

Indira Day¹, Floris van Blankenstein¹, Michiel Westenberg², Wilfried Admiraal¹

¹ICLON, Leiden University Graduate School of Teaching, The Netherlands,

²Leiden University Faculty of Social and Behavioural Sciences, The Netherlands

Introduction

Universities implement intermediate assessments, which are assessments that take place during the course period before the final exam. Intermediate assessment, often also called frequent or continuous assessment, is implemented to encourage students to start their study work earlier in the course period and to study more often, which in turn will lead to improved results. Furthermore, it gives teachers opportunities to assess different goals than with one final exam, and results from the assessment can be used to improve teaching practice.

Research in cognitive psychology has identified two effects underlying the positive effects of intermediate assessment. The first is the testing effect, which states that repeated testing of information helps retention in memory (Roediger & Karpicke, 2006) and the second is the spacing effect, which states that spacing learning activities over time improves memory (Kornell, 2009). The spacing effect is especially interesting since Michaels and Miethe (1998) found that additional study time is only effective if students spread their study time throughout the semester, instead of cramming for the final exam.

In addition to these cognitive psychology studies, which often take place in a lab, there have been several studies that show the positive effects of intermediate assessment on final exam results in actual classes (e.g. Admiraal, Wubbels & Pilot, 1999; Pennebaker, Gosling & Ferrel, 2013). These studies often compare a situation with intermediate assessment with a situation without assessment, but not different forms or characteristics of intermediate assessment. Therefore, there is no clear evidence whether some forms or characteristics of assessment may yield better results than others.

The nature of higher education often prevents teachers from comparing different forms of assessment. It is often inadvisable to radically change a course every year, and it may not be ethical to offer different students different opportunities to meet the course goals. However, instructors can benefit from the assessment experiences of their colleagues and peers at other higher education institutions. Therefore, information on what characteristics of intermediate assessment work should be exchanged.

In the current study, we conduct a systematic review of the body of research focusing on different characteristics of intermediate assessment in higher education and try to answer the following research question: "What characteristics of intermediate assessment in higher education are related to student grades?"

The main purpose of this study is to provide an overview of assessment characteristics and their relation to student grades. To the best of our knowledge such an overview does not currently exist, but it will be very useful to inform teaching practice, and show gaps in our knowledge about intermediate assessment.

Methods

Using Web of Science, we searched for relevant articles by individually combining the adjectives "intermediate", "frequent", "continuous", "programmatic", "in-between", "formative" and "summative" with the noun "assessment". Results were restricted to time range 2000 – march 2016, and refined to showing only articles relating to Education & Educational Research. This yielded 2296 results, of which the title, abstract, and keywords were scanned to determine whether the context was higher education and the research focused on some form of intermediate assessment. Only articles written in English were selected. Ultimately, 275 articles were selected. These were empirical as well as theoretical articles, reviews or commentaries. Of these, 258 were available as full text for our institution. Inclusion criteria for the review included that the studies needed to be empirical, that

student grades needed to be one of the outcomes and sufficient information about the assessment should be provided. There are no a priori criteria for the methodology of studies.

To determine assessment characteristics discussed in the articles, relevant sections (methods, context) were summarised in detail. After this initial exploration of assessment characteristics, the relevant empirical articles will be re-analysed to get more detailed information about the assessment characteristics and their relation to student outcomes. Furthermore, detailed information on the methodology of the article will also be written down.

Articles will be grouped according to clusters of assessment characteristics, and subsequently, the relation between these characteristic and student outcomes will be examined. Ultimately this will lead to a matrix of assessment characteristics and their relation to outcomes, which can inform assessment practice.

Preliminary results

After reading the context and method sections of the articles 130 papers were identified as being relevant for the current review. During this identification process a large variety in assessment methods between the different studies was found. Studies differed in for example, the frequency and number of assessments, the type of assessment offered, the assessor, the grading of the assessment, or the use of feedback.

Research indicates special interest for several characteristics. First of all, the use of formative assessment. Formative assessment focuses on improving learning by feedback, which leads to improved results compared to summative assessment (Black & William, 2004).

Second, Bangert-Drowns, Kulik & Kulik, (1991) indicated that the effect of assessment diminishes with each additional assessment that is added, which makes the number of assessments a characteristic of interest. Comparing several studies with different numbers of assessment can indicate an “optimal” number of assessments.

Other characteristics that can be interesting to teachers are for example the duration of the assessment, the assessor, and whether or not an assessment is voluntary.

The final review aims to provide a comprehensive overview of how different assessment characteristics relate to student results. This overview can subsequently be used in educational practice, where teachers can choose to vary different characteristics to help improve students' grades.

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.

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C.-L. C. (1991). Effects of Frequent Classroom Testing. *The Journal of Educational Research*, 85(2), 89-99. doi: 10.1080/00220671.1991.10702818

Black, P., & William, 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

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

Michaels, J. W., & Miethe, T. D. (1989). Academic Efforts and College Grades. *Social Forces (University of North Carolina Press)*, 68(1), 309.

Pennebaker, J. W., Gosling, S. D., & Ferrell, J. D. (2013). Daily Online Testing in Large Classes: Boosting College Performance while Reducing Achievement Gaps. *PLoS ONE*, 8(11), e79774. doi: 10.1371/journal.pone.0079774

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