integral part of the course, as portrayed in Figure 4.13.

4.3.2.3 *Current assessment is more transformational.*

One teacher (1) expressed more transformational conceptions for the current assessment, a research report, then for her ideal assessment. However, when discussing her ideal assessment, she brainstormed about several assessment types, some with more reproductive conceptions, like multiple choice or open-ended exams. This is apparent in Figure 4.14, which shows a larger variety of conceptions for the ideal assessment. Ultimately, this teacher focused on an ideal assessment of discussing written work and the ‘Harvard’ classroom assessment, where she displayed several transformational conceptions as well.

When addressing her assessment beliefs, no discussion of the role of the ideal assessment in teaching and learning came up, but regarding her feedback beliefs this teacher focused on using feedback to monitor learning, possibly even across courses.

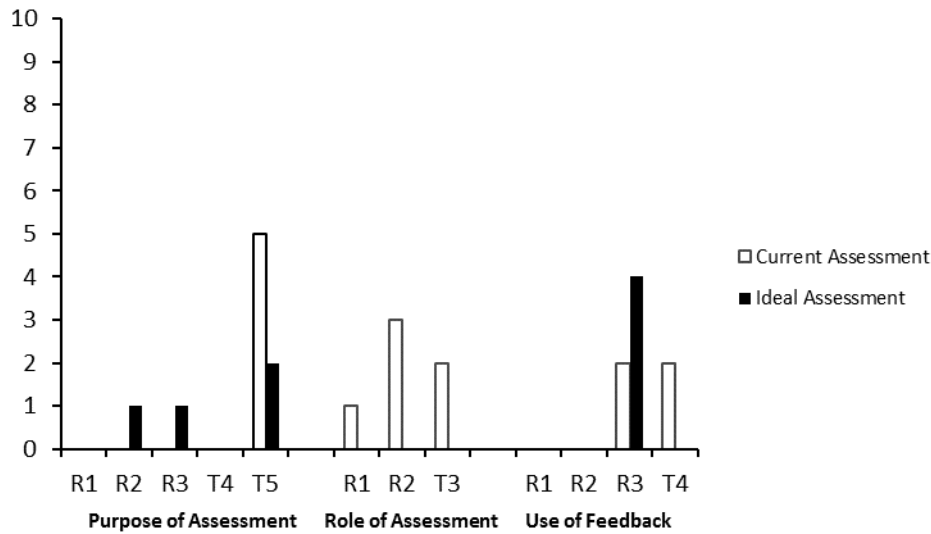


Figure 4.14. Number of quotes discussing assessment conceptions and beliefs of teacher 1. See Table 4.1 for an extensive description of each code.

'Do you give the feedback [in a half semester course] for them to use in that course and write a paper, or do you try to teach first-year students how to write a paper and give them extensive feedback that they can use for the rest of their education?'

4.3.3 Reasons for discrepancy between current and ideal assessment

Several teachers gave rationalisations for the fact that their ideal assessment is not their current assessment. A main explanation that came up was that the ideal assessment usually added workload: 'But that would take an hour per essay, it would take 600 hours and I only work here 25 hours per week' (Teacher 3). Further explanations were 'I am not sure whether that's really suited for first-year students' (Teacher 4), or 'The study culture over there is different' (Teacher 9), coincidentally both referring to 'Harvard' style classroom assessment.

4.4 Discussion

In the current study, we tried to discover teachers' conceptions of intermediate assessment by asking them to reflect on their current assessment practice as well as their ideal assessment. Results indicate that the majority of teachers have transformational assessment conceptions, and that, overall, the conceptions for the ideal assessment are more transformational than those for the current assessment. Teachers posed that development of understanding and support of learning should be the main focus of intermediate assessment. They would like to employ assessments like essays or continuous testing to achieve this goal.

It is encouraging to see that teachers strongly focused on the student learning aspect of assessment. When Watkins et al. (2005) interviewed teachers about how assessment influences student learning (the backwash effect), not all teachers referred to content and deep learning, aspects that did come up often during our interviews. Interestingly, the amount of utterances focusing on deep learning and students' own thinking went down from the current to the ideal assessment, as shown in the purpose of assessment T4 column in Figure 4.1. A possible explanation for this is that for the ideal assessment teachers wanted to focus more on the development of students' own thinking, which would make their utterances fall under purpose of assessment T5. This may be a downside of our decision to show all conceptions portrayed by teachers instead of just the highest level they portrayed.

Black and Wiliam (2004) pose that formative assessment is the optimal assessment for student learning. Even though strict discussion of formative assessment did not occur in our study, teachers explicitly focused on learning effects of the assessment, as well as on summative functions of their assessment. In this light, it is interesting that Segers and Tillema (2011) found that secondary school teachers did not differentiate between the summative and formative functions of assessment. Their results show teachers' conception that assessment should inform students and teachers about students' performance, whether it be through summative or formative assessments. This conception overlaps with Postareff et al.'s (2012) conception five, and Samuelowicz and Bain's (2002) use of feedback dimension, which were often occurring codes in our interviews.

According to Gibbs and Simpson (2004) assessment supports learning when students receive timely feedback that they attend to. This is endorsed by teachers in our study, who stress the importance of feedback in monitoring student learning and helping them improve. Beaumont, O'Doherty and Shannon (2011) found that even though university tutors reported that they provide extensive written feedback, students often do not perceive this as useful. There may be a disconnect between 'after the fact' feedback, and feedback as a more dialogical process during which students are guided through the assessment. In the current study, teachers also discussed feedback after the fact, but because of the nature of intermediate assessment, this may have a more dialogical effect than after the fact feedback for end-of-term assessment.

Our results contrast with those of Postareff et al. (2012), who found that a majority of teachers has reproductive conceptions. A possible explanation for this discrepancy is that the current study specifically focused on intermediate assessment. When contrasting intermediate and end-of-term assessment it becomes clear that, by design, intermediate assessment is more suitable to support student learning throughout the term, whereas end-of-term assessment should measure students' achievement against a set of criteria.

This study also found a discrepancy between teachers' conceptions and their practices even though previous research has found these two to be highly aligned (Postareff 2012). Reimann and Sadler (2017) argue that teachers' understandings of assessment cannot be fully discovered without looking at how they enact these understandings, indicating that teachers in the current study may not have full understanding of their ideal assessment. However, for their ideal assessment, several teachers drew inspiration from courses they taught

in higher years, indicating that these are assessment practices, just not for first-year courses. Possibly, the conceptions of teachers who only have experience teaching first-year courses could be much different. Furthermore, conceptions and practice may not always align. Sin, Tavares and Amaral (2017), for example, found that teachers incorporated measures to further employability into their curriculum, even when they did not feel that ensuring employability is an important goal of higher education.

Teachers often mentioned the workload as a constraint for using the transformational assessments that were their ideal. Myers and Myers (2015) have also identified that workload is often a factor for teachers when deciding on assessments. This is also in concordance with results from Chapter 3, where teachers mentioned they took workload into account when designing their intermediate assessments. Four teachers participated in both the previous and the current study, indicating that workload is still an important factor for them, over two years later.

In general, transformational assessment is regarded as appealing to higher order cognitive processes (Kratwohl 2002), whereas reproductive assessment appeals to lower order processes. Therefore, transformational assessments are often seen as more ideal in supporting student learning. However, transformational assessments often also pose a higher workload for teachers. Samuelowicz and Bain (2002) suggest that some teachers may reject more transformational assessments not just because of workload, but because their beliefs about assessment do not correspond with transformational assessment. However, our current study indicates that teachers see the benefit of transformational assessments, but they often do not put this into practice because of practical constraints. Postareff et al. (2012) suggest that teachers should be made aware of the influence of assessment on student learning, but our results indicate that teachers are aware of these positive effects, but are unable to translate them into practice.

4.4.1 Directions for future research

Since our study was a preliminary exploration of university teachers' conceptions of intermediate assessment, additional research should be conducted to further investigate and corroborate these conceptions. Further research could focus on the differences in conception for intermediate and end of term assessment, instead of only one of the two.

With regard to the discrepancy between teachers' conceptions and their practice, future studies could focus on investigating the specific needs of teachers to bring their ideas into practice.

Furthermore, our study only focused on teachers in first-year psychology, criminology, or law courses, and future studies can expand on the domain, as well as the educational year of the courses.

4.4.2 Concluding remarks

If teachers in higher education do have transformational assessment conceptions, but feel constrained by outside factors, extensive thought should be put into how we can support teachers in their assessment practice and encourage them to use more transformational assessments.

Chapter 5

**Explaining Individual Student Success Using
Intermediate Assessment Types and Student
Characteristics**

Individual student success is influenced by the educational environment and student characteristics. One adaptation of the educational environment to improve student success is the introduction of intermediate, or continuous, assessment. Previous research already identified several student characteristics that are related to student success as measured by student achievement, like previous achievement, motivation, self-efficacy, and gender. The two facets are investigated in a group of first-year undergraduate Law students in the Netherlands, by examining the relationship of different types of intermediate assessment and student characteristics with academic achievement. A questionnaire, measuring demographic information, self-regulation, and motivational constructs, was completed by 94 students and their grades were requested from the student administration. Repeated measures ANCOVAs with assessment type as the within-subject factor identified that student achievement is not dependent on the type of intermediate assessment. Students with higher high school GPAs got higher scores across assessment types. Male students performed worse than their female peers in courses without intermediate assessment, but in courses using any type of intermediate assessment this gender difference disappeared. Intrinsic motivation was a negative predictor for achievement in courses using writing assignments and mandatory homework assignments. Results from the current study indicate that intermediate assessment may be a potent measure to improve male students' success by closing the gender achievement gap, and that students with high levels of intrinsic motivation do not benefit from intermediate assessment.

This chapter has been published in adapted form as: Day, I. N. Z., van Blankenstein, F. M., Westenberg, P. M., & Admiraal, W. F. (2018) Explaining Individual Student Success Using Continuous Assessment Types and Student Characteristics. *Higher Education Research and Development*, Advance Online Publication. Doi: 10.1080/07294360.2018.1466868

5.1 Introduction

Student success in higher education has been a topic of interest for several decades (e.g. McKenzie & Schweitzer, 2001; Pascarella & Terenzini, 1991). Prior to this, Tinto (1975) started to develop his model on student drop out, and Feldman and Newcomb (1969) investigated how college education affects student outcomes. Yet student success is still a topic of research and is defined in many ways. Studies in the Netherlands have used drop out (van den Bogaard, 2012), study progress (Kamphorst, Hofman, Jansen, & Terlouw, 2013; van den Bogaard, 2012), and perceived competence (Kamphorst et al., 2013) as definitions. Some other examples are employability (Qenani, MacDougall, & Sexton, 2014), or academic achievement (McKenzie & Schweitzer, 2001). Research on student success is not just about operationalisation, but also about what variables influence student success. Student success can be influenced by the *university environment* (e.g., van Berkel, Jansen & Bax, 2012), *student characteristics* (e.g., Richardson, Abraham, & Bond, 2012), or both. Van Berkel et al. focus on student success in terms of graduation rates, and Richardson et al. in terms of GPA. However, Qenani et al. (2014) pose that employability can also be influenced by factors in the university environment as well as in students. In the current paper, we will focus on academic achievement as a measure of student success and investigate the university environment as well as student characteristics as explaining factors for academic achievement.

5.1.1 The University Environment and Student Success

Several facets of the university environment can play a role in student success. Tinto's (1975) model, for example, focuses on interactions between students and faculty. Additionally, Thomas (2002) posits that what she terms *institutional habitus*, the norms and practices of the institution, can influence student retention, and that retention is greatest when students' habitus corresponds with the institutional habitus. According to van Berkel and colleagues (2012), it is a university's responsibility to shape the curriculum in a way that optimises student success. In their book, several curriculum optimisation measures are presented, like preventing competition of several different course activities, introducing active learning activities, and adjusting the assessment program. The current paper will explore this final measure, and more specifically, the use of intermediate assessment, since previous research has shown that an adjusted

assessment programme is a potent driver of student learning (Cohen-Schotanus, 1999). Furthermore, using a 'range of assessment tools' is also one of the measures for adapting the institutional habitus proposed by Thomas (2002, p. 439).

Intermediate assessment refers to the use of one or several assessments during the course period, instead of a single final exam in the last weeks of the semester. It is also referred to as continuous or frequent assessment (e.g., Rezaei, 2015). Intermediate assessment in higher education can be used to improve student learning (e.g., Rezaei, 2015) as well as student engagement (e.g., Holmes, 2015). In both cases intermediate assessment can be used to provide feedback to students (e.g., De Kleijn, Bouwmeester, Ritzen, Ramaekers, & Van Rijen, 2013) and teachers (e.g., Domenech, Blazquez, de la Poza, & Muñoz-Miquel, 2015). Furthermore, intermediate assessment can be used as a reward system for desired studying behaviour (Admiraal, Wubbels, & Pilot, 1999), which also relates to the cognitive principle of reinforcement learning (Daw & Frank, 2009). Additionally, several of Gibbs and Simpson's (2004) conditions that assessment must meet to support learning correspond to factors of intermediate assessment. Chapter 3 indicates that, at Leiden University, teachers employ intermediate assessment to keep students working during the course period and to be able to assess different knowledge and skills. With this second goal in mind, it is apparent that intermediate assessments can have different types, like essays, presentations, as well as partial exams. Intermediate assessments can be either voluntary or mandatory. However, using voluntary assessments may promote self-selection. Thomas et al. (2017), for example, were unsure whether increased usage of online self-tests could explain higher grades, or whether high achieving students chose to use self-tests more often. To overcome this problem with self-selection, in the current study a constraint for intermediate assessment is that the assessment is mandatory and completion is checked by the teacher.

Intermediate assessment has two main cognitive benefits. First, there is the testing effect (Roediger & Karpicke, 2006) which states that repeated testing of information leads to better retention of this information. According to Butler (2010), the testing effect also extends to final assessments with new information, denoting a transfer of knowledge. The second benefit can be referred to as the spacing effect (Kornell, 2009), spreading your studying across the study period leads to longer retention than last minute cramming does. Dunlosky, Rawson, Marsh, Nathan, and Willingham (2013) cited evidence from the lab and the classroom and stated that practice testing (testing effect) and distributed

practice (spacing effect) are the most beneficial study methods. Furthermore, intermediate assessment leaves students with time to reflect on their learning and their results. According to Moon (1999) “reflection makes deeper and better considered knowledge available to us.” (p.155).

Several studies have found that using intermediate assessment in higher education courses improves student achievement (e.g., Domenech et al. 2015; Nelson, Robison, Bell, & Bradshaw, 2009; Tuunila & Pulkkinen, 2015). However, this research usually does not contrast different types of intermediate assessment. Therefore, there is no information whether some types of intermediate assessment are more beneficial to student achievement than others.

In sum, intermediate assessment can lead to more effective study behaviour and promote student academic achievement. After discussing intermediate assessment as a change in the educational environment to promote student success, we now continue with the role student characteristics play in academic achievement.

5.1.2 Student Characteristics and Student Success

Research into the relationship between student characteristics and academic achievement has identified a wide variety of predictors. Student characteristics include motivational constructs, previous achievement, and more demographic information. McKenzie and Schweitzer (2001), for example, found that previous achievement, self-efficacy, and whether students had a job were significant predictors of academic achievement. An oft cited paper discussing student characteristics related to academic achievement is the meta-analysis by Richardson et al. (2012). This meta-analysis identified 41 characteristics that are correlated with academic achievement. These were cognitive characteristics, like high school GPA, as well as non-cognitive characteristics, like motivation and self-regulation. To narrow down the list of correlates the current paper focuses on the strongest correlates which are high school GPA, academic self-efficacy, effort regulation and performance self-efficacy, showing medium to large correlations. In addition, we focus on a few of the conceptually related smaller correlates like learning goal orientation, academic intrinsic motivation, and metacognition. Furthermore, we also include gender, a small correlate in Richardson et al.'s study.

5.1.3 Intermediate Assessment and Student Characteristics

Intermediate assessment and student characteristics can influence academic achievement independently, but they can also influence each other. Possible interplays between intermediate assessment and student characteristics are specifically interesting in the light of optimising the curriculum for student success. When different groups of students get different benefits, this may present a case for more individualised assessment paths.

The most apparent case that intermediate assessment and student characteristics relate to each other may be that of students who lack the self-regulation skills for independent study throughout the semester. Teachers and students praised the fact that intermediate assessments help to keep students on track (see Chapter 3) and in Peat and Franklin's (2002) study, students remarked mainly using self-assessment modules as learning guides and not as assessment tools.

Looking at student ability and intermediate assessment, research shows that higher achieving students benefit more from intermediate exams (de Paola & Scoppa, 2011) or that lower achieving students perform better each intermediate assessment, while higher achieving students started regressing to the mean (Kerdijk, Tio, Mulder, & Cohen-Schotanus, 2013).

When relating motivation to intermediate assessment and achievement, Ibabe and Jauregizar (2010) found that students with higher motivation made more use of the online self-assessment tool and that students who used the tool had higher achievement. However, even students with lower motivation levels used online self-assessment.

Several researchers have looked at gender differences in academic achievement. When looking at general achievement, Richardson et al. (2012) identified that female students perform better than their male peers. In the case of intermediate assessment, this picture is less clear. Domenech et al. (2015) found no significant gender differences for students taking frequent cumulative tests. However, they discerned a trend where women got higher grades but men had higher exam passing rates. Research by Cano (2011) suggests that when students have the opportunity to choose whether they want to participate in intermediate assessments, women more often opt-in than men. Furthermore, female engineering students rate themselves and their female peers lower on peer and self-assessment tasks than their male counterparts do (Torres-Guijarro & Bengoechea, 2017).

To summarise, there seems to be interplay between intermediate assessment and several student characteristics. Unfortunately, this relationship is still largely unclear. Therefore, in the current study, we try to answer the following two research questions:

- (1) To what extent does the type of intermediate assessment relate to academic achievement?
- (2) What role do gender, high school achievement, motivation and self-regulation play in this relationship?

5.2 Methods

5.2.1 Context

The study was conducted during the 2014-2015 academic year at the undergraduate law school of a research university in the Netherlands. This law school offers bachelor degrees in Criminology, Law, Fiscal Law, Notarial Law, Business, International Business Law, and Law and Economics. During the first year, which is a foundation year, the majority of courses are the same for all law majors and about 45% of courses is also a part of the criminology program. A full overview of the courses in the program and, when applicable, their intermediate assessment, can be found in Table 5.1. To reiterate, intermediate assessments are checked for completion by teachers, assessments that are graded are marked in the table. Courses without intermediate assessment generally do have required readings or homework assignments, but there is no check to see if these are actually completed. All course information was gathered from the university's e-prospectus. Both majors take courses amounting to 1680 hours of study work.

Table 5.1. Overview of the first-year courses in the 2014-2015 academic year, their intermediate assessment and final exam results.

Course No. and Description of Intermediate assessment	Major	N	M (SD)
LLP	L + C		
1. -	L + C	85	7.06 (1.21)
2. Short written assignment(s)	L + C	86	5.73 (1.09)
3. Partial Exam (Case, open ended)†	L	79	6.03 (1.37)
4. Homework assignments (e.g., debate, plea, case)	L + C	85	5.66 (1.23)
5. -	L	76	5.61 (1.59)
6. -	L	79	7.06 (1.15)
7. Short written assignment(s)	L + C	89	7.29 (1.30)
8. Written assignment†*	L	38	7.63 (0.91)
9. Written assignment†*	L	16	7.19 (0.75)
10. Three Written assignments†*	L	1	8.00 (-)
11. Written assignment†*	L	13	6.23 (1.01)
12. -	L	74	5.78 (1.17)
13. -	L	74	6.32 (1.26)
14. -	L	72	6.85 (1.07)
15. Portfolio of homework assignments†	C	8	6.38 (0.74)
16. Presentation, prepositions, mini experiment and report	C	8	5.75 (1.40)
17. Written Assignments†	C	8	5.88 (0.64)
18. Partial Exam (essay questions), presentation†	C	7	6.63 (1.41); 5.86 (1.57)
19. Paper based on interview†	C	7	6.29 (1.11)
20. Three written assignments	C	7	7.29 (0.76)

Note. L is Law, C is Criminology

†Denotes courses where the intermediate assessment counts towards the overall grade.

* Course 8 is taken by Law, Fiscal Law and Notarial Law majors, Course 9 is taken by Business majors, Course 10 is taken by Economic majors, and Course 11 is taken by International Business Law majors.

5.2.2 Participants

Ninety-four first year students (42.6% male) completed the full questionnaire. The majority of students majored in Law (64.9% Fiscal, Notarial or Law, 24.5% other), whereas only eight percent were Criminology majors. Over three quarters (77.7%) of students were 18 or 19 years old, with the remainder of students being older. Eighty-one percent of students are of Dutch origin and almost 90% entered undergraduate studies directly after high school.

5.2.3 Materials

5.2.3.1 Student Characteristics

Demographics, self-regulation, and motivation, were measured using a slightly adapted version of the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, Smith, Garcia, & McKeachie, 1993). The following eight scales were used: Intrinsic Goal Orientation, Extrinsic Goal Orientation, Task Value, Control of Learning Beliefs, Self-Efficacy of Learning and Performance, Time and Study Environment, Metacognitive Self-Regulation, and Effort Regulation. All questions were translated to Dutch based on the translation used by Blom, Severiens, Broekkamp, and Hoek (2004) and adapted to be applicable to the whole course program instead of one specific course. A translation back-translation procedure was used to check the accuracy of the translated items. All items were answered on a Likert scale ranging from 1 (*not at all applicable to me*) to 5 (*very applicable to me*). The eight MSLQ scales encompassed a total of 50 questions. Table 5.2 shows an overview of the reliabilities of the scales. In addition to the 50 MSLQ questions students were asked to answer questions about their major, age, cultural background, high school exam grade, and prior education. For all these questions the expected most frequent answers were supplied as multiple-choice options, with an open ended “other” option added.

Table 5.2. Reliabilities and Mean Scores for Motivated Strategies for Learning Questionnaire Scales

MSLQ Scale	Item N	Reliability	M (SD)
Intrinsic Goal Orientation	4	0.654*	3.53 (0.55)
Extrinsic Goal Orientation	4	0.710*	3.57 (0.64)
Task Value	6	0.731	3.98 (0.42)
Control of Learning Beliefs	4	0.755*	3.83 (0.57)
Self-Efficacy of Learning and Performance	8	0.819	3.75 (0.51)
Time and Study Environment	8	0.702	3.65 (0.57)
Effort Regulation	4	0.748*	3.63 (0.65)
Meta-cognitive Self-Regulation	12	0.640	3.20 (0.43)

* Spearman-Brown predicted reliability for scale length six items.

5.2.3.2 Student Achievement

To get a measure of student achievement, first-try final exam grades were collected from the student administration. Based on the assessment type they use, the courses can be classified in six groups. These are: ‘no intermediate assessment’ (course $N=6$), ‘written assignment(s)’ ($N=8$), partial exam ($N=2$), ‘mandatory homework assignments’ ($N=2$), ‘interview and paper’ ($N=1$), and ‘presentation, proposition and mini-experiment’ ($N=1$). For each of these groups, an average grade on all courses was calculated as a composite score.

5.2.4 Procedure

The questionnaire was handed out during the coffee break of a lecture of Course 7, where approximately 275 students attended (response rate 34.2%). This course was chosen because it is taught to all majors simultaneously and takes place during the second semester of the academic year. Therefore, students already got a full impression of what their major was like, and early drop outs were not going to participate in the research. The objectives of the study were introduced briefly before the break in a plenary announcement by the first author.

5.2.5 Ethics

The current research was approved by the ethical committee of the psychology department of our university. The first page of the questionnaire was an informed consent letter that explained additional information about the research and asked students’ permission to access their grades. Students were asked to fill in their

student ID number, to be able to connect questionnaire data to student results, but confidentiality of results was guaranteed. Only questionnaires including a signed consent form were included in the study.

5.3 Results

5.3.1 Descriptive Statistics

Mean exam scores for the courses can be found in Table 5.1, and mean scores on all MSLQ scales can be found in Table 5.2. Mean scores for the assessment type composite scores can be found in Table 5.3. The composite scores for interview and presentation were excluded from further analysis since both have a student *N* lower than 10.

Table 5.3. Descriptive Statistics for Assessment Type Composite Scores

Assessment Type Composite Score	Course N	Student N	M (SD)
No Intermediate Assessment (Courses 1,5,6,12,13,14)	6	89	6.39 (0.95)
Written Assignment (Courses 2,7,8,9,10,11,17,20)	8	89	6.68 (0.91)
Partial Exam (Courses 3,18)	2	87	6.04 (1.35)
Mandatory Homework Assignment (Courses 4, 15)	2	85	5.73 (1.15)
Interview and Paper (Course 19)	1	7	6.29 (1.11)
Presentation, Proposition, Mini-Experiment (Course 16)	1	8	5.75 (1.39)

Note. Course numbers for each assessment type composite score correspond to course numbers in Table 5.1.

5.3.2 Preliminary Regression Analyses

Hierarchical regression analyses were run for each assessment type composite score individually to investigate which predictors were related to achievement. Variables were included in the model based on the paper by Richardson et al. (2012). The strongest predictors high-school GPA, self-efficacy and effort regulation were added in the first step. The weaker correlates gender, age, intrinsic goal orientation, extrinsic goal orientation, task value, metacognitive self-regulation, and time and study environment were included in the second step. The third and final step included the variable control of learning beliefs, which is not discussed by Richardson et al. Outcomes (not pictured) show that the only significant predictors were High-School GPA, Intrinsic Goal Orientation, Task Value and Gender. These four variables were added to the repeated measures ANCOVAs as between-subjects variables.

5.3.3 Student Characteristics, Assessment Type, and Student Achievement

To fully investigate the relationship between student characteristics, assessment characteristics, and student performance, two repeated measures ANCOVAs were conducted. To be able to fully investigate the relationship between the different variables, individual regression parameters were requested in SPSS.

5.3.3.1 Contrasting Courses with and without Intermediate Assessment

In the first analysis, investigating the role of student characteristics in courses with and without intermediate assessment, the within-subject variable assessment had two levels. The between-subject variables were high school GPA, intrinsic goal orientation, task value and gender, the latter of the four is dichotomous, the other variables are continuous and therefore added as covariates.

Results from this analysis indicate that students' achievement is not dependent on whether their course has intermediate assessment, $F(1,79) = .021, p > .05$. Main effects were found for three variables. First, high school GPA, $F(1,79) = 36.09, p < .001$, partial $\eta^2 = .314$, which indicates that students' who had higher previous achievement also have higher achievement in university. Second, intrinsic goal orientation, $F(1,79) = 7.10, p = .009$, partial $\eta^2 = .084$ where higher levels of intrinsic goal orientation lead to lower achievement. The third and final significant main effect is that of gender, $F(1,79) = 5.28, p = .023$, partial $\eta^2 = .064$ indicating that female students perform better than their male peers. This gender effect, however, is characterised by an assessment by gender

interaction effect, $F(1,79) = 7.68, p = .007$, partial $\eta^2 = .089$ where there is only a gender difference on courses that do not have intermediate assessment. For courses that do have intermediate assessments there is no difference in score for men and women. The individual influence of each variable on the two types of courses can be found in Table 5.4.

Table 5.4. Parameter Estimates for ANCOVA Comparing Courses with and without Intermediate Assessment.

Use of Intermedi- ate Assessment	Parameter	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	Par- tial η^2
No Intermediate Assessment	Intercept	2.24	1.10	20.3	.046	.050
	High School GPA	0.80	0.15	5.49	<.001	.276
	Male Gender	-0.57	0.18	-3.17	.002	.113
	Intrinsic Goal Orientation	-0.40	0.17	-2.35	.021	.065
	Task Value	0.05	0.24	0.22	.829	.001
Intermediate As- sessment	Intercept	2.16	1.06	2.05	.044	.05
	High School GPA	0.77	0.14	5.50	<.001	.277
	Male Gender	-0.18	0.17	-1.02	.311	.013
	Intrinsic Goal Orientation	-0.43	0.16	-2.58	.012	.078
	Task Value	0.08	0.23	0.36	.718	.002

5.3.3.2 Investigating the Three Types of Intermediate Assessment

We subsequently ran another Repeated Measures ANCOVA, with a three level within-subject variable to investigate whether there are different outcomes for different assessment types. The three levels were *written assignments*, *partial exam*, and *mandatory homework assignments*. The same four between-subject variables as in the previous analysis were included.

The assumption of sphericity was violated for assessment type, $\chi^2(2) = 18.62, p < .001$, therefore, Huynh-Feldt estimates of sphericity were used to correct the degrees of freedom ($\epsilon = .88$).

There was still no main effect of assessment type $F(1.76,131.76) = 1.49, p > .05$, indicating that students scored similarly on courses with different assessment types.

There was a main effect of High School GPA $F(1,75) = 37.26, p < .001$, partial $\eta^2 = .332$ indicating that a higher high school GPA was related to higher

university grades. There is no main effect of gender in this analysis, $F(1,75) = 2.78, p > .05$, which shows that men's scores do not differ from those of their female peers across all three intermediate assessment types.

There were no main effects for intrinsic goal orientation and task value, but both these variables interacted with the assessment type, $F(1.76,131.76) = 3.79, p = .03$ and $F(1.76,131.76) = 4.95, p = .011$, respectively. Investigation of the parameter estimates indicates that intrinsic goal orientation is a negative predictor of students' grades on courses with written assignments, suggesting that students with higher levels of intrinsic goal orientation for their Law program get lower grades for these courses. This contrasts with the result from the comparison of courses with and without intermediate assessment, where intrinsic goal orientation was a negative predictor for all courses. Task value is a negative predictor of students' grades on courses with mandatory homework assignments, again suggesting that students who have a higher task value of their studies score lower on courses with mandatory homework. A full overview of all parameter estimates for each assessment type composite score can be found in Table 5.5.

Table 5.5. Parameter Estimates for ANCOVA comparing the three assessment type composite scores.

Assessment Type Composite Score	Parameter	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	Par- tial η^2
Written Assign- ment	Intercept	2.55	1.05	2.42	.018	.073
	High School GPA	0.76	0.14	5.47	<.001	.285
	Male Gender	-0.19	0.17	-1.11	.270	.016
	Intrinsic Goal Orientation	-0.48	0.16	-2.94	.004	.103
	Task Value	0.17	0.24	0.69	.490	.006
Partial Exam	Intercept	0.06	1.76	0.04	.97	<.001
	High School GPA	1.01	0.23	4.36	<.001	.202
	Male Gender	-0.46	0.29	-1.63	.108	.034
	Intrinsic Goal Orientation	-0.52	0.27	-1.90	.062	.046
	Task Value	0.24	0.40	0.60	.551	.005
Mandatory Home- work Assignment	Intercept	2.27	1.51	1.50	.14	.029
	High School GPA	0.91	0.20	4.57	<.001	.217
	Male Gender	-0.25	0.25	-1.00	.319	.013
	Intrinsic Goal Orientation	0.12	0.24	0.50	.616	.003
	Task Value	-0.83	0.34	-2.42	.018	.072

5.4 Discussion

This paper focused on two research questions. The first was to what extent the type of intermediate assessment relates to academic achievement, and the second investigated the role of gender, high school achievement, motivation and self-regulation in this relationship.

Results from the current study indicate that the type of intermediate assessment does not influence academic achievement. This result suggests, first of all, that students do not perform differently depending on whether they need to complete written assignments, a partial exam, or homework assignments.

However, the second suggestion of the lack of a main effect of assessment type is that students do not perform better on courses whether these courses use intermediate assessment or not. This contrasts with previous research that discovered that, in most cases, intermediate assessment positively influences students' achievement (e.g. Domenech et al., 2015; Ibabe & Jauregizar, 2010; Rezaei, 2015). A possible explanation for the lack of results is the structure of the curriculum. Cognitive advantages of intermediate assessment like distributed practice (Dunlosky et al. 2013) or time for reflection (Moon, 1999), could be cancelled out by the fact that all courses have distributed educational meetings. Students may have prepared for meetings irrespective whether they had intermediate assessments or not, independently distributing their practice throughout the semester.

With respect to the second research question, we see results for four student characteristics. Surprisingly, the seven other characteristics in the research did not relate to student achievement. Based on the results we can paint the following picture.

First of all, students with a higher high school GPA score higher on courses with all different assessment types. High school GPA is one of the stronger correlates of university achievement (Richardson et al., 2012), and this paper presents more evidence for this case.

The second characteristic that plays a role in intermediate assessment is gender. On average male students get lower grades than their female peers. However, in the present study, this difference was only significant in the case of courses that use no intermediate assessment. This result is interesting in the light of previous research (Richardson et al., 2012) that suggests that the achievement of male students lags behind. The fact that one gender outperforms the other is often

called the gender achievement gap. Several studies found that female students perform better than their male counterparts, not just in higher education, but across all educational levels (Machin & McNally, 2005; Richardson et al, 2012). However, depending on the discipline, the gender achievement gap may be reversed (Miyake et al., 2010). A gender achievement gap is generally unwanted, and several measures to bridge this gap are researched. Miyake et al. (2010), for example, used a values affirmation intervention to improve female performance. Our results show that introducing intermediate assessment into the curriculum may be a potent intervention in supporting male students. However, when introducing intermediate assessment to bridge the gender achievement gap, gender differences in assessment achievement should be considered. Like mentioned before, Torres-Guijarro and Bengoechea (2017) found that female students do not score as well on peer and self-assessments as their male peers. So it seems that some types of intermediate assessment only benefit men and not women, and probably vice versa. Supporting male achievement by introducing intermediate assessment should not be simultaneously detrimental to female achievement.

For the third characteristic the results indicate that students with a higher level of intrinsic goal orientation get lower scores on courses using writing assignments as intermediate assessment. This contrasts with the results of Richardson and her colleagues (2012), who found that intrinsic goal orientation is a positive correlate of achievement.

The fourth, and final, characteristic is task value, which exhibits a negative relationship to student achievement for courses that use homework assignments. Again, this result is the opposite of the results suggested by Richardson et al. (2012).

Both intrinsic goal orientation and task value are aspects of student motivation. Intrinsic goal orientation is comparable to learning goal, or mastery orientation and task value to academic intrinsic motivation, and both measures are small correlates of academic achievement (Richardson et al, 2012). However, there are also several studies that identified different relationships. Neroni, Meijs, Leontjevas, Kirschner, and de Groot (2017) for example, discovered that mastery approach goals were no significant predictor of student success, measured as achievement, for the distant education students in their sample. One possible explanation they give for this lack of a relationship is that distant education students with a mastery orientation possibly are not driven by grades at all, and

only enrol in the courses for their own interest, subsequently not participating in the final exams. Additionally, Baker (2004) also did not find influence of any motivational construct on student achievement. Her hypothesis for this lack of an effect is that motivation may have influenced achievement indirectly, via perceived stress and adjustment.

A major difference between the two aforementioned studies and our results is that where those found an absence of a relationship between motivation and achievement, our study actually found a significant negative association for motivation on two types of intermediate assessment. One explanation for this is that the first year is a foundation year. In this foundation year students are presented with courses that introduce them to the different facets of their major, which may not all hold their interest. Since the questionnaire is formulated on a course program level, this difference in interest for specific disciplines could have influenced the way students answered the questions. For example, a student with a large interest in criminal law may have reported high levels of interest in their course program with criminal law courses in mind, but subsequently not achieved very well on the other foundation courses. Additionally, individual course difficulty levels also may have influenced student achievement.

The fact that motivation is related to lower achievement for only two out of four assessment type composite scores complicates the situation even further. It is notable that this relationship does not occur in the courses using a partial exam.

According to Macfarlane (2015), the current higher education climate makes several demands of students. They are expected to attend obligatory class meetings, and to complete assessments during these meetings, a process he calls *presenteeism*. Furthermore, students need to show active participation in the meetings and assessments, which Macfarlane deems *learnerism*. Macfarlane posits that these two processes negate student autonomy to shape their own educational process. Under self-determination theory, a lack of autonomy leads to lower motivation and results (Deci & Ryan, 2000), which can explain the negative impact of motivation in the current study. That is to say, students who report high levels of motivation in the current study may have been demotivated by the lack of autonomy their courses offer. The results of Ibabe and Jauregizar (2010) where motivated students chose to use the self-assessment tool more often also links in to this case of autonomy.

Another explanation for the negative relationship of motivation for writing and homework assignments may be due to a perceived lack of alignment. Results from Chapter 3 indicate that students prefer intermediate assessments that clearly relate to the final exam. It can be argued that this is especially the case in courses that have a partial exam, and less for the other two types of intermediate assessment discussed. In the current study, we cannot comment on this possible lack of alignment, since we did not observe the classes and course materials.

It is striking that the current study did not find support for a relationship between intermediate assessment, student characteristics and achievement for six out of eight MSLQ scales. Two of these (self-efficacy and effort regulation) were marked as medium strength correlates by Richardson and her co-authors (2012). The lack of influence of effort regulation may be explained by the design of the curriculum. Peat and Franklin (2002) already mentioned that students used the intermediate self-assessments as study guides instead of as assessment tools, and items measuring effort regulation focus on persistence in studying. However, all intermediate assessments in the current study were mandatory, and there was no option for students to not persist. In a way, their effort was regulated for them. A possible explanation for the lack of self-efficacy may be that, even though the second semester had already started, students were still unsure of their self-efficacy for their course programs, because of their limited experience with studying in these programs.

5.4.1 Limitations and Directions for Future Research

The main limitation of the current study is the low response rate of almost 35%. The authors suspect that this is partly explained by the fact that the questionnaire was presented as the first step in planned additional research. This additional research requested a more prolonged time investment of students, which may have deterred participation. Low response rates usually lead to biased samples, but when inspecting the average final exam results, the sample does not seem to consist of exceptionally high or low scoring students. The current study could be repeated without the additional requirements, hopefully boosting participation rates.

Furthermore, our study only focused on a subset of the student characteristics investigated in the meta-analysis of Richardson et al. (2012). To expand our results, the relationship between intermediate assessment types and other student characteristics like socio-economic status or personality traits should be studied.

As an expansion on our definition of student success focussing on academic achievement, future research could also investigate how intermediate assessment and student characteristics relate to other student success outcomes like employability, or perceived competence.

Future research also should take a more extended view into courses by not only examining student achievements, but also looking at course materials in depth. Furthermore, qualitative data in the form of teacher and student observations or interview could be included.

Another direction for future research should focus on motivational development during foundation years, and how intermediate assessments relate to this development. This research should also extend to other disciplines, to investigate whether the current outcomes hold true for students studying science and humanities as well.

5.4.2 Concluding Remarks

The results of the current paper indicate the following:

- (1) Intermediate assessment supports student success of male students more than that of female students.
- (2) There does not seem to be a particular type of assessment responsible for this gender difference.
- (3) Writing assignments and mandatory homework assignments may be detrimental for students' motivation.

We believe that teachers who want to improve student achievement by introducing intermediate assessment can benefit by carefully aligning the intermediate assessment with the final examination of the course. Additionally, perceived usefulness of the intermediate assessment is of importance to keep students motivated as well. When these points are taken into account, intermediate assessment can be a potent measure to improve student achievement.

Chapter

6

General discussion

6.1 Introduction

In this dissertation I investigated intermediate assessment characteristics, concentrating on three focal points (implementation, perceptions, and effects). Chapter 2 is an overview of the literature, whereas Chapters 3 – 5 focus on the introduction and subsequent use of intermediate assessment in the curricula of four programs at Leiden University. To reiterate, intermediate assessment is defined as all assessment that takes place during the course period, instead of at the end. In Chapters 3 – 5, an added requirement for intermediate assessment was that the assessment was mandatory and needed to be handed in to the teacher, to prevent self-selection biases that are associated with voluntary assessments.

6.2 Main findings per chapter

Chapter 2. A Review of the Characteristics of Intermediate Assessment and their Relationship with Student Grades

This chapter provides an overview of the literature on intermediate assessments in higher education, with a specific focus on several assessment characteristics. The research question central to this chapter is ‘what characteristics of intermediate assessment are related to student grades.’ In total 88 articles using various types of intermediate assessment were synthesised. These articles showcased the variety of intermediate assessment in higher education: 24 different assessment types were used across many different disciplines. There was also a large variation in frequency of assessment, ranging from a single assessment to 38 assessments. Four main characteristics were further discussed. These were use of feedback, mandatory assessments, the assessor, and the rewards for the assessment. In general, the majority of studies found positive effects of intermediate assessment on student grades. When looking at the four main characteristics, it seems that studies using corrective feedback found positive influences on student grades most often, but literature (Shute, 2008) suggests that elaborate feedback may be more beneficial for some groups. There does not seem to be a definitive answer whether an assessment should be mandatory, because a similar proportion of articles using mandatory assessment as articles using voluntary assessments found positive results for student grades. Peer and self-assessment showed promising results in the studies employing it, but it should be kept in mind that

not all assessment is suitable for peer or self-assessment, and furthermore the number of articles using teacher assessment largely outnumbered those using peer or self-assessment. With regard to the reward for assessment, studies providing students with course credit as a reward for completing intermediate assessment showed the most positive results. This chapter ends with three scenarios for intermediate assessment, to further exemplify how intermediate assessment can be deployed in higher education.

Chapter 3. Teacher and Student Perceptions of Intermediate Assessment in Higher Education

Chapter 3 is a first exploration of the perceptions of intermediate assessment and their types. The study described in this chapter was conducted in the first year of the educational reform introducing intermediate assessment. Teachers involved in teaching first-year courses using intermediate assessment were interviewed and information from student evaluations of teaching and student interviews was used to gauge these perceptions. The research questions for this chapter are:

- (1) What types of intermediate assessment are used in the programmes under investigation?
- (2) How are these types of intermediate assessment perceived by teachers and students?

Contrary to the expectations, there were no differences in perceptions for different assessment types. Overall it can be said that teachers were positive about intermediate assessment as a way to get students to keep up with study work, and as a possibility to measure a variety of knowledge and skills. However, they were not all in agreement with the obligatory nature of intermediate assessment, and some felt that the fact that students could compensate their intermediate assessment grade was not desirable. Furthermore, teachers remarked a few practical concerns, focusing on the added workload of correcting assessments, and administration that came with the assessment. Other practical concerns were relating to possible plagiarism and resits of the assessment. Students were also generally positive about intermediate assessment, but for different reasons. The main attraction for students seemed to be that intermediate assessment kept them on track, that it deepened their knowledge, and that it lowered the stakes of the final exam. In complete opposition to teachers' appreciation for measuring a variety of knowledge and skills, the main critique of students was that sometimes

the intermediate assessment was not related to the final exam. Furthermore, students also complained about the workload of the intermediate assessments.

Chapter 4. University Teachers' Conceptions of their Current and Ideal Intermediate Assessment

Following the exploration of perceptions in Chapter 3, Chapter 4 delves deeper into teachers' ideas about assessment. To research the question 'What differences in conceptions of intermediate assessment do university teachers display when discussing their current and ideal intermediate assessment?' thirteen course coordinators of first-year Law, Criminology, and Psychology courses were interviewed on their ideas about their current assessment and how they would employ intermediate assessment in an ideal situation. Teachers' responses were coded using codes from two existing sources (Postareff Virtanen, Katajavuori, and Lindblom-Ylänne, 2012; Samuelowicz and Bain, 2002). In both these sources teachers' ideas can be identified as either reproductive (focusing on memorisation and reproduction of knowledge) or transformational (focusing on knowledge construction). Overall it could be said that teachers had transformational goals for their current as well as their ideal assessment, but goals for the ideal assessment were more transformational. Teachers were divided into subgroups based on the differences in conceptions for the current and ideal assessment. One teacher had no transformational goals for both, three teachers had a mix of transformational and reproductive goals for both, and one teacher had only transformational goals. Furthermore, seven teachers had more transformational goals for their ideal assessment, and one teacher had more transformational goals for her current assessment.

Chapter 5. Explaining Individual Student Success Using Intermediate Assessment Types and Student Characteristics

In this chapter assessment type was used to quantitatively investigate the relationship between assessment characteristics and student grades. Furthermore, student characteristics were added to this equation, to get a fully informed vision of the subject. The two research questions central to this chapter were:

- (1) To what extent does the type of intermediate assessment relate to academic achievement?
- (2) What role do gender, high school achievement, motivation and self-regulation play in this relationship?

To investigate this, an adapted version of the motivated strategies for learning questionnaire (MSLQ; Pintrich, Smith, Garcia, & McKeachie, 1993) was filled in by 94 first-year law and criminology students. Their final course grades were accessed and courses were divided in six assessment type groups (no intermediate assessment, written assignment, homework assignments, partial exam, interview and paper, presentation propositions and mini experiment). The final two type groups were excluded from analysis, because less than ten students in the sample completed the course. After initial regression analyses indicated which variables were related to outcomes, repeated measures ANCOVA's were conducted to discover whether assessment type and student characteristics influenced student grades. Results indicate that assessment type does not influence students' grades and high school GPA is the strongest predictor of achievement. Furthermore, male students scored worse on courses without intermediate assessment, but this gender achievement gap disappeared in courses using any type of intermediate assessment. Lastly, two measures for intrinsic motivation (i.e., task value and intrinsic goal orientation) were negative predictors for two types of intermediate assessment, suggesting that some types of intermediate assessment may be detrimental for students with higher levels of intrinsic motivation.

6.3 General findings

The overview of the literature in Chapter 2 shows the wide variety of ways in which intermediate assessment can be implemented in the curriculum. A closer look at the implementation at Leiden University in Chapters 3-5 also shows several different types of assessment characteristics. Furthermore, teachers' ideas about their ideal assessment in Chapter 4 show that there could be even more different assessment types, teachers had full control of the educational environment and its circumstances.

With regard to the perceptions of intermediate assessment, chapters 3 and 4 show that teachers generally have positive perceptions. Chapter 3 shows that teachers appreciate that intermediate assessment can help keep students on track and that they can measure a variety of knowledge and skills. Additionally, Chapter 4 shows that teachers are very focused on using intermediate assessment to develop student learning, and that they would focus on student learning even more in their ideal situation.

The effects of intermediate assessment show a less unequivocal picture. The literature review in Chapter 2 indicates that intermediate assessment generally has a positive effect on students' grades, but the results from the quantitative study described in Chapter 5 do not show any influence of assessment. In this study students' high school GPA was the main predictor of university grades. Whether a course had intermediate assessment or not, or what type of assessment this was, if there was one, did not seem to matter.

6.4 Reflections on the design of intermediate assessment

Since this dissertation is specifically focused on the characteristics of intermediate assessment, the following section will go deeper into the perceptions and results specifically related to two important assessment characteristics.

6.4.1 Assessment Type

A main finding is that the type of intermediate assessment is not a defining characteristic. This is evidenced in the literature review (Chapter 2) as well as in two of the empirical studies (Chapters 3 and 5). Both the review and Chapter 5 indicate that the assessment type does not influence student grades. In the review it is evidenced that for the majority of assessment types there are more studies showing positive effects of intermediate assessment than there are studies that do not portray these effects. There are some assessment types where this finding is reversed (like cases and exercises) and a few where there are only negative results. However, these are usually small groups of five or less studies. For the three largest assessment type groups, 'general assessment', quiz, and writing assignment, there is a similar ratio where two-thirds of the studies have found positive results. In Chapter 5, the lack of influence of assessment type is shown by the fact that the type of assessment is not a significant factor in the repeated measures ANCOVA, indicating that students get similar final exam scores regardless of the type of intermediate assessment employed in their course. Chapter 3 shows that this lack of influence of assessment type reaches further than just student results. Teachers across courses using different types of intermediate assessment all had similar perceptions about intermediate assessment. This indicates that, at least for assessment type, it does not really matter *how* you employ intermediate assessment. Results from Chapter 5 even

seem to suggest that it does not matter *whether* you use intermediate assessment, contrasting with results from previous research (e.g., Larsen, Butler, & Roediger, 2008; McDaniel, Anderson, Derbish, & Morrisette, 2007).

A possible explanation for the current lack of an effect for (different types of) intermediate assessment, that was also raised in Chapter 5, is that all types of intermediate assessment promote distributed studying. Dunlosky, Rawson, Marsh, Nathan, and Willingham (2013) have shown that distributed practice improves student results. Furthermore, distributed studying may also lead to increased study time, which in turn may also lead to better results (Doumen, Broeckmans, & Masui, 2014). However, since university curricula are usually set up with distributed educational meetings that have required readings, it stands to reason that the distributed practice effect is enhancing results in *all* university courses, not just those using intermediate assessment.

Another explanation could be that students did not perceive the intermediate assessment as aligned with, and therefore useful for, the final exam. Kahu (2013) proposes that assessment can influence students' engagement, and subsequently their study behaviour and results. Students' preference for a strong connection between the intermediate and final assessment already came up in Chapter 3, and it stands to reason that students are less engaged if they perceive this connection to be absent.

6.4.2 Use of Feedback

With regard to other characteristics, the importance of feedback is underlined. Chapter 2 shows that almost all studies using intermediate assessment provided some form of feedback, and that corrective feedback seemed the most positive. When addressing their beliefs about assessment in Chapter 4, teachers focused on student learning and the importance of providing feedback in this process.

Feedback is a construct that is thoroughly investigated, whether in the case of formative assessment (Nicol & Macfarlane-Dick, 2006), or when referring to learning and education in general (Hattie & Timperley, 2007). Gibbs and Simpson (2004) mention ten conditions that assessment has to meet to improve learning, seven of which focus on the influence of feedback. The recommendations for instructive feedback given by different authors often overlap. Several authors, for example, suggest providing task or process focused feedback (Gibbs & Simpson, 2004; Hattie & Timperley, 2007; Shute, 2008). The finding that corrective feedback may be sufficient is partly corroborated by

Shute (2008) who suggests that feedback should be simple but specific enough. Other suggestions include that feedback should inform students on how they can attain the desired performance (Nicol & Macfarlane-Dick, 2006; Hattie & Timperley, 2007, refer to this as feed-forward), and that teachers should be able to reshape their teaching with feedback information (Nicol & Macfarlane-Dick, 2006). Several teachers in Chapter 4 also referenced how feedback could help them improve their teaching.

6.5 Limitations

None of the chapters of this dissertation have established base-line measures, or control groups to compare results to. Establishing a baseline was not possible, because the educational reform was studied in real time during its roll-out. Similarly, in educational research, using true control groups is often difficult, because students are all offered the same curriculum.

A second limitation for all four studies described in Chapters 2 to 5 is issues with selection. Teachers and students could self-select for participation in the interviews and questionnaires. It is possible that teachers or students who felt very negative about the concept of intermediate assessment declined to participate, and therefore their opinions are not included in this dissertation. However, for both interview studies (Chapters 3 and 4) the teacher participation rate is approximately 70 percent, indicating that even if teachers who were very negative about intermediate assessment chose not to participate, the positive results portrayed in this dissertation do represent the majority of teachers. Student participation rates (Chapter 5) are much smaller, but the results show that participating students displayed a range of final exam grades. Similarly, the literature review in Chapter 2 suffers from its own selection issue: publication bias. In general, articles displaying positive results are more likely to be accepted for publication in scientific journals. Mahoney (1977), for example, already showed that reviewers deem articles with positive outcomes to be more methodologically sound. Shadish, Doherty, and Montgomery (1989) estimated that there may be as many unpublished studies in their field as there are published ones. Since the review did not include unpublished articles, we cannot be sure that the full spectrum of intermediate assessment characteristics in higher education is represented in the review.

In addition to the problems with selection, the three empirical studies all describe relatively small samples of teachers or students from a limited number of educational programs at Leiden University. These programs are all large scale, with few contact hours. It stands to reason that these results may not generalise to programs with a smaller number of students or more contact hours. Several STEM programs, for example, have lectures in the morning and practical labs in the afternoon, eliminating at least the need to introduce intermediate assessment as a means to get students to work. However, the review in Chapter 2 does describe studies from different educational fields, which often have positive results.

6.6 Implications for Practice

6.6.1 Encourage use of intermediate assessment

In general, intermediate assessment seems to be beneficial for students' performance. The overview of the literature presented in Chapter 2 evidences this. Teachers and students presented several benefits of intermediate assessment in Chapter 3, including the ability to measure different skills, and to take pressure of the final exam. Results from previous studies also show benefits when looking at affective or behavioural outcomes, intermediate assessment can, for example, be used to engage students (Holmes, 2015) or encourage them to spend more time studying (Admiraal, Wubbels & Pilot, 1999).

6.6.2 Minimise institutional assessment guidelines

The studies described in this dissertation suggest that a multitude of different assessment characteristics can influence student results. Therefore, it seems important to keep requirements for intermediate assessment as minimal as possible, to facilitate teachers' independence in deciding on the type and characteristics of the assessment they want to provide. When course coordinators were interviewed by Goos, Gannaway and Hughes (2011), one of their main problems with assessment was 'increasingly centralised and bureaucratic assessment requirements' (p. 99). The current results seem to advise against constraining assessment practice too firmly.

6.6.3 Support teachers' assessment practice

A main conclusion from Chapters 3 and 4 is that teachers are willing to implement intermediate assessment. Often, the teachers have positive perceptions and ambitious ideas on how they would like to incorporate intermediate assessment in the courses they are teaching. However, to keep this enthusiasm intact, it is important that the assessment practice is supported. Goos et al. (2011) identified that, on top of the problems with bureaucratic requirements discussed in the previous section, teachers' main problem with regard to assessment was related to workload and lack of tailored support.

Workload was identified as a constraining factor by teachers quoted in this dissertation as well. Several teachers mentioned, for example, the heavy workload of providing students with individual feedback. Most university courses are taught by staff whose primary task is research, creating a tension between research and teaching. When an institution wants to fully utilise the advantages of intermediate assessment, it should be prepared to invest in measures to alleviate workload. A utopian solution would be to increase the amount of teaching staff, but more realistically, institutions can invest in (digital) tools to support teachers' assessment practice. Several tools exist that can alleviate the burden of workload, for example by easing the feedback process.

Additionally, it is important to support teachers' knowledge about assessment. Postareff et al. (2012) suggested that teachers need to be made aware of the benefits of transformational assessment. Chapter 4 suggests that most teachers in the sample are already aware of these benefits, but nonetheless, assessment knowledge could be further supported. Some teachers quoted in Chapter 3 mentioned being unsure about what type of assessment was the best fit for their course, indicating that their assessment knowledge could be broadened. Samuelowicz and Bain (2002) mention that there is a large chance that several teachers enrolled in staff development classes will hold widely different beliefs on assessment, which may mean that a one size fits all assessment knowledge course will not work. A potent way to have teachers exchange their assessment knowledge is by using communities of practice, or another group-like constellation where teachers can learn from each other's experiences.

6.7 Directions for future research

Future studies into the characteristics of intermediate assessment can expand on this dissertation in several ways. First of all, this research can focus more quantitatively on the behavioural and affective outcomes of intermediate assessment. It could, for example, take a similar approach as in Chapter 5, but to measure more whether specific types of intermediate assessment elicit more student studying behaviour than others do.

A second direction for future research could focus further on the assessment types. The majority of studies cited in Chapter 2, and the studies conducted in Chapters 3 to 5, focused on comparing the effects of assessment types *between* courses. Perdigones, Garcia, Valino, & Raposo (2009), for example, investigated what they called comparable courses and found that a system of writing an article for a professional journal and presenting this article elicited higher pass rates for their course than a system of continuous assessment did. However, these 'comparable' courses differed in the number of credits, the university they were taught at, and the number of students enrolled. Furthermore, it stands to reason that several other course characteristics may differ between courses as well, thereby influencing the outcomes. Future studies could focus on comparing different assessment types *within* a course, either in a cohort design (keeping everything constant between cohorts, except for the assessment) or by comparing several groups in a quasi-experiment, where students are either assigned to a different assessment type or options for intermediate assessment are given, and students choose their preferred assessment type.

Results from Chapter 5 indicate that several student characteristics may influence the relationship between intermediate assessment and student results. Teachers could choose to offer tailored assessment for different student groups, and subsequent research could investigate whether this tailored assessment leads to improved performance for these groups. Moreover, teachers could offer several intermediate assessment options and let students choose their own assessment program. In this case, it is interesting to study whether different groups of students choose different types of assessment.

Finally, future research could further explore the influence of student characteristics. Richardson, Abrahams, and Bond (2012) identified 41 predictors of student achievement in their meta-analysis, of which only ten were discussed in the current dissertation. It stands to reason that several of the other

characteristics are also related to intermediate assessment and student outcomes. Future research could focus on (self-regulated) learning strategies (which are somewhat related to the motivation and self-regulation discussed in Chapter 5), or on personality traits, in a similar design as Chapter 5 to see whether these characteristics predict student results.

6.8 Concluding Remarks

Overall, it can be said that various types of assessment all have positive influences on student results, and that maybe the act of providing students with an intermediate assessment is more important than what the exact assessment is. University teachers often have thorough ideas and ideals about intermediate assessment, but often lack the means to put these into practice. Institutions should support teachers with assessment knowledge courses or opportunities to exchange assessment knowledge with colleagues. Additionally, they should minimise assessment constraints and simultaneously provide teachers with the freedom to design assessments that fit the courses that are being taught.

References

References

Articles marked by an * were included in the literature review.

- Admiraal, W., Wubbels, T., & Pilot, A. (1999). COLLEGE TEACHING IN LEGAL EDUCATION: Teaching Method, Students' Time-on-Task, and Achievement. *Research in Higher Education*, 40(6), 687-704.
- *Anziani, H., Durham, J., & Moore, U. (2008). The relationship between formative and summative assessment of undergraduates in oral surgery. *European Journal of Dental Education*, 12(4), 233-238. doi:10.1111/j.1600-0579.2008.00524.x.
- *Azzi, A. J., Ramnanan, C. J., Smith, J., Dionne, E., & Jalali, A. (2015). To quiz or not to quiz: Formative tests help detect students at risk of failing the clinical anatomy course. *Anatomical Sciences Education*, 8(5), 413-420. doi:10.1002/ase.1488.
- Bälter, O., Enstrom, E., & Klingenberg, B. (2013). The effect of short formative diagnostic web quizzes with minimal feedback. *Computers & Education*, 60(1), 234-242. doi:10.1016/j.compedu.2012.08.014.
- Baeten, M., Dochy, F., & Struyven, K. (2012). Using students' motivational and learning profiles in investigating their perceptions and achievement in case-based and lecture-based learning environments. *Educational Studies*, 38(5), 491-506. doi:10.1080/03055698.2011.643113.
- Baker, S. R. (2004). Intrinsic, extrinsic, and amotivational orientations: Their role in university adjustment, stress, well-being, and subsequent academic performance. *Current Psychology* 23(3): 189-202.
- *Baleghizadeh, S., & Zarghami, Z. (2012). The Impact of Conferencing Assessment on EFL Students' Grammar Learning. [Impacto de la evaluación mediante conferencias en el aprendizaje de la gramática en estudiantes de inglés como lengua extranjera]. *Profile Issues in Teachers' Professional Development*, 14(2), 131-144.
- Beaumont, C., O'Doherty, M., & Shannon, L. (2011). Reconceptualising assessment feedback: a key to improving student learning? *Studies in Higher Education*, 36(6), 671-687. doi:10.1080/03075071003731135.
- *Behrouzi, P., & Mowlavian, R. (2015). THE EFFECT OF ASSESSMENT PORTFOLIOS ON THE PARAGRAPH WRITING DEVELOPMENT OF IRANIAN INTERMEDIATE EFL LEARNERS. *Modern Journal of Language Teaching Methods*, 5(3), 306-320.
- Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5-25. doi:10.1080/0969594X.2010.513678
- van Berkel, H., Jansen, E., & Bax, A. (Eds.). (2012). *Studiesucces bevorderen: het kan en het is niet moeilijk [Improving Study Success: It's Possible and It's Not Difficult]*. The Hague, the Netherlands: Boom Lemma.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347-364. doi:10.1007/BF00138871.

- *Bijol, V., Byrne-Dugan, C. J., & Hoenig, M. P. (2015). Medical student web-based formative assessment tool for renal pathology. *Medical Education Online*, 20. doi:10.3402/meo.v20.26765.
- *Birjandi, P., & Hamidi Tamjid, N. (2012). The role of self-, peer and teacher assessment in promoting Iranian EFL learners' writing performance. *Assessment & Evaluation in Higher Education*, 37(5), 513-533. doi:10.1080/02602938.2010.549204.
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. doi:10.1080/0969595980050102.
- Black, P., & Wiliam, D. (2004). The Formative Purpose: Assessment Must First Promote Learning. *Yearbook of the National Society for the Study of Education*, 103(2), 20-50. doi:10.1111/j.1744-7984.2004.tb00047.x.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, 21(1), 5. doi:10.1007/s11092-008-9068-5.
- Blom, S., Severiens, S., Broekkamp, H., & Hoek, D. (2004). *Zelfstandig leren van allochtone en autochtone leerlingen in het Studiehuis [Independent learning of non-native and native Dutch students in the Studiehuis]*. Amsterdam: ILO, UvA.
- Bongers, V. (2015). Betutteling ligt op de loer: Studenten klagen over hoeveelheid Toetsen [Paternalism Looms: Students Complain About the Amount of Tests] *Mare*, February 19. <http://mareonline.nl>
- van den Bogaard, M. (2012). Explaining student success in engineering education at Delft University of Technology: a literature synthesis. *European Journal of Engineering Education*, 37(1), 59-82. doi:10.1080/03043797.2012.658507
- *Bookhan, V., Warren, N., & Schoeman, H. (2014). Do precursor tests influence the performance of Oral Hygiene students in subsequent semester tests? *SADJ: Journal of the South African Dental Association*, 69(10), 464-467.
- Bouwmeester, R. A. M., de Kleijn, R. A. M., Freriksen, A. W. M., van Emst, M. G., Veeneklaas, R. J., van Hoeij, M. J. W., ... van Rijen, H. V. M. (2013). Online formative tests linked to microlectures improving academic achievement. *Medical Teacher*, 35(12), 1044-1046. doi:10.3109/0142159X.2013.818633.
- Brookhart, S. M. (2001). Successful Students' Formative and Summative Uses of Assessment Information. *Assessment in Education: Principles, Policy & Practice*, 8(2), 153-169. doi:10.1080/09695940123775.
- *Brown, G. A., Bice, M. R., Shaw, B. S., & Shaw, I. (2015). Online quizzes promote inconsistent improvements on in-class test performance in introductory anatomy and physiology. *Advances in Physiology Education*, 39(2), 63-66. doi:10.1152/advan.00064.2014.
- *Buchanan, T. (2000). The efficacy of a World-Wide Web mediated formative assessment. *Journal of Computer Assisted Learning*, 16(3), 193-200. doi:10.1046/j.1365-2729.2000.00132.x.

- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1118-1133. doi:10.1037/a0019902.
- *Burch, V. C., Seggie, J. L., & Gary, N. E. (2006). Formative assessment promotes learning in undergraduate clinical clerkships. *Samj South African Medical Journal*, 96(5), 430-433.
- *Cacchione, A. (2015). Creative use of Twitter for Dynamic Assessment in Language Learning classroom at the university. *Interaction Design and Architectures* (24), 145-161.
- Cano, M. D. (2011). Students' Involvement in Continuous Assessment Methodologies: A Case Study for a Distributed Information Systems Course. *IEEE Transactions on Education*, 54(3), 442-451. doi:10.1109/TE.2010.2073708.
- *Carnegie, J. (2015). Use of Feedback-Oriented Online Exercises to Help Physiology Students Construct Well-Organized Answers to Short-Answer Questions. *Cbe-Life Sciences Education*, 14(3). doi:10.1187/cbe-14-08-0132.
- *Carrillo-de-la-Pena, M. T., Bailles, E., Caseras, X., Martinez, A., Ortet, G., & Perez, J. (2009). Formative assessment and academic achievement in pre-graduate students of health sciences. *Advances in Health Sciences Education*, 14(1), 61-67. doi:10.1007/s10459-007-9086-y.
- *Carrillo-de-la-Pena, M. T., & Perez, J. (2012). Continuous Assessment Improved Academic Achievement and Satisfaction of Psychology Students in Spain. *Teaching of Psychology*, 39(1), 45-47. doi:10.1177/0098628311430312.
- Clouder, L., Broughan, C., Jewell, S., & Steventon, G. (2012). *Improving student engagement and development through assessment*. London: Routledge.
- Cohen-Schotanus, J. (1999). Student assessment and examination rules. *Medical Teacher*, 21(3), 318-321. doi:10.1080/01421599979626.
- *Covic, T., & Jones, M. K. (2008). Is the essay resubmission option a formative or a summative assessment and does it matter as long as the grades improve? *Assessment & Evaluation in Higher Education*, 33(1), 75-85. doi:10.1080/02602930601122928.
- *Cruz, É., Dias, H., & Kortemeyer, G. (2011). The effect of formative assessment in Brazilian university physics courses. [Efeito da avaliação formativa em cursos de física em universidades brasileiras]. *Revista Brasileira de Ensino de Física*, 33(4), 4315-4315. doi:10.1590/s1806-11172011000400016.
- *Cukusic, M., Garaca, Z., & Jadric, M. (2014). Online self-assessment and students' success in higher education institutions. *Computers & Education*, 72, 100-109. doi:10.1016/j.compedu.2013.10.018.
- Daw, N. D., & Frank, M. J. (2009). Reinforcement learning and higher level cognition: Introduction to special issue. *Cognition*, 113(3), 259-261. doi:https://doi.org/10.1016/j.cognition.2009.09.005

- Day, I. N. Z., van Blankenstein, F. M., Westenberg, M., & Admiraal, W. (2016, May). *Literatuurreview: Kenmerken van tussentijds toetsen in het hoger onderwijs en hun relatie met opbrengst bij studenten. [Literature review: Characteristics of intermediate assessment in higher education and their relation to student outcomes.]* Roundtable session presented at the annual meeting of the Netherlands Educational Research Association, Rotterdam, the Netherlands.
- De Paola, M., & Scoppa, V. (2011). Frequency of examinations and student achievement in a randomized experiment. *Economics of Education Review*, 30(6), 1416-1429. doi:10.1016/j.econedurev.2011.07.009.
- *Deane, R. P., Joyce, P., & Murphy, D. J. (2015). Team Objective Structured Bedside Assessment (TOSBA) as formative assessment in undergraduate Obstetrics and Gynaecology: a cohort study. *Bmc Medical Education*, 15. doi:10.1186/s12909-015-0456-5.
- Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, 11(4), 227-268. doi:10.1207/S15327965PLI1104_01.
- van Dinther, M., Dochy, F., Segers, M., & Braeken, J. (2014). Student perceptions of assessment and student self-efficacy in competence-based education. *Educational Studies*, 40(3), 330-351. doi:10.1080/03055698.2014.898577.
- Dirkx, K. J. H. (2014). *Putting the Testing-effect to the Test: Why and When is Testing effective for Learning in Secondary School?* Maastricht, The Netherlands: Datawyse, Universitaire Pers Maastricht.
- Dirkx, K. J. H., Kester, L., & Kirschner, P. A. (2014). The Testing Effect for Learning Principles and Procedures from Texts. *The Journal of Educational Research*, 107, 357-364. doi:10.1080/00220671.2013.823370
- *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. doi:10.1152/advan.90162.2008.
- *Domenech, J., Blazquez, D., de la Poza, E., & Munoz-Miquel, A. (2015). Exploring the impact of cumulative testing on academic performance of undergraduate students in Spain. *Educational Assessment Evaluation and Accountability*, 27(2), 153-169. doi:10.1007/s11092-014-9208-z.
- Doumen, S., Broeckmans, J., & Masui, C. (2014). The role of self-study time in freshmen's achievement. *Educational Psychology*, 34(3), 385-402. doi:10.1080/01443410.2013.785063.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Psychological Science in the Public Interest*, 14(1), 4-58. doi:10.1177/1529100612453266.
- *English, R., Brookes, S. T., Avery, K., Blazeby, J. M., & Ben-Shlomo, Y. (2006). The effectiveness and reliability of peer-marking in first-year medical students. *Medical Education*, 40(10), 965-972. doi:10.1111/j.1365-2929.2006.02565.x.

- Entwistle, N. J. Approaches to learning and perceptions of the learning environment. *Higher Education*, 22(3), 201-204. doi:10.1007/bf00132287.
- *Ernst, H., & Colthorpe, K. (2008). Expanding voluntary active-learning opportunities for pharmacy students in a respiratory physiology module. *American Journal of Pharmaceutical Education*, 72(2).
- *Fakcharoenphol, W., Potter, E., & Stelzer, T. (2011). What students learn when studying physics practice exam problems. *Physical Review Special Topics-Physics Education Research*, 7(1). doi:10.1103/PhysRevSTPER.7.010107.
- *Fautch, J. M. (2015). The flipped classroom for teaching organic chemistry in small classes: is it effective? *Chemistry Education Research and Practice*, 16(1), 179-186. doi:10.1039/c4rp00230j.
- *Fisher, R., Cavanagh, J., & Bowles, A. (2011). Assisting transition to university: using assessment as a formative learning tool. *Assessment & Evaluation in Higher Education*, 36(2), 225-237. doi:10.1080/02602930903308241.
- *Gauci, S. A., Dantas, A. M., Williams, D. A., & Kemm, R. E. (2009). Promoting student-centered active learning in lectures with a personal response system. *Advances in Physiology Education*, 33(1), 60-71. doi:10.1152/advan.00109.2007.
- Gibbs, G., & Simpson, C. (2004). Conditions Under Which Assessment Supports Students' Learning. *Learning and Teaching in Higher Education*, 1(1), 3-31.
- *Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: a multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning*, 31(5), 435-449. doi:10.1111/jcal.12096.
- *Gill, M., & Greenhow, M. (2008). How effective is feedback in Computer-Aided Assessments? *Learning Media and Technology*, 33(3), 207-220. doi:10.1080/17439880802324145.
- *Glass, A. L. (2009). The effect of distributed questioning with varied examples on exam performance on inference questions. *Educational Psychology*, 29(7), 831-848. doi:10.1080/01443410903310674.
- Goos, M., Gannaway, D., & Hughes, C. (2011). Assessment as an equity issue in higher education: comparing the perceptions of first year students, course coordinators, and academic leaders. *The Australian Educational Researcher*, 38(1), 95-107. doi:10.1007/s13384-010-0008-2.
- *Gums, T. H., Kleppinger, E. L., & Urlick, B. Y. (2014). Outcomes of Individualized Formative Assessments in a Pharmacy Skills Laboratory. *American Journal of Pharmaceutical Education*, 78(9).
- *Haffling, A.-C., Beckman, A., & Edgren, G. (2011). Structured feedback to undergraduate medical students: 3 years' experience of an assessment tool. *Medical Teacher*, 33(7), E349-E357. doi:10.3109/0142159x.2011.577466.
- *Hancock, T. M. (2010). Use of audience response systems for summative assessment in large classes. *Australasian Journal of Educational Technology*, 26(2), 226-237.

- *Hardy, J., Bates, S. P., Casey, M. M., Galloway, K. W., Galloway, R. K., Kay, A. E., . . . McQueen, H. A. (2014). Student-Generated Content: Enhancing learning through sharing multiple-choice questions. *International Journal of Science Education*, 36(13), 2180-2194. doi:10.1080/09500693.2014.916831.
- Harland, T., McLean, A., Wass, R., Miller, E., & Sim, K. N. (2015). An assessment arms race and its fallout: high-stakes grading and the case for slow scholarship. *Assessment & Evaluation in Higher Education*, 40(4), 528-541. doi:10.1080/02602938.2014.931927.
- *Harwood, I. (2006). Managing summative computer-aided assessment failures: further evidence for 'optional substitution'. *British Journal of Educational Technology*, 37(4), 637-641. doi:10.1111/j.1467-8535.2006.00583.x.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81-112. doi:10.3102/003465430298487.
- *Henly, D. C. (2003). Use of Web-based formative assessment to support student learning in a metabolism/nutrition unit. *European journal of dental education: official journal of the Association for Dental Education in Europe*, 7(3), 116-122. doi:10.1034/j.1600-0579.2003.00310.x.
- *Henly, D. C., & Reid, A. E. (2001). Use of the web to provide learning support for a large metabolism and nutrition class. *Biochemistry and Molecular Biology Education*, 29(6), 229-233.
- *Hernick, M. (2015). Test-Enhanced Learning in an Immunology and Infectious Disease Medicinal Chemistry/Pharmacology Course. *American Journal of Pharmaceutical Education*, 79(7).
- Holmes, N. (2015). Student perceptions of their learning and engagement in response to the use of a continuous e-assessment in an undergraduate module. *Assessment & Evaluation in Higher Education*, 40(1), 1-14. doi:10.1080/02602938.2014.881978.
- *Hooshyar, D., Ahmad, R. B., Yousefi, M., Fathi, M., Horng, S.-J., & Lim, H. (2016). Applying an online game-based formative assessment in a flowchart-based intelligent tutoring system for improving problem-solving skills. *Computers & Education*, 94, 18-36.
- *Ibabe, I., & Jauregizar, J. (2010). Online self-assessment with feedback and metacognitive knowledge. *Higher Education*, 59(2), 243-258. doi:10.1007/s10734-009-9245-6.
- *Isaksson, S. (2008). Assess as you go: the effect of continuous assessment on student learning during a short course in archaeology. *Assessment & Evaluation in Higher Education*, 33(1), 1-7. doi:10.1080/02602930601122498.
- *Jacoby, J. C., Heugh, S., Bax, C., & Branford-White, C. (2014). Enhancing learning through formative assessment. *Innovations in Education and Teaching International*, 51(1), 72-83. doi:10.1080/14703297.2013.771970.
- *Jain, V., Agrawal, V., & Biswas, S. (2012). Use of formative assessment as an educational tool. *Journal of Ayub Medical College, Abbottabad: JAMC*, 24(3-4), 68-70.

- Jessop, T., El Hakim, Y., & Gibbs, G. (2014). The whole is greater than the sum of its parts: a large-scale study of students' learning in response to different programme assessment patterns. *Assessment & Evaluation in Higher Education*, 39(1), 73-88. doi:10.1080/02602938.2013.792108.
- Kahu, E. R. (2013). Framing student engagement in higher education. *Studies in Higher Education*, 38(5), 758-773. doi:10.1080/03075079.2011.598505
- Kamphorst, J. C., Hofman, W. H. A., Jansen, E. P. W. A., & Terlouw, C. (2013). The relationship between perceived competence and earned credits in competence-based higher education. *Assessment & Evaluation in Higher Education*, 38(6), 646-661. doi:10.1080/02602938.2012.680015.
- Karpicke, J. D., Butler, A. C., & Roediger, H. L., III (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, 17(4), 471-479. doi:10.1080/09658210802647009.
- Karpicke, J. D., & Roediger, H. L., III. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 704-719. doi:10.1037/0278-7393.33.4.704
- *Kennelly, B., Considine, J., & Flannery, D. (2011). Online Assignments in Economics: A Test of Their Effectiveness. *Journal of Economic Education*, 42(2), 136-146. doi:10.1080/00220485.2011.555696.
- *Kerdijk, W., Tio, R. A., Mulder, B. F., & Cohen-Schotanus, J. (2013). Cumulative assessment: strategic choices to influence students' study effort. *Bmc Medical Education*, 13. doi:10.1186/1472-6920-13-172.
- *Kibble, J. (2007). Use of unsupervised online quizzes as formative assessment in a medical physiology course: effects of incentives on student participation and performance. *Advances in Physiology Education*, 31(3), 253-260. doi:10.1152/advan.00027.2007.
- *Kibble, J. D., Johnson, T. R., Khalil, M. K., Nelson, L. D., Riggs, G. H., Borrero, J. L., & Payer, A. F. (2011). Insights Gained from the Analysis of Performance and Participation in Online Formative Assessment. *Teaching and Learning in Medicine*, 23(2), 125-129. doi:10.1080/10401334.2011.561687.
- *Kim, M., & Ryu, J. (2013). The development and implementation of a web-based formative peer assessment system for enhancing students' metacognitive awareness and performance in ill-structured tasks. *Educational Technology Research and Development*, 61(4), 549-561. doi:10.1007/s11423-012-9266-1.
- Kingston, N., & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice*, 30(4), 28-37. doi:10.1111/j.1745-3992.2011.00220.x
- *van der Kleij, F. M., Eggen, T. J. H. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, 58(1), 263-272. doi:10.1016/j.compedu.2011.07.020.

- van der Kleij, F. M., Vermeulen, J. A., Schildkamp, K., & Eggen, T. J. H. M. (2015). Integrating data-based decision making, Assessment for Learning and diagnostic testing in formative assessment. *Assessment in Education: Principles, Policy & Practice*, 22(3), 324-343. doi:10.1080/0969594X.2014.999024
- de Kleijn, R. A. M., Bouwmeester, R. A. M., Ritzen, M. M. J., Ramaekers, S. P. J., & Van Rijen, H. V. M. (2013). Students' motives for using online formative assessments when preparing for summative assessments. *Medical Teacher*, 35(12), E1644-E1650. doi:10.3109/0142159x.2013.826794
- Kling, N., McCorkle, D., Miller, C., & Reardon, J. (2005). The Impact of Testing Frequency on Student Performance in a Marketing Course. *Journal of Education for Business*, 81(2), 67-72.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284. doi:10.1037/0033-2909.119.2.254.
- Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, 23(9), 1297-1317. doi:10.1002/acp.1537.
- *Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell biology education*, 4(4), 298-310. doi:10.1187/05-06-0082.
- *Krasne, S., Wimmers, P. F., Relan, A., & Drake, T. A. (2006). Differential effects of two types of formative assessment in predicting performance of first-year medical students. *Advances in Health Sciences Education*, 11(2), 155-171. doi:10.1007/s10459-005-5290-9.
- Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory into Practice*, 41(4), 212-218. doi:10.1207/s15430421tip4104_2.
- *Krontiris-Litowitz, J. (2008). Using truncated lectures, conceptual exercises, and manipulatives to improve learning in the neuroanatomy classroom. *Advances in Physiology Education*, 32(2), 152-156. doi:10.1152/advan.00103.2007.
- Larsen, D. P., Butler, A. C., & Roediger, H. L., III. (2008). Test-enhanced learning in medical education. *Medical Education*, 42(10), 959-966. doi:10.1111/j.1365-2923.2008.03124.x
- *Lawrence, N. K. (2013). Cumulative Exams in the Introductory Psychology Course. *Teaching of Psychology*, 40(1), 15-19. doi:10.1177/0098628312465858.
- Leeming, F. C. (2002). The Exam-A-Day Procedure Improves Performance in Psychology Classes. *Teaching of Psychology*, 29(3), 210-212.
- *Lemus-Zuniga, L. G., Montanana, J. M., Buendia-Garcia, F., Poza-Lujan, J. L., Posadas-Yaguee, J. L., & Benlloch-Dualde, J. V. (2015). Computer-assisted method based on continuous feedback to improve the academic achievements of first-year students on computer engineering. *Computer Applications in Engineering Education*, 23(4), 610-620. doi:10.1002/cae.21633.
- Lizzio, A., Wilson, K., & Simons, R. (2002). University Students' Perceptions of the Learning Environment and Academic Outcomes: Implications for theory and practice. *Studies in Higher Education*, 27(1), 27-52. doi:10.1080/03075070120099359.

- *MacDermott, R. J. (2013). The Impact of Assessment Policy on Learning: Replacement Exams or Grade Dropping. *Journal of Economic Education*, 44(4), 364-371. doi:10.1080/00220485.2013.825114.
- Macfarlane, B. (2015). Student performativity in higher education: converting learning as a private space into a public performance. *Higher Education Research & Development*, 34(2), 338-350. doi:10.1080/07294360.2014.956697.
- Machin, S., & McNally, S. (2005). Gender and Student Achievement in English Schools. *Oxford Review of Economic Policy*, 21(3), 357-372. doi:10.1093/oxrep/gri021.
- Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, 1(2), 161-175. doi:10.1007/bf01173636
- *Maley, M. A. L., Harvey, J. R., De Boer, W. B., Scott, N. W., & Arena, G. E. (2008). Addressing current problems in teaching pathology to medical students: blended learning. *Medical Teacher*, 30(1), E1-E9. doi:10.1080/01421590701753575.
- *Marden, N. Y., Ulman, L. G., Wilson, F. S., & Velan, G. M. (2013). Online feedback assessments in physiology: effects on students' learning experiences and outcomes. *Advances in Physiology Education*, 37(2), 192-200. doi:10.1152/advan.00092.2012.
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19(4-5), 494-513. doi:10.1080/09541440701326154.
- McDaniel, M. A., K.M. Wildman, and J.L Anderson. 2012. "Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study." *Journal of Applied Research in Memory and Cognition* 1: 18-26. doi: 10.1016/j.jarmac.2011.10.001.
- McKenzie, K., & Schweitzer, R. (2001). Who Succeeds at University? Factors predicting academic performance in first year Australian university students. *Higher Education Research & Development*, 20(1), 21-33. doi:10.1080/07924360120043621
- *McNulty, J. A., Espiritu, B. R., Hoyt, A. E., Ensminger, D. C., & Chandrasekhar, A. J. (2015). Associations Between Formative Practice Quizzes and Summative Examination Outcomes in a Medical Anatomy Course. *Anatomical Sciences Education*, 8(1), 37-44. doi:10.1002/ase.1442.
- Michaels, J. W., & Miethe, T. D. (1989). Academic Efforts and College Grades. *Social Forces*, 68(1), 309.
- *Micklewright, D., Pearsall, L., Sellens, M., & Billam, N. (2010). Changes in approaches to learning among undergraduate sports science students following a programme of weekly online assessments. *Journal of Hospitality Leisure Sport & Tourism Education*, 9(2), 141-155. doi:10.3794/johlste.92.216.
- *Mitra, N. K., & Barua, A. (2015). Effect of online formative assessment on summative performance in integrated musculoskeletal system module. *Bmc Medical Education*, 15. doi:10.1186/s12909-015-0318-1.

- Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the Gender Achievement Gap in College Science: A Classroom Study of Values Affirmation. *Science*, 330 (6008), 1234-1237. doi:10.1126/science.1195996
- Moon, J.A. (1999). *Reflection in learning & professional development: theory & practice*. London: Kogan Page.
- *Mulder, R., Baik, C., Naylor, R., & Pearce, J. (2014). How does student peer review influence perceptions, engagement and academic outcomes? A case study. *Assessment & Evaluation in Higher Education*, 39(6), 657-677. doi:10.1080/02602938.2013.860421.
- *Munson, A., & Pierce, R. (2015). Flipping Content to Improve Student Examination Performance in a Pharmacogenomics Course. *American Journal of Pharmaceutical Education*, 79(7).
- Myers, C. B., & Myers, S. M. (2015). The use of learner-centered assessment practices in the United States: the influence of individual and institutional contexts. *Studies in Higher Education*, 40(10), 1904-1918. doi:10.1080/03075079.2014.914164.
- *Nelson, J., Robison, D. F., Bell, J. D., & Bradshaw, W. S. (2009). Cloning the Professor, an Alternative to Ineffective Teaching in a Large Course. *Cbe-Life Sciences Education*, 8(3), 252-263. doi:10.1187/cbe.09-01-0006.
- Neroni, J., Meijs, C., Leontjevas, R., Kirschner, P. A., & De Groot, R. H. M. (2017). *Goal Orientation and Academic Performance in Adult Distance Education*. Manuscript submitted for publication
- *Nicol, D. (2009). Assessment for learner self-regulation: enhancing achievement in the first year using learning technologies. *Assessment & Evaluation in Higher Education*, 34(3), 335-352. doi:10.1080/02602930802255139.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218. doi:10.1080/03075070600572090
- *Olson, B. L., & McDonald, J. L. (2004). Influence of online formative assessment upon student learning in biomedical science courses. *Journal of Dental Education*, 68(6), 656-659.
- *Palmer, E. J., & Devitt, P. G. (2008). Limitations of student-driven formative assessment in a clinical clerkship. A randomised controlled trial. *Bmc Medical Education*, 8. doi:10.1186/1472-6920-8-29.
- *Palmer, E., & Devitt, P. (2014). The assessment of a structured online formative assessment program: a randomised controlled trial. *Bmc Medical Education*, 14. doi:10.1186/1472-6920-14-8.
- Palmer, E. L. (1974). FREQUENCY OF TESTS AND GENERAL SUBJECT-AREA MASTERY. *Psychological Reports*, 35(1), 422-422. doi:10.2466/pr0.1974.35.1.422
- *Paschal, C. B. (2002). Formative assessment in physiology teaching using a wireless classroom communication system. *Advances in Physiology Education*, 26(4), 299-308. doi:10.1152/advan.00030.2002.

- Peat, M., & Franklin, S. (2002). Supporting student learning: the use of computer-based formative assessment modules. *British Journal of Educational Technology*, 33(5), 515-523. doi:10.1111/1467-8535.00288
- *Perdigones, A., Garcia, J. L., Valino, V., & Raposo, C. (2009). Assessing heterogeneous student bodies using a methodology that encourages the acquisition of skills valued by employers. *Assessment & Evaluation in Higher Education*, 34(4), 389-400. doi:10.1080/02602930802071056.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (Mslq). *Educational and Psychological Measurement*, 53(3), 801-813. doi:10.1177/0013164493053003024
- Popham, W. J. (2001). Teaching to the Test? *Educational leadership*, 58(6), 16-21.
- Postareff, L., Virtanen, V., Katajavuori, N., & Lindblom-Ylänne, S. (2012). Academics' conceptions of assessment and their assessment practices. *Studies in Educational Evaluation*, 38(3), 84-92. doi:http://dx.doi.org/10.1016/j.stueduc.2012.06.003.
- *Priya, M., Muthu, M. S., Amarlal, D., & Thomas, E. (2012). Continuous Assessment of Undergraduate Students at a Dental College in India. *Journal of Dental Education*, 76(4), 501-508.
- Prosser, M., Walker, P., & Millar, R. (1996). Differences in students' perceptions of learning physics. *Physics Education*, 31(1), 43.
- Qenani, E., MacDougall, N., & Sexton, C. (2014). An empirical study of self-perceived employability: Improving the prospects for student employment success in an uncertain environment. *Active Learning in Higher Education*, 15(3), 199-213. doi:10.1177/1469787414544875.
- *Ramirez, B. U. (2010). Effect of self-assessment on test scores: student perceptions. *Advances in Physiology Education*, 34(3), 134-136. doi:10.1152/advan.00021.2010.
- *Raupach, T., Brown, J., Anders, S., Hasenfuss, G., & Harendza, S. (2013). Summative assessments are more powerful drivers of student learning than resource intensive teaching formats. *Bmc Medicine*, 11. doi:10.1186/1741-7015-11-61.
- Reimann, N., & Sadler, I. (2017). Personal understanding of assessment and the link to assessment practice: the perspectives of higher education staff. *Assessment & Evaluation in Higher Education*, 42(5), 724-736. doi:10.1080/02602938.2016.1184225.
- *Rezaei, A. R. (2015). Frequent collaborative quiz taking and conceptual learning. *Active Learning in Higher Education*, 16(3), 187-196. doi:10.1177/1469787415589627.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353-387. doi:10.1037/a0026838.
- Roediger, H. L., III, & Karpicke, J. D. (2006). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, 1(3), 181-210. doi:10.1111/j.1745-6916.2006.00012.x

- Sadler, I., & Reimann, N. (2018). Variation in the development of teachers' understandings of assessment and their assessment practices in higher education. *Higher Education Research & Development*, 37(1), 131-144. doi:10.1080/07294360.2017.1344199.
- *Sample, S. J., Bleedorn, J. A., Schaefer, S. L., Mikla, A., Olsen, C. W., & Muir, P. (2014). Students' Perception of Case-Based Continuous Assessment and Multiple-Choice Assessment in a Small Animal Surgery Course for Veterinary Medical Students. *Veterinary Surgery*, 43(4), 388-399. doi:10.1111/j.1532-950X.2014.12145.x.
- Samuelowicz, K., & Bain, J. D. (2002). Identifying academics' orientations to assessment practice. *Higher Education*, 43(2), 173-201. doi:10.1023/a:1013796916022.
- *Sargent, C. S., & Curcio, A. A. (2012). Empirical Evidence that Formative Assessments Improve Final Exams. *Journal of Legal Education*, 61(3), 379-405.
- *Schlegel, E. F. M., & Selfridge, N. J. (2014). Fun, collaboration and formative assessment: Skinquization, a class wide gaming competition in a medical school with a large class. *Medical Teacher*, 36(5), 447-449. doi:10.3109/0142159x.2014.888409.
- Segers, M., & Tillema, H. (2011). How do Dutch secondary teachers and students conceive the purpose of assessment? *Studies in Educational Evaluation*, 37(1), 49-54. doi: 10.1016/j.stueduc.2011.03.008.
- Shadish, W. R., Doherty, M., & Montgomery, L. M. (1989). How many studies are in the file drawer? An estimate from the family/marital psychotherapy literature. *Clinical Psychology Review*, 9(5), 589-603. doi:https://doi.org/10.1016/0272-7358(89)90013-5
- Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78(1), 153-189. doi:10.3102/0034654307313795.
- Sin, C., Tavares, O., & Amaral, A. (2017). Accepting employability as a purpose of higher education? Academics' perceptions and practices. *Studies in Higher Education*, 1-12. doi:10.1080/03075079.2017.1402174.
- *Siweya, H. J., & Letsoalo, P. (2014). Formative assessment by first-year chemistry students as predictor of success in summative assessment at a South African university. *Chemistry Education Research and Practice*, 15(4), 541-549. doi:10.1039/c4rp00032c.
- *Snowball, J. D., & Mostert, M. (2013). Dancing with the devil: formative peer assessment and academic performance. *Higher Education Research & Development*, 32(4), 646-659. doi:10.1080/07294360.2012.705262.
- *Stellmack, M. A., Keenan, N. K., Sandidge, R. R., Sippl, A. L., & Konheim-Kalkstein, Y. L. (2012). Review, Revise, and Resubmit: The Effects of Self-Critique, Peer Review, and Instructor Feedback on Student Writing. *Teaching of Psychology*, 39(4), 235-244. doi:10.1177/0098628312456589.
- *Sullivan, P. B., Gregg, N., Adams, E., Rodgers, C., & Hull, J. (2013). How much of the paediatric core curriculum do medical students remember? *Advances in Health Sciences Education*, 18(3), 365-373. doi:10.1007/s10459-012-9375-y.

- *Taradi, S. K., Taradi, M., & Radic, K. I. (2005). Integration of online formative assessments into medical education: Experience from University of Zagreb Medical School, Croatia. *National Medical Journal of India*, 18(1), 39-40.
- Taras, M. (2009). Summative assessment: the missing link for formative assessment. *Journal of Further and Higher Education*, 33(1), 57-69. doi:10.1080/03098770802638671
- *Thome, G., Hovenberg, H., & Edgren, G. (2006). Portfolio as a method for continuous assessment in an undergraduate health education programme. *Medical Teacher*, 28(6), E171-E176. doi:10.1080/01421590600776511.
- Thomas, J. A., Wadsworth, D., Jin, Y., Clarke, J., Page, R., & Thunders, M. (2017). Engagement with online self-tests as a predictor of student success. *Higher Education Research & Development*, 36(5), 1061-1071. doi:10.1080/07294360.2016.1263827
- Thomas, L. (2002). Student retention in higher education: the role of institutional habitus. *Journal of Education Policy*, 17(4), 423-442. doi:10.1080/02680930210140257
- Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research*, 45(1), 89-125. doi:10.3102/00346543045001089
- Topping, K. (1998). Peer Assessment Between Students in Colleges and Universities. *Review of Educational Research*, 68(3), 249-276. doi:10.3102/00346543068003249.
- Torres-Guijarro, S., & Bengoechea, M. (2017). Gender differential in self-assessment: a fact neglected in higher education peer and self-assessment techniques. *Higher Education Research & Development*, 36(5), 1072-1084. doi:10.1080/07294360.2016.1264372
- *Tuunila, R., & Pulkkinen, M. (2015). Effect of continuous assessment on learning outcomes on two chemical engineering courses: case study. *European Journal of Engineering Education*, 40(6), 671-682. doi:10.1080/03043797.2014.1001819.
- *Velan, G. M., Jones, P., McNeil, H. P., & Kumar, R. K. (2008). Integrated online formative assessments in the biomedical sciences for medical students: benefits for learning. *Bmc Medical Education*, 8. doi:10.1186/1472-6920-8-52.
- van der Vleuten, C. P. M., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., & van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher*, 34(3), 205-214. doi:10.3109/0142159X.2012.652239.
- Vos, P. (1992). Het ritme van het Rooster [The Rythm of the Schedule]. *Onderzoek van Onderwijs* 21: 51-53.
- *Wakefield, C., Adie, J., Pitt, E., & Owens, T. (2014). Feeding forward from summative assessment: The Essay Feedback Checklist as a learning tool. *Assessment & Evaluation in Higher Education*, 39(2), 253-262. doi:10.1080/02602938.2013.822845.
- Watkins, D., B. Dahlin & M. Ekholm (2005). "Awareness of the backwash effect of assessment: A phenomenographic study of the views of Hong Kong and Swedish lecturers." *Instructional Science* 33(4): 283-309. DOI: 10.1007/s11251-005-3002-8.
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What Influences Learning? A Content Analysis of Review Literature. *Journal of Educational Research*, 84(1), 30.

- Wiliam, D. (2011). "What is assessment for learning?" *Studies in Educational Evaluation* 37(1): 3-14. DOI: <https://doi.org/10.1016/j.stueduc.2011.03.001>.
- *Williams, N. A., Bland, W., & Christie, G. (2008). Improving student achievement and satisfaction by adopting a blended learning approach to inorganic chemistry. *Chemistry Education Research and Practice*, 9(1), 43-50. doi:10.1039/b801290n.
- *Wilson, K., Boyd, C., Chen, L., & Jamal, S. (2011). Improving student performance in a first-year geography course: Examining the importance of computer-assisted formative assessment. *Computers & Education*, 57(2), 1493-1500. doi:10.1016/j.compedu.2011.02.011.
- *Wininger, S. R. (2005). Using your tests to teach: Formative summative assessment. *Teaching of Psychology*, 32(3), 164-166. doi:10.1207/s15328023top3203_7.
- *Zhang, N., & Henderson, C. N. R. (2015). Can formative quizzes predict or improve summative exam performance? *Journal of Chiropractic Education*, 29(1), 16-21. doi:10.7899/jce-14-12.
- Zijlstra, H. (2012). *Prestatieafspraken met Hogescholen en Universiteiten*. [Performance agreements with Universities of Applied Sciences and Research Universities]. Retrieved from <http://www.rijksoverheid.nl/documenten-en-publicaties/kamerstukken/2012/03/07/kamerbrief-over-prestatieafspraken-met-hogescholen-en-universiteiten.html>.
- van Zundert, M., Sluijsmans, D., & van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, 20(4), 270-279. doi:<https://doi.org/10.1016/j.learninstruc.2009.08.004>.

Summary

Introduction

Assessment is an integral part of higher education and is often used to measure learning (summative assessment) as well as to support learning (formative assessment). In practice, summative and formative processes often overlap, by measuring students' knowledge and subsequently providing feedback to improve this knowledge. One way to use assessment is intermediate assessment, assessing students throughout the semester instead of just in the final exam week. Intermediate assessment can take various forms, and research has indicated that it can improve student performance. Therefore, it was introduced at Leiden University as a measure to improve student success.

Intermediate assessment can influence cognitive (e.g. grades), affective (e.g. motivation), and behavioural outcomes (e.g. study time). The three outcomes are often intertwined, for example in the case that students who are more engaged in a course will spend more time studying and subsequently get higher grades.

Intermediate assessment may improve students' results through four mechanisms. The first two are the testing and spacing effects. The testing effect posits that repeated testing creates stronger retention in memory than repeated studying does. The spacing effect, or distributed practice, entails that spreading study time through the semester strengthens retention, compared to cramming. These two effects can reinforce each other. Research has indicated that distributed testing has better results than cramming using testing. The other two mechanisms are increased time-on-task, and time for reflection. Intermediate assessment encourages students to spend more time on studying, which subsequently leads to better grades. Furthermore, getting results from intermediate assessment allows students time to reflect on their learning, which can also improve performance.

There are several ways to implement intermediate assessment. There is a variety of different characteristics, like, for example, the amount of assessments, the frequency of assessment, the type of assessment, or the rewards for the assessment (for a full overview, see Table 1.1.). At Leiden University, intermediate assessment was introduced with minimal constraints, leaving teachers with sufficient freedom to vary the different characteristics. At Leiden Law School, teachers teaching a 10 EC course were obligated to offer a partial exam, but all other teachers were free to choose whether they wanted to offer intermediate assessment. At the institute of Psychology, teachers needed to

offer plural assessment, i.e. several different assessment types, and teachers often chose some form of intermediate assessment as their plural assessment.

In this dissertation I try to answer the question ‘in what ways can intermediate assessment in higher education be designed to improve study behaviour and study results?’. In answering this question, each chapter of the dissertation focuses on at least one of the following three focal points central to the investigation. The first is the implementation of intermediate assessment, with a specific focus on different characteristics (which comes up in Chapters 2-5). The second is student and teacher perceptions of intermediate assessment (Chapters 3 & 4), and the third is the effects of intermediate assessment on student grades (Chapters 2 & 5).

Chapter 2

This chapter describes the results from a literature review answering the question ‘What characteristics of intermediate assessment are related to student grades?’. The focal points of the investigation central to this chapter are implementation and results. In total, 88 empirical articles were included in the review. For all these articles, eleven different assessment characteristics were extracted. These are the type of assessment, the frequency of assessment, the duration of the assessment, the number and type of questions, the scoring of the assessment, the type of feedback students get, who the assessor is, what the assessment medium is, where the assessment took place, whether the assessment was mandatory, and any other relevant characteristics.

Results indicate that intermediate assessment is studied around the world, but the majority of studies was conducted in the United States of America, the United Kingdom, and Australia. When looking at disciplines Medicine is a main contender. Other large disciplines are STEM, Social Sciences, and Law, Business & Economics. In total, the 88 articles describe 24 different assessment types. The most prevalent types are midterms/exams, quizzes, and writing assignments, all three ‘generic’ assessment types that can be employed across disciplines. Only quizzes do not occur across all disciplines in the current sample. There is a large variety in frequency of the assessment, ranging from 1 to 38 intermediate assessments in a single course, although the majority of courses has between one and ten assessments.

Four assessment characteristics were chosen to discuss in more detail, using exemplar articles. These are the use of feedback, whether an assessment is mandatory, who the assessor is, and what the reward for the assessment was.

First, feedback is provided in some way in the majority of the studies. This feedback varies from simple corrective feedback, to using rubrics or feedback forms, and using elaborate qualitative feedback. Contrary to expectation, it seems that studies providing students with just the correct answers, and not elaborate feedback, get the most positive results.

Second, whether an assessment is mandatory is almost evenly divided between studies and a similar proportion of studies in both the mandatory and the voluntary group get positive results from intermediate assessment.

The third characteristic is who the assessor is. A subset of the studies used peer or self-assessment as a means of intermediate assessment. Studies using self-assessment showed more positive results than those using peer assessment, which in turn showed more positive results than the studies using teacher assessment. However, a large proportion of studies did not describe who the assessor was, and the groups using peer or self-assessment were substantially smaller than those using teacher assessment.

The fourth and final characteristic under investigation was the reward students got for the assessment. The most positive results occurred in studies rewarding intermediate assessment with a percentage of the final grade, but other rewards include bonus points, or just knowledge of the attained score without any reward.

The chapter ends with three scenarios for intermediate assessment, to further explore how characteristics of intermediate assessment can be combined in practice.

Chapter 3

After surveying the literature on assessment in higher education in Chapter 2, Chapters 3 to 5 focus on the context of Leiden University. In this chapter, teacher and student perceptions of intermediate assessment were investigated. In Chapter 1, I had already identified several positive effects of intermediate assessment, but intermediate assessment can also lead to teaching to the test or strategic student behaviour. Since teachers and students are important in evaluating the implementation of intermediate assessment, the study reported in

Chapter 3 focused on (1) What types of intermediate assessment are used in the programmes under investigation? and (2) How are these types of intermediate assessment perceived by teachers and students?

To investigate these research questions, eighteen university teachers representing fifteen first-year Law, Criminology or Psychology courses were interviewed. Furthermore, students' evaluations of teaching (SET) were investigated for student perceptions, and a small sample of psychology students was interviewed for a more detailed account on the results of the SETs.

Teachers employed a large diversity of intermediate assessment in their courses, because they wanted to measure diverse skills, or because they wanted to experiment with assessment methods. Not all teachers were 'first-time' intermediate assessment users, about half of them had used some form of intermediate assessment before the 2013 curriculum adaptation.

Teacher perceptions indicate that, in general and irrespective of the type of intermediate assessment they use, teachers were quite positive about intermediate assessment. Main reasons that were discussed for this positivity were the possibility to measure a variety of knowledge and skills and the fact that intermediate assessment forces students to start studying earlier in the course period. Teachers hoped, and some reported this hope as fulfilled, that starting earlier would lead to more student engagement and active participation in class.

However, besides their positivity, teachers also had some conceptual criticisms about intermediate assessment. Not all teachers felt that intermediate assessment should be necessary, and some were dissatisfied with the preconditions that were set for the assessment. Additionally, teachers also had practical considerations, mainly related to the added workload of intermediate assessment. Some teachers also worried about plagiarism, or the practicality of resits.

Students rated all courses using intermediate assessment favourably, evidenced by results from the SETs. Nevertheless, their reasons for positivity were somewhat different than those of teachers. Students did agree with teachers about the fact that intermediate assessment helped to keep them on track, but they also valued that intermediate assessments improved their subject knowledge or that it helped them to practice. Of course, students also had conceptual criticisms, mainly focusing on the fact that they preferred intermediate assessments that were strongly aligned with the final exam, and practical considerations, focusing on the workload or the planning of the assessment.

The results of this chapter indicate that both teachers and students have favourable views of intermediate assessment, regardless of the assessment type employed in the course. There does, however, seem to be a disconnection between teachers' and students' goals for the assessment.

Chapter 4

After the exploration of teachers' perceptions of intermediate assessment in Chapter 3, this chapter further investigates these. Previous research (e.g. Postareff, Virtanen, Katajavuori, and Lindblom-Ylänne, 2012; Samuelowicz and Bain, 2002; Watkins, Dahlin, and Ekholm, 2005) has investigated teachers' conceptions, ideas, or beliefs about assessment. These authors interviewed teachers and classified their responses as focussing on reproduction of knowledge (assessment focused on memorisation, reproduction of information, etc.) or on transformation of knowledge (assessment focused on deep understanding, and the development of understanding). These studies investigated teachers' ideas about assessment in general, but this chapter specifically focuses on intermediate assessment. Since external factors may constrain teachers' current assessment practice, the current chapter focused on their conceptions for their current and ideal intermediate assessment and tried to answer the following research question: 'What differences in conceptions of intermediate assessment do university teachers display when discussing their current and ideal intermediate assessment?'

Thirteen course coordinators of twelve first-year Psychology, Law, or Criminology courses were interviewed. In the interviews, teachers were asked about their current intermediate assessment and their ideal intermediate assessment. Additionally, they were asked to reflect on preliminary results of the review discussed in Chapter 2, to further prime their ideas about the possibilities of intermediate assessment. All interviews were coded using coding schemes taken from Postareff et al. (2012) and Samuelowicz and Bain (2002). For each teacher, a bar chart was constructed displaying the full variety of different codes.

Across the twelve courses, there are six types of current intermediate assessment. The ideal assessment can also be classified in six groups, but there are some differences between these two. Current assessments, for example, included partial exams and presentations, whereas ideal assessment included

essay writing and classroom assessment. In general, it can be said that teachers have quite transformational conceptions and beliefs for assessment, and that their ideas for the ideal assessment were even more transformational.

Subsequently, teachers are classified into five groups, based on the differences in conception between the current and ideal assessments. The first three groups displayed no difference in conception between the current and ideal assessment and were *no transformational conceptions for both* (N=1), *a mix of transformational and reproductive assessments for both* (N=3), and *only transformational conceptions for both* (N=1). The last two groups included teachers that showed difference in conception between the two and were *ideal assessment is more transformational* (N=7) and *current assessment is more transformational* (N=1). Teachers often mentioned the workload of their ideal assessment as a reason why it was not their current assessment.

The ideas portrayed by teachers indicate that, overall, these teachers are very focused on student learning and aware of the assessments that support this. Teachers' focus on student learning is an indication that assessment support for university teachers maybe should not focus on the fact that transformational assessment is more beneficial, but on ways in which teachers could be supported to actually implement these transformational assessments.

Chapter 5

After exploring the prevalence of intermediate assessment in Chapter 2, and teacher and student perceptions in Chapters 3 and 4, this chapter focuses on the effects of intermediate assessment by seeing how it can explain student results. Student grades are one possible measure of student success, which can be influenced by the university environment (van Berkel, Jansen & Bax, 2012), by student characteristics (Richardson, Abraham, & Bond, 2012), or a combination of both. This chapter uses intermediate assessment as a measure of the university environment, and gender, high school achievement, motivation, and self-regulation as measures of student characteristics to answer the following two research questions. (1) To what extent does the type of intermediate assessment relate to academic achievement? (2) What role do gender, high school achievement, motivation and self-regulation play in this relationship?

Ninety-four first year students (42.6% male, 77.7% under twenty) enrolled in the bachelor programmes of Law (including Fiscal, Notarial, Business, International Business Law, and Law and Economics) and Criminology participated in the research by answering an adapted eight-scale version of the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, Smith, Garcia, & McKeachie, 1993). Furthermore, these students allowed for their first-try final exam grades to be accessed.

Based on the type of intermediate assessment they used, the twenty first-year courses could be divided into six groups. Six courses used no intermediate assessment, eight used one or more written assignments, two used a partial exam, two used mandatory homework assignments, one used an interview and paper, and one used a presentation, proposition, and mini-experiments. Assessment type composite scores were calculated for each assessment type group, but the final two groups were excluded from final analysis due to small student participation numbers.

After preliminary regression analyses to determine which characteristics were significant predictors of student grades, two separate repeated measures analyses of covariance were conducted. The first ANCOVA had a two level within subject variable, contrasting courses with and without intermediate assessment. The between-subject variables were high school GPA, intrinsic goal orientation, task value, and gender. The results show that whether or not a course has intermediate assessment does not influence students' grades. Students with higher GPAs in high school performed better, and those with higher levels of intrinsic goal orientation performed worse. Female students performed better, but an interaction effect indicates that this is only the case for courses without intermediate assessment. In courses that use an intermediate assessment, there is no gender difference.

Subsequently, a second repeated measures ANCOVA was conducted to see whether there were differences for the three assessment types (written assignment, partial exam, mandatory homework). The between subject variables were the same as in the previous analysis. The outcomes from this analysis indicate that there is no difference between the three assessment types. Furthermore, a higher high school GPA leads to higher grades. This analysis shows no differences in grades for female or male students. Intrinsic goal orientation is a negative predictor for courses using written assignments, and task value negatively predicts scores for courses using mandatory homework assignments.

In conclusion, this chapter indicates that the type of intermediate assessment used in a course does not explain differences in student success. However, using intermediate assessment seems to support the grades of male students, suggesting that employing intermediate assessment in a course can close the gender achievement gap. Both motivational measures explored in this study were negatively related to student outcomes. This suggests that teachers should not indiscriminately add intermediate assessment to their courses, but consider how these will affect motivated students.

Discussion

This chapter summarises the findings from the previous chapters. Furthermore, it presents general findings and reflections on those findings. The chapter ends with a discussion of some general limitations, implications for practice, and directions for future research.

With regards to the focal points discussed in the introduction (implementation, perceptions, results), Chapters 2 to 5 show the possibilities for implementation of intermediate assessment, with a legion of different types and characteristics. Perceptions of intermediate assessment are generally positive. Teachers appreciate the possibilities for measuring different skills and the fact that intermediate assessment can help support student learning. When looking at the effects, Chapter 2 suggests that intermediate assessment improves student results, but this is not repeated in the empirical study described in Chapter 5.

When looking at assessment type, Chapter 2 suggests that how you use intermediate assessment is less important than the fact *that you use it*. In Chapter 5, however, it seems that students perform similarly on courses with and without intermediate assessment. Teachers in Chapter 4 stressed the importance of supporting student learning, usually through feedback. Similarly, Chapter 2 shows that the majority of studies make use of some form of feedback. Corrective feedback appears to be more beneficial than elaborate feedback, but, according to previous studies lower achieving students may benefit more from elaborate feedback. Since the literature also suggests that feedback is an important factor in student learning, intermediate assessment should always be accompanied by some form of feedback.

This chapter suggests three implications for practice about the implementation of intermediate assessment. The first is to encourage the use of intermediate assessment. Several studies have shown the beneficial effects, as is evidenced in Chapter 2. The second implication is that institutional boards should minimise assessment constraints. A multitude of studies, using a multitude of intermediate assessments with different characteristics, have shown that intermediate assessment can improve students' performance. Therefore, it is more important that teachers' independence in deciding on an intermediate assessment is supported than it is to force teachers to adhere to a range of guidelines. The third and final implication is to support teachers' assessment practice. Overall, it seems that teachers view intermediate assessment quite positively, but that they get discouraged by added workload or constraints. Furthermore, it is important to support their assessment knowledge, for example by installing communities of practice. Overall, this dissertation has shown that a wide variety of intermediate assessments are beneficial for students and that teachers feel committed to improving student learning through assessment.

Samenvatting



Hoofdstuk 1: Introductie

Toetsing is een integraal onderdeel van het hoger onderwijs en wordt zowel ingezet om leren te meten (summatieve toetsing) als om het leren te ondersteunen (formatieve toetsing). In de praktijk overlappen summatieve en formatieve processen elkaar vaak, bijvoorbeeld door de kennis van studenten te meten en vervolgens feedback te geven om de kennis te vergroten. Een manier om toetsing in te zetten is tussentijds toetsen, het gebruik van toetsen gedurende het semester in plaats van alleen in de tentamenweek. Tussentijds toetsen kan verscheidene vormen aannemen en onderzoek heeft aangetoond dat het de prestaties van studenten kan verbeteren. Daarom is het op de Universiteit Leiden ingevoerd als een maatregel om het studiesucces te vergroten.

Tussentijds toetsen kan cognitieve uitkomsten (bijvoorbeeld cijfers), affectieve uitkomsten (bijvoorbeeld motivatie) en gedragsuitkomsten (bijvoorbeeld studietijd) beïnvloeden. Deze drie uitkomsten hangen vaak met elkaar samen, bijvoorbeeld dat studenten die meer geëngageerd zijn door een vak, meer tijd zullen besteden aan studeren en zo hogere cijfers zullen halen.

Tussentijds toetsen verbetert de resultaten van studenten door middel van vier verschillende processen. Als eerst is er sprake van het *testing effect*, het herhaaldelijk toetsen van informatie zorgt ervoor dat deze beter onthouden wordt. Ten tweede speelt *distributed practice* een rol, het spreiden van studeertijd over het semester zorgt voor sterkere retentie in het geheugen. Deze twee effecten kunnen elkaar versterken. Onderzoek heeft aangetoond dat het beter werkt als studenten zichzelf verspreid toetsen dan wanneer ze blokken door herhaaldelijk te toetsen kort voor het tentamen. De andere twee effecten zijn toegenomen tijdsbesteding en tijd voor reflectie. Tussentijds toetsen moedigt studenten aan meer tijd aan hun studie te besteden, wat tot betere cijfers leidt. Daarnaast bieden de resultaten van tussentijdse toetsing studenten de mogelijkheid om te reflecteren op hun leerproces, wat hun prestaties kan verbeteren.

Er zijn verschillende manieren om tussentijds toetsen te implementeren. Er is een verscheidenheid aan verschillende kenmerken, zoals, bijvoorbeeld, de hoeveelheid toetsen, de frequentie van de toetsen, het soort toets of de beloning voor de toets (zie voor een volledig overzicht Tabel 1.1). Op de Universiteit Leiden werd tussentijds toetsen met minimale randvoorwaarden ingevoerd, waardoor docenten veel vrijheid hebben de toetsing te laten variëren op de kenmerken die zij belangrijk vinden. Op de faculteit Rechtsgeleerdheid waren

docenten met vakken van 10 EC verplicht een deeltentamen aan te bieden, maar alle andere docenten hadden de vrijheid om te besluiten al dan niet een tussentijdse toets aan te bieden. Bij het instituut Psychologie moesten alle docenten *meervoudig toetsen*, dat wil zeggen meerdere toetsvormen aanbieden binnen een vak. Docenten kozen in dit geval vaak voor een vorm van tussentijds toetsen.

In dit proefschrift probeer ik een antwoord te geven op de vraag ‘Op welke manieren kan tussentijds toetsen in het hoger onderwijs worden ontworpen om studeergedrag en prestaties te verbeteren?’. Bij het beantwoorden van deze vraag heeft ieder hoofdstuk betrekking op minimaal een van de volgende drie kernpunten. Als eerste is er de implementatie van tussentijds toetsen, met een specifieke focus op toetskenmerken (dit komt terug in Hoofdstukken 2-5). Het tweede kernpunt is student en docent percepties van tussentijds toetsen (Hoofdstukken 3 & 4) en het derde is de effecten van tussentijds toetsen op studentcijfers (Hoofdstukken 2 & 5).

Hoofdstuk 2: Een overzicht van de literatuur over tussentijds toetsen in het Hoger Onderwijs

In dit hoofdstuk wordt een overzicht van de literatuur over tussentijds toetsen gegeven en de vraag ‘Welke kenmerken van tussentijds toetsen zijn gerelateerd aan cijfers van studenten?’ beantwoord. De kernpunten die in dit hoofdstuk centraal staan zijn implementatie en effecten. Uiteindelijk werden 88 empirische artikelen in het overzicht geïncludeerd. Uit deze artikelen werden elf verschillende toetskenmerken gedestilleerd. Deze kenmerken zijn het type toets, de frequentie van de toets, de duur van de toets, het aantal en soort vragen, de scoring van de toets, het soort feedback dat studenten krijgen, wie de assessor is, wat het toetsmedium is, waar de toets plaatsvond, of de toets verplicht was en eventuele andere relevante kenmerken.

De resultaten laten zien dat er over de hele wereld onderzoek gedaan wordt naar tussentijds toetsen, maar de meerderheid van de studies is uitgevoerd in de Verenigde Staten van Amerika, het Verenigd Koninkrijk en Australië. Als er gekeken wordt naar disciplines wordt het meeste onderzoek uitgevoerd binnen de geneeskunde, maar natuurwetenschappen, sociale wetenschappen en rechten en bedrijfswetenschappen zijn ook veelvoorkomende disciplines.

In de 88 artikelen worden in totaal 24 verschillende toetstypen beschreven. De meest voorkomende toetstypen zijn deeltentamens, *quizzes* (korte toetsen) en schrijfopdrachten, alle drie 'generieke' toetstypen die in verschillende disciplines ingezet kunnen worden. Van deze drie is de quiz de enige die niet in alle verschillende disciplines in de huidige steekproef voorkomt. Er is een grote verscheidenheid in de hoeveelheid toetsen, variërend van 1 tot 38 tussentijdse toetsen in een vak. De meerderheid van de vakken heeft tussen de een en tien toetsen.

Vier toetskenmerken werden in meer detail beschreven, met behulp van voorbeeldartikelen. Deze kenmerken zijn het gebruik van feedback, of een toets verplicht is, wie de assessor is en wat de beloning voor een toets is.

Met betrekking tot het eerste kenmerk, feedback, gebruikt de meerderheid van de studies enige vorm van feedback, variërend van simpele correctieve feedback, naar feedbackformulieren en uitgebreide kwalitatieve feedback. Tegen de verwachtingen in worden de meeste positieve resultaten gerapporteerd voor studies waar studenten alleen de juiste antwoorden, maar geen uitgebreide feedback, krijgen.

Wat betreft het tweede kenmerk, de verplichting van de toets, zijn in de steekproef studies waar de tussentijdse toets verplicht is en studies waar dat niet zo is bijna gelijk verdeeld. Bovendien laat in zowel de verplichte als de vrijwillige groep een vergelijkbare proportie van studies positieve resultaten zien.

Het derde kenmerk is wie de assessor is. Een deel van de studies gebruikte *peer of self-assessment* als een middel voor tussentijdse toetsing. De studies waar studenten zichzelf toetsten lieten positievere resultaten zien dan die waar studenten door hun medestudenten getoetst werden, terwijl deze weer positievere resultaten lieten zien dan studies waar de docent de assessor was. Hier moet echter wel opgemerkt worden dat een groot deel van de artikelen niet specificeerde wie de assessor was en dat de groep van studies die *peer of self-assessment* gebruikte substantieel kleiner was dan de groep die gebruik maakte van de docent als assessor.

Het vierde en laatste kenmerk dat onderzocht werd is de beloning die studenten kregen voor de toets. De meeste positieve resultaten werden gevonden in studies die studenten beloonden met een percentage van het eindcijfer. Andere auteurs beloonden studenten met bonuspunten, of gaven hen alleen kennis van de behaalde score zonder verdere beloning.

Het hoofdstuk eindigt met drie scenario's voor tussentijds toetsen, als verdere verkenning van hoe de kenmerken van tussentijds toetsen in de praktijk gevarieerd kunnen worden.

Hoofdstuk 3: Percepties over tussentijds toetsen van docenten en studenten

Na het overzicht van de literatuur over tussentijds toetsen in het hoger onderwijs gepresenteerd in Hoofdstuk 2, richten Hoofdstukken 3 tot en met 5 zich op de context van de Universiteit Leiden. In dit hoofdstuk werden de percepties over tussentijds toetsen van docenten en studenten onderzocht. In Hoofdstuk 1 zijn al een aantal positieve effecten van tussentijds toetsen genoemd, maar tussentijds toetsen kan ook leiden tot *teaching to the test* of strategisch gedrag van studenten. Aangezien docenten en studenten van belang zijn bij het evalueren van de implementatie van tussentijds toetsen, wordt in dit hoofdstuk onderzocht (1) Welke tussentijdse toetstypen worden gebruikt in de onderzochte opleidingen? En (2) Hoe worden deze toetstypen gepercipieerd door docenten en studenten?

Om deze vragen te onderzoeken werden achttien docenten geïnterviewd. Deze docenten vertegenwoordigden vijftien verschillende eerstejaars vakken van de opleidingen Rechtsgeleerdheid, Criminologie en Psychologie. Daarnaast werden studentevaluaties onderzocht om achter studentpercepties te komen en werd een kleine steekproef van psychologiestudenten geïnterviewd voor meer details over de resultaten van de studentevaluaties.

Docenten gebruiken een diverse set tussentijdse toetsen in hun vakken, omdat ze diverse vaardigheden wilden toetsen, of om te kunnen experimenteren met verschillende toetstypen. Ongeveer de helft van de docenten was onervaren op het gebied van tussentijds toetsen, de andere helft had al enige vorm van tussentijds toetsen gebruikt voor de aanpassing van het curriculum in 2013.

De docentenpercepties laten zien dat, over het algemeen en onafhankelijk van het type toets dat gebruikt werd, docenten vrij positief tegenover tussentijds toetsen staan. Hun belangrijkste redenen daarvoor waren dat tussentijds toetsen de mogelijkheid biedt verschillende kennis en vaardigheden te toetsen, en dat tussentijds toetsen studenten dwingt eerder in het blok te beginnen met studeren. Docenten hoopten dat studenten die eerder begonnen met studeren actiever mee zouden doen tijdens contactonderwijs, een hoop die door een aantal docenten bevestigd werd.

Naast hun positieve oordeel hadden docenten ook een aantal conceptuele kritieken over tussentijds toetsen. Niet alle docenten vonden dat tussentijds toetsen echt noodzakelijk was en sommigen waren ontevreden met de randvoorwaarden die voor de toetsing gesteld waren. Daarnaast hadden docenten ook nog praktische overwegingen, voornamelijk gerelateerd aan de extra werklast die tussentijds toetsen met zich meebrengt. Een paar docenten maakten zich ook zorgen over plagiaat en de uitvoering van eventuele herkansingen.

Resultaten van de studentevaluaties laten zien dat studenten de cursussen met tussentijdse toetsen positief waarderen. Hun redenen voor deze positiviteit zijn echter deels anders dan die van docenten. Studenten waren het met docenten eens dat tussentijds toetsen helpt om hen op de rit te houden, maar ze vonden het ook belangrijk dat tussentijdse toetsing hun vakkennis vergrootte, of hen hielp oefenen met de stof. Natuurlijk hadden studenten ook conceptuele kritieken, die zich voornamelijk focusten op het feit dat studenten de voorkeur gaven aan tussentijdse toetsing die sterk gerelateerd was aan het tentamen. Wat betreft de praktische overwegingen hadden de studenten opmerkingen over de werklast en de planning van de toetsing.

De resultaten van dit hoofdstuk laten zien dat zowel docenten als studenten een gunstig beeld van tussentijds toetsen hebben, onafhankelijk van het toetstype dat gebruikt werd. Er lijkt echter wel een verschil te zitten in de doelen die docenten en studenten met de toetsing hebben.

Hoofdstuk 4: Opvattingen van docenten over hun huidige en ideale tussentijdse toets

Na de eerste verkenning van de docentenpercepties in Hoofdstuk 3, gaat dit hoofdstuk dieper op deze percepties in. De opvattingen, ideeën en overtuigingen van docenten over toetsen zijn al door verschillende auteurs onderzocht (bijvoorbeeld Postareff, Virtanen, Katajavuori, en Lindblom-Ylänne, 2012; Samuelowicz en Bain, 2002; Watkins, Dahlin, en Ekholm, 2005). Deze auteurs interviewden docenten en classificeerden hun antwoorden als gefocust op reproductie van kennis (toetsen gericht op herinnering, het reproduceren van informatie, etc.) of op de transformatie van kennis (toetsen gericht op diep begrip en het ontwikkelen van begrip). Al deze studies onderzochten de ideeën van docenten over toetsing in het algemeen, maar dit hoofdstuk richt zich

specifiek op tussentijds toetsen, aangezien deze vorm van toetsing vaak andere doelen met zich meebrengt. Deze andere doelen kunnen eventueel ook tot andere ideeën leiden. Aangezien de toetspraktijk van docenten beperkt wordt door externe factoren, richtte het huidige hoofdstuk zich op de opvattingen voor zowel de huidige als de ideale tussentijdse toets en probeerde de volgende onderzoeksvraag te beantwoorden: ‘Welk verschil in opvattingen over tussentijds toetsen laten docenten zien als ze het hebben over hun huidige en hun ideale tussentijdse toets?’

Dertien vakcoördinatoren van twaalf eerstejaarsvakken van de opleidingen Psychologie, Rechtsgeleerdheid of Criminologie werden geïnterviewd. In de interviews werd docenten naar hun huidige en hun ideale tussentijdse toets gevraagd. Daarnaast werd hen gevraagd te reflecteren op voorlopige uitkomsten van het literatuuroverzicht van Hoofdstuk 2, om hun ideeën over de mogelijkheden van tussentijds toetsen verder te inspireren. Alle interviews zijn gecodeerd met behulp van codeerschema's afgeleid uit Postareff et al. (2012) en Samuelowicz en Bain (2002). Voor elke docent werd een staafdiagram met de volledige variëteit aan toegekende codes opgesteld.

In de twaalf vakken werden zes types tussentijdse toetsen gebruikt, en de ideale tussentijdse toetsen konden ook ingedeeld worden in zes groepen. Tussen de zes huidige en zes ideale toetsen zaten wel enkele verschillen. De huidige toetsen waren bijvoorbeeld deeltentamens en presentaties, terwijl de ideale toetsen essay opdrachten en *classroom assessment* waren. Over het algemeen kan het gezegd worden dat docenten vrij transformationele opvattingen en overtuigingen over tussentijds toetsen hadden en dat hun ideeën voor de ideale toets in nog grotere mate transformationeel waren.

Docenten kunnen geclassificeerd worden in vijf groepen, gebaseerd op de verschillen in opvattingen tussen de huidige en de ideale toets. De eerste drie groepen toonden geen verschil in opvattingen tussen beide toetsen en kunnen beschreven worden als *geen transformationele opvattingen voor beide* (N=1), *een mix van transformationele en reproductieve opvattingen voor beiden* (N=3) en *alleen transformationele opvattingen voor beide* (N=1). De laatste twee groepen bestaan uit docenten die een verschil lieten zien in opvattingen tussen de huidige en de ideale toets en kunnen beschreven worden als *de ideale toets is transformationeler* (N=7) en *de huidige toets is transformationeler* (N=1). Als reden waarom de ideale toets niet in praktijk gebracht werd als de huidige toets gaven docenten vaak aan dat de werklast van de ideale toets te hoog was.

De ideeën die door docenten besproken werden laten zien dat ze, over het algemeen, zeer gefocust zijn op het leren van studenten en zich bewust zijn van toetsvormen die dit leren ondersteunen. Dit geeft aan dat toetsondersteuning voor docenten op universiteiten zich misschien niet moet richten op het feit dat transformationele toetsen gunstiger zijn, maar op manieren waarop docenten deze transformationele toetsen ook echt kunnen implementeren.

Hoofdstuk 5: De invloed van tussentijds toetsen en studentkenmerken op studiesucces

Na het verkennen van de prevalentie van tussentijdse toetsing in Hoofdstuk 2, en docent- en studentpercepties in Hoofdstukken 3 en 4, richt dit hoofdstuk zich op de effecten van tussentijds toetsen, door te kijken hoe het studentresultaten kan verklaren. Cijfers zijn een mogelijke indicator van studiesucces. Volgens onderzoek kan studiesucces beïnvloed worden door de universitaire omgeving (van Berkel, Jansen & Bax, 2012), door studentkenmerken (Richardson, Abraham, & Bond, 2012), of door een combinatie van beide. In dit hoofdstuk wordt tussentijds toetsen gebruikt als factor van de universitaire omgeving, en worden de studentkenmerken geslacht, prestatie op de middelbare school, motivatie en zelfregulatie gebruikt om antwoord te geven op de volgende twee onderzoeksvragen: (1) In hoeverre relateert het soort tussentijdse toets aan academische prestaties? (2) Welke rol spelen geslacht, prestaties op de middelbare school, motivatie en zelfregulatie in deze relatie?

Vierennegentig eerstejaars studenten (42,6% mannen; 77,7% jonger dan 20) die Rechtsgeleerdheid (inclusief Fiscaal Recht, Notarieel Recht, Bedrijfswetenschappen, Economie en International Business Law) of Criminologie studeerden namen deel aan het onderzoek door een aangepaste versie van de Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, Smith, Garcia & McKeachie, 1993) in te vullen. Bovendien gaven deze studenten toestemming dat hun cijfers voor de eerste kans van de tentamens opgevraagd werd.

Op basis van de soort tussentijdse toets die gebruikt werd konden de eenentwintig eerstejaarsvakken verdeeld worden in zes groepen. Zes vakken maakten geen gebruik van tussentijdse toetsing, acht gebruikten een of meer schrijfopdrachten, twee gebruikten een deeltentamen, twee verplichte

huiswerkopdrachten, één een interview en paper, en het laatste vak gebruikte een presentatie, stelling en een mini-experiment. Voor elke groep werd een samengestelde score berekend, maar de laatste twee groepen werden van analyse uitgesloten vanwege lage studentenaantallen.

Na verkennende regressieanalyses om te bepalen welke kenmerken significante voorspellers van cijfers waren, werden twee aparte covariantie analyses voor herhaalde metingen (Repeated Measures ANCOVA) uitgevoerd. De eerste ANCOVA vergeleek vakken met en zonder tussentijdse toets. De voorspellende variabelen in deze analyse waren gemiddelde eindexamencijfer, geslacht, intrinsieke doel oriëntatie en taakwaarde (beide maten van motivatie). De resultaten lieten zien dat het voor de cijfers van een student niet uitmaakt of een vak wel of geen tussentijdse toets gebruikt. Studenten met hogere examencijfers presteerden beter, en studenten met hogere intrinsieke doel oriëntatie presteerden slechter. Vrouwelijke studenten scoorden beter, maar een interactie-effect toont aan dat dit alleen het geval is in vakken zonder tussentijdse toetsing. In vakken die gebruik maken van tussentijdse toetsing is er geen verschil in prestatie tussen mannen en vrouwen.

Vervolgens werd een tweede repeated measures ANCOVA uitgevoerd om te zien of er verschillen waren tussen de drie toetstypen (schrijfopdrachten, deeltentamen, verplichte huiswerkopdrachten). In deze analyse werden dezelfde voorspellers gebruikt als in voorgaande analyse. Uitkomsten van deze analyse tonen dat er geen verschil is tussen de drie toetstypen. Een hoger eindexamencijfer leidt tot hogere prestaties. In deze analyse zijn er geen verschillen tussen mannen en vrouwen. Intrinsieke doel oriëntatie is een negatieve voorspeller voor vakken die schrijfopdrachten gebruiken, en taakwaarde is een negatieve voorspeller voor vakken die verplichte huiswerkopdrachten gebruiken.

Concluderend laat dit hoofdstuk zien dat het type tussentijdse toets dat gebruikt wordt in een vak de verschillen in studiesucces niet kan verklaren. Het gebruik van tussentijdse toetsing lijkt echter de cijfers van mannen te ondersteunen, wat suggereert dat tussentijds toetsen ingezet kan worden om de *gender achievement gap* te verkleinen. Allebei de motivationele constructen die in deze studie verkend werden waren negatief gerelateerd aan studentuitkomsten. Dit suggereert dat docenten moeten overwegen hoe tussentijdse toetsing gemotiveerde studenten beïnvloedt, in plaats van het klakkeloos in te voeren.

Hoofdstuk 6: Discussie

In dit hoofdstuk worden de resultaten van de voorgaande hoofdstukken samengevat. Bovendien worden de algemene bevindingen en reflecties op deze bevindingen gepresenteerd. Het hoofdstuk eindigt met het bespreken van de beperkingen van het onderzoek, implicaties voor de praktijk en aanbevelingen voor aanvullend onderzoek.

Als gekeken wordt naar de kernpunten die in de introductie genoemd zijn (implementatie, percepties en effecten), laten Hoofdstukken 2 tot en met 5 de mogelijkheden voor implementatie van tussentijds toetsen zien, met een grote verscheidenheid aan types en kenmerken. De percepties over tussentijds toetsen zijn over het algemeen positief. Docenten waarderen de mogelijkheid verschillende vaardigheden te toetsen en het feit dat tussentijds toetsen het leren van studenten kan ondersteunen. Als er naar de effecten gekeken wordt, suggereert Hoofdstuk 2 dat tussentijds toetsen studentresultaten bevordert, maar deze uitkomst wordt niet herhaald in het empirische onderzoek beschreven in Hoofdstuk 5.

Met betrekking tot toetstype suggereert Hoofdstuk 2 dat hoe je tussentijds toetsen inzet minder belangrijk is dan het feit *dat je het inzet*. In Hoofdstuk 5 lijken studenten echter vergelijkbaar te presteren op vakken met en zonder tussentijdse toets. In Hoofdstuk 4 benadrukten de docenten het belang van het ondersteunen van studenten, meestal door middel van feedback. Ook in Hoofdstuk 2 is te zien dat de meerderheid van de studies gebruik maakt van enigerlei vorm van feedback. Correctieve feedback lijkt effectiever te zijn dan uitgebreide feedback, maar eerder onderzoek toont aan dat vooral lager presterende studenten meer gebaat zijn bij uitgebreide feedback. Aangezien de literatuur ook laat zien dat feedback een belangrijke factor is in het leren van studenten, zou tussentijds toetsen altijd vergezeld moeten worden door een vorm van feedback.

Dit hoofdstuk stelt drie praktische implicaties voor, die betrekking hebben op de implementatie van tussentijds toetsen. De eerste is om het gebruik van tussentijds toetsen aan te moedigen. Verschillende studies hebben de gunstige effecten van tussentijds toetsen laten zien, zoals te lezen is in Hoofdstuk 2. De tweede implicatie is dat instituutsbesturen de randvoorwaarden voor tussentijds toetsen moeten minimaliseren. Een grote hoeveelheid studies, die een grote variëteit aan tussentijdse toetsen gebruikt, heeft aangetoond dat tussentijds toetsen de prestaties van studenten kan bevorderen. Daarom

is het belangrijker dat de beslissingsvrijheid wat betreft tussentijdse toetsing van docenten ondersteund wordt, dan om docenten te dwingen aan een reeks richtlijnen te voldoen. De derde en laatste implicatie is dat de toetspraktijk van docenten ondersteund moet worden. Over het algemeen lijkt het dat docenten vrij positief tegenover tussentijds toetsen staan, maar dat ze ontmoedigd worden door de extra werklast of randvoorwaarden. Bovendien is het belangrijk om hun toetskennis te ondersteunen, bijvoorbeeld door middel van communities of practice. Samengevat heeft dit proefschrift laten zien dat een grote variëteit aan tussentijds toetsen gunstig voor studenten is, en dat docenten zich geroepen voelen het leren van studenten te bevorderen door middel van toetsing.

List of publications and presentations

Scientific Publications

- Day, I.N.Z., van Blankenstein, F.M., Westenberg, P.M., & Admiraal, W.F. (2017). Teacher and student perceptions of intermediate assessment in higher education. *Educational Studies*. Advance online publication. doi:10.1080/03055698.2017.1382324
- Day, I.N.Z., van Blankenstein, F.M., Westenberg, P.M., & Admiraal, W.F. (2017). A review of the characteristics of intermediate assessment and their relationship with student grades. *Assessment & Evaluation in Higher Education*. Advance online publication. doi:10.1080/02602938.2017.1417974
- Day, I.N.Z., van Blankenstein, F.M., Westenberg, P.M., & Admiraal, W.F. (2018). Explaining Individual Student Success Using Continuous Assessment Types and Student Characteristics. *Higher Education Research and Development*. Advance online publication. doi: 10.1080/07294360.2018.1466868
- Day, I.N.Z., van Blankenstein, F.M., Westenberg, P.M., & Admiraal, W.F. (under revision). University teachers' conceptions of their current and ideal intermediate assessment: An A+ is good, but speaking your mind is better.

Professional Publications

- Day, I.N.Z., Huisman, B.A., Vereijken, M.W.C. (2017). Exploring freedom and control in global higher education: een conferentie verslag van de jaarlijkse bijeenkomst van de Society for Research into Higher Education [Exploring freedom and control in global higher education: A report of the annual meeting of the Society for Research into Higher Education], *Onderzoek van Onderwijs*, 46, 19-21.

Paper Presentations

- Day I.N.Z., Van Blankenstein, F.M., Westenberg P.M. & Admiraal W.F. (2015, August). *Teacher and student experiences with intermediate assessment in higher education*. Paper presented at the JURE Conference, 23-24 august, Limassol, Cyprus.
- Day I.N.Z., Blankenstein F.M van., Westenberg P.M. & Admiraal W.F. (2016, July). *Intermediate Assessment in Undergraduate Law Programs: The Relation between Test Characteristics, Student Characteristics and Study Results*. Paper presented at the Higher Education Conference, 13-15 July Amsterdam, the Netherlands.

Day I.N.Z., Van Blankenstein, F.M., Westenberg P.M. & Admiraal W.F. (2016, December). *Characteristics of intermediate assessment and student results: A literature review*. Paper presented at the annual meeting of the Society for Research into Higher Education, 6-9 December, Newport, Wales.

Poster Presentations and Roundtables

Day, I.N.Z., van Blankenstein, F.M., Westenberg, P.M., & Admiraal, W.F. (2014 June). *Tussentijds toetsen en studiesucces in het Hoger Onderwijs: een beschrijvende studie [Intermediate assessment and study success in Higher Education: a descriptive study]*. Poster presented at the Onderwijs Research Dagen, 11-13 June, Groningen, the Netherlands.

Day, I.N.Z., van Blankenstein, F.M., Westenberg, P.M., & Admiraal, W.F. (2015, June). *Tussentijds toetsen in de bacheloropleidingen rechtsgeleerdheid en criminologie [Intermediate Assessment in the Bachelor programs Law and Criminology]*. Poster presented at the Onderwijs Research Dagen, 17-19 June, Leiden, the Netherlands.

Day, I.N.Z., Van Blankenstein, F.M., Westenberg, P.M., & Admiraal, W.F. (2015, August). *Teacher and student experiences with intermediate assessment in higher education*. Poster presented at the EARLI Conference, 25-29 August, Limassol, Cyprus.

Day, I.N.Z., van Blankenstein, F.M., Westenberg, P.M., & Admiraal, W.F. (2016, May). *Literatuurreview: Kenmerken van tussentijds toetsen in het hoger onderwijs en hun relatie met opbrengst bij studenten [Literature review: Characteristics of intermediate assessment in higher education and their relationship to student grades]*. Roundtable presentation at the Onderwijs Research Dagen, 25-27 May, Rotterdam, the Netherlands.

Curriculum Vitae



Indira Day was born in Leiden on September 27, 1988. After graduating from Stedelijk Gymnasium Leiden in 2005, she studied pedagogical sciences at Radboud University in Nijmegen, graduating with a bachelor's diploma in 2008. Subsequently, Indira completed the Research Master Behavioural Science at the same university. She wrote her master's thesis at the Learning and Plasticity research group and graduated in 2011. Immediately after graduation, Indira started working as a teacher at the Educational Science department of the Radboud University, supervising bachelor theses and introduction to educational science tutorial meetings. From 2011 to 2013, she worked at Vrije Universiteit Amsterdam, supervising tutorial meetings on methodology and statistics for psychology and pedagogical science students. In September 2013 Indira started working as a PhD student at the Leiden University Graduate School of Teaching (ICLON). She currently works as a researcher at ICLON, studying the use of peer feedback on presentations using video, and as a tutorial supervisor for psychology students at the University of Amsterdam.

Dankwoord



Hoewel een promotietraject een pad is wat je grotendeels alleen aflegt, zijn er toch verschillende mensen elementair geweest in het proces. Hier wil ik hen graag bedanken. Als eerst Jan van Driel, die me uitnodigde te solliciteren voor dit project. Zonder zijn aanmoediging zou ik überhaupt niet hier beland zijn.

Waarschijnlijk de belangrijkste mensen in mijn promotie waren mijn (co) promotoren, Wilfried Admiraal, Michiel Westenberg en Floris van Blankenstein. Wilfried, je hebt jezelf een betrokken promotor getoond, die altijd tijd voor me wist te vinden in je volle agenda. Michiel, jouw perspectief als buitenstaander op mijn promotieonderwerp leidde vaak tot scherpe vragen, waar mijn stukken beter van werden. Floris, jouw ondersteuning bij de kwalitatieve analyses en in het proces hebben me, zeker de laatste maanden, geholpen alles op een rijtje te houden.

Zeker in het begin van het traject heb ik veel contact gehad met de onderwijsdirecties van Psychologie en de faculteit Rechtsgeleerdheid. De directeuren en onderwijskundig medewerkers van beide instituten, Arjaan Wit, Pauline Schuyt (later Peter van Es), Marlies van der Meer en Gerda Korevaar wil ik bedanken voor hun bijdrage aan de totstandkoming van mijn onderzoek. Zij waren een belangrijke brug tussen mij en de docenten en studenten die deelgenomen hebben aan mijn onderzoek. Natuurlijk wil ik ook alle docenten die de tijd hebben genomen om (soms tweemaal) geïnterviewd te worden bedanken. Ook de studenten die de vragenlijst hebben ingevuld of hebben deelgenomen aan een interview ben ik dankbaar voor hun bijdrage aan mijn onderzoek.

Het ICLON is een fantastische werkomgeving geweest, met een gevarieerde en stimulerende onderzoeksgroep. To all researchers on the fourth floor: Thanks for letting me be your lunch lady for the past four years. In het bijzonder bedank ik Mayke en Bart. Jullie zijn fantastische kamergenoten geweest en ik ben blij dat we deze ervaring met elkaar hebben mogen delen. Alle andere (oud)promovendi, we waren een fijne klik en ik ga jullie missen.

Buiten de universiteit ben ik vooral dankbaar voor alle mensen die me de mogelijkheid boden aan iets anders dan mijn proefschrift te denken. Lieve familie en vrienden, zonder de afleiding had ik het niet gered. Pap en mam, bedankt voor jullie steun, sorry dat ik niet meer over mijn proefschrift vertelde. Als laatst wil ik mijn zussen bedanken, in het bijzonder Shanti, voor je kunstwerk voor mijn kaft en Zyfra, dat ik altijd met alles bij je terecht kan.

ICLON PhD
dissertation series

- Hoeflaak, A. (1994). *Decoderen en interpreteren: een onderzoek naar het gebruik van strategieën bij het beluisteren van Franse nieuwsteksten.*
- Verhoeven, P. (1997). *Tekstbegrip in het onderwijs klassieke talen.*
- Meijer, P. C. (1999). *Teachers' practical knowledge: Teaching reading comprehension in secondary education.*
- Zanting, A. (2001). *Mining the mentor's mind: The elicitation of mentor teachers' practical knowledge by prospective teachers.*
- Uhlenbeck, A. M. (2002). *The development of an assessment procedure for beginning teachers of English as a foreign language.*
- Oolbekkink-Marchand, H.W. (2006). *Teachers' perspectives on self-regulated learning: An exploratory study in secondary and university education.*
- Henze-Rietveld, F. A. (2006). *Science teachers' knowledge development in the context of educational innovation.*
- Mansvelder-Longayroux, D. D. (2006). *The learning portfolio as a tool for stimulating reflection by student teachers.*
- Meirink, J.A. (2007). *Individual teacher learning in a context of collaboration in teams.*
- Nijveldt, M.J. (2008). *Validity in teacher assessment: An exploration of the judgement processes of assessors.*
- Bakker, M.E.J. (2008). *Design and evaluation of video portfolios: Reliability, generalizability, and validity of an authentic performance assessment for teachers.*
- Oonk, W. (2009). *Theory-enriched practical knowledge in mathematics teacher education.*
- Visser-Wijnveen, G.J. (2009). *The research-teaching nexus in the humanities: Variations among academics.*
- Van der Rijst, R.M. (2009). *The research-teaching nexus in the sciences: Scientific research dispositions and teaching practice.*
- Platteel, T.L. (2010). *Knowledge development of secondary school L1 teachers on concept-context rich education in an action-research setting.*
- Kessels, C.C. (2010). *The influence of induction programs on beginning teachers' well-being and professional development.*
- Min-Leliveld, M.J. (2011). *Supporting medical teachers' learning: Redesigning a program using characteristics of effective instructional development.*
- Dobber, M. (2011). *Collaboration in groups during teacher education.*
- Wongsopawiro, D. (2012). *Examining science teachers pedagogical content knowledge in the context of a professional development program.*

- Belo, N.A.H. (2013). *Engaging students in the study of physics: An investigation of physics teachers' belief systems about teaching and learning physics.*
- De Jong, R.J. (2013). *Student teachers' practical knowledge, discipline strategies, and the teacher-class relationship.*
- Verberg, C.P.M. (2013). *The characteristics of a negotiated assessment procedure to promote teacher learning.*
- Van Kan, C.A. (2013). *Teachers' interpretations of their classroom interactions in terms of their pupils' best interest: A perspective from continental European pedagogy.*
- Dam, M. (2014). *Making educational reforms practical for teachers: Using a modular, success-oriented approach to make a context-based educational reform practical for implementation in Dutch biology education.*
- Hu, Y. (2014). *The role of research in university teaching: A comparison of Chinese and Dutch teachers.*
- Vink, C.C. (2014). *Mapping for meaning: Using concept maps to integrate clinical and basic sciences in medical education.*
- De Hei, M.S.A. (2016). *Collaborative learning in higher education: design, implementation and evaluation of group learning activities.*
- Louws, M.L. (2016). *Professional learning: what teachers want to learn.*
- Moses, I. (2017). *Student-teachers' commitment to teaching.*
- Veldman, I. M.J. (2017). *Stay or leave? Veteran teachers' relationships with students and job satisfaction.*
- Chen, D. (2017). *Intercultural identities of English language teachers: An exploration in China and the Netherlands.*
- Vereijken M.W.C. (2018) *Student engagement in research in medical education.*
- Stollman, S.H.M. (2018). *Differentiated instruction in practice: A teacher perspective.*