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
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, Dirx 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

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

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.