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
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.
Summary

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 GPA’s 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.