

University of Groningen

Implementing assessment innovations in higher education

Boevé, Anna Jannetje

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:
2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Boevé, A. J. (2018). *Implementing assessment innovations in higher education*. [Thesis fully internal (DIV), University of Groningen]. Rijksuniversiteit Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter

1



Introduction

1.1 Introduction

Student assessment plays an important role in higher education. Two recent developments in educational assessment were important for the research conducted in this thesis: (1) the massification of higher education and (2) the digital developments in higher education. Due to the massification of higher education, teachers are faced with large classes of, sometimes, hundreds of students (Hornsby & Osman, 2014). As a result the teacher-student ratio becomes very small, leaving teachers with little time and resources to monitor their students' learning process. As a result of the digital developments, it is not possible for students in the Netherlands to access university courses without having access to the Internet and computer or smart devices: from enrolment to course participation to the access of final grades, virtually all information is accessible and stored online. Although there are important new technological developments, students are often still required to tick the appropriate box on a paper-and-pencil answer sheet when completing exams. Thus the challenge facing universities is how to integrate digital technologies in a way that contributes to improving learning and assessment.

Another important aspect that inspired the research conducted in this thesis was the new system of performance-based funding that was recently introduced in the Netherlands in the form of an agreement between each university and the government (de Boer et al., 2015). Performance based funding of universities is a policy change that has occurred in the past few decades in countries across the world in different ways (De Boer et al., 2015). In the Netherlands, the agreement is that if after a certain period, the objectives stipulated in the agreement between a university and government have not been met that university will receive less funding. Important performance indicators agreed upon were the student dropout rates in the first year and the 4-year bachelor graduation rate. Although recognizing that maintaining quality in higher education should not be limited to these two indicators (Bussemaker, 2014), other indicators in the agreement remained vague. In line with the digital revolution, the minister of education also noted that: "New developments like open online education provide the opportunity to further improve the quality of higher education. Higher education should give the right attention to all these (existing and new) challenges" (Bussemaker, 2014).

In addition to the introduction of performance-based funding, the Dutch Inspectorate of Education recently evaluated the quality of assessment in higher education (Inspectie van het Onderwijs, 2016). One of the recommendations was to increase research and evaluation concerning assessment quality in higher education. This thesis is a collection of studies investigating the implementation of innovations in assessment in higher education.

1.1.1 Assessment

Assessment in education refers to the entire process of collecting information concerning students' knowledge, skills, and/or abilities (Cizek, 2009). A test or exam is a "systematic sample of a person's knowledge, skill or ability" (Cizek, 2009, pp.64). Test results can be used for different purposes, such as determining the strengths and weaknesses of students, guiding instruction, and making decisions about students (Cizek, 2009). A test is therefore a potential part of assessment, but assessment does not necessarily include tests. In this

thesis, some chapters are more focused on exams, and exam results (chapters 2, 3, and 6), while other chapters are more broadly related to assessment throughout a course (chapters 4 and 5).

An important distinction can be made between summative and formative functions of tests (Black & William, 2003; William & Black, 1996). The aim of a summative assessment is to evaluate student learning after instruction by comparing it to some standard, whereas the aim of formative assessment is to modify teaching and learning activities to improve the learning process. Examples of summative tests are final exams used to decide whether a student has sufficiently achieved the learning goals, or admission tests used to determine whether a candidate is sufficiently skilled to enter a particular education program. Examples of formative tests are diagnostic tests that aim to map parts of the literature that a student does not master yet. Thus, formative tests are assessments *for* learning rather than assessment *of* learning (Schuwirth & Van Der Vleuten, 2011). In this thesis, the focus in chapters 2 and 6 is on summative assessment. Often in educational practice, the distinction between formative and summative assessment is not always clear. This is a challenge for teachers in the implementation of assessment innovations as will be illustrated in chapters 3, 4, and 5.

Another, though subtler, distinction in the literature is between large-scale testing and classroom testing. In the Netherlands, examples of large scale testing are the national end-of-primary school tests (delivered by various organizations, e.g., CITO), and the national exams at the end of secondary education. In the US, a well-known example of large scale high-stakes testing is the SAT that is administered at the end of high school and used by colleges to select students. Other large-scale tests are admission tests such as the Law School Admissions Test (LSAT), and the Graduate Record Examinations (GRE). Large-scale tests are subject to very stringent quality criteria, requiring large amounts of resources. Furthermore, given the scale of testing, advanced statistical methods have been developed to evaluate the characteristics and quality of these tests and test items. Classroom testing, on the other hand, occurs in the classroom in the context of a learning process. The aim of classroom testing is to facilitate and evaluate the learning process of students. Tests used in the classroom are selected or developed by teachers, on a much smaller scale, and teachers generally do not have the resources, time, or amount of students to extensively evaluate the quality of tests. There is a growing interest in combining the strengths of both large-scale - and classroom testing, as a first convention on classroom testing was organized by the National Council on Measurement in Education in the US in 2017. In this thesis, the focus is on classroom testing, but statistical techniques developed in the context of large-scale testing are used in chapters 3 and 4.

1.1.2 The context of the research conducted in this thesis

All the studies in this dissertation were conducted at the University of Groningen, the Netherlands, and most chapters considered courses in the first year of the bachelor program. Study results in the first year are important because they have strong predictive validity for later study success (e.g., Niessen, Meijer, & Tendeiro, 2016) and because there is a binding study advice (BSA, in Dutch: bindend studie advies), that is, Dutch law requires

bachelor degree programs to inform students at the end of the first year whether they are allowed to continue their study. In practice, this advice is translated into rules about the number of credits students must minimally obtain; for example, some universities require that students have to obtain 45 out of the 60 European Credit Transfer and Accumulation System (ECTS) points at the end of their first year. If students do not attain sufficient ECTS points, they are not allowed to continue their study. In the Netherlands there is also a financial incentive for students to perform well: If they decide to quit after January tuition fees are not reimbursed. As a result of the binding study advice, in the first year of university education the stakes are high for students. Most of the first year courses are assessed by means of a final exam, largely consisting of multiple choice questions. The studies in this thesis were conducted in collaboration with teachers seeking to improve or innovate their assessment by implementing changes, often by means of technology.

1.2 Short introduction to each chapter of the thesis

This dissertation is organized as follows: In chapter 2 the implementation of computer-based exams is studied. Digital testing such as computer-based or web-based testing has the potential to ease and improve assessment in higher education. Nevertheless, there have been concerns about equality of test-modes, the fairness, and the stress students might experience (e.g., Whitelock, 2009). In order to ensure a smooth transition from traditional paper-based exams to computer-based exams in higher education, it is important that students perform equally well on computer-based and paper-based administered exams. If, for example, computer-based administration would result in consistently lower scores than paper-based administration, due to unfamiliarity with the test mode or due to technical problems this would result in biased measurement. Thus, it is important that sources of error, or construct irrelevant variance (Huff & Sireci, 2001), which may occur as a result of administration mode, are prevented or minimized as much as possible in high-stakes exams. The main research questions guiding this chapter were: How do students experience computer-based exams? And, is there a difference in student performance depending on mode of examination or preference for mode of examination?

In Chapter 3 the question was addressed whether and when reporting sub-test scores (or subscores) of assessment in higher education is useful. Given the limited time and resources teachers in higher education have given the large classes, teachers and management are often interested in efficient ways of giving students diagnostic feedback. Providing information on the basis of subscores is one method that is often used in large-scale standardized testing, and is more and more often under consideration in classroom testing in higher education as well. After a discussion of recent psychometric literature that warns against the use of subscores in addition to the use of total scores, I illustrate how the added value of subscores can be evaluated using two college exams: A multiple choice exam and a combined open-ended question and multiple choice exam. These formats are often used in higher education and represent cases in which using subscores may be informative.

In chapter 4 the focus is on implementing practice tests. There is a wealth of research on the positive effect of assessment on learning (e.g., Roediger & Karpicke, 2006), and the use of formative assessment to improve student learning is generally recommended. Nevertheless, it is unclear how to most effectively implement practice tests in college psychology education. In chapter 4 students' use of practice test resources was explored and evaluated in light of student performance. First, the relationship between student use of practice test resources and exam results was investigated in three courses with different types of implementations of practice tests. In a follow-up study for one of the courses, the performance of cohorts with practice test resources was compared to a cohort without practice test resources by means of test equating, a technique developed in the context of large-scale testing.

In chapter 5 the focus is on implementing the flipped classroom (e.g., Abeysekera & Dawson, 2015; Street, Gilliland, McNeil, & Royal, 2015). The flipped classroom is becoming more popular as a means to support student learning in higher education by requiring students to prepare before lectures and actively engage in the lectures. While some research has been conducted with respect to student performance in the flipped classroom, not much is known about students' study behaviour throughout a flipped course. Students' study behaviour throughout a flipped and regular course was explored in chapter 5 by means of bi-weekly diaries. Furthermore, student references to their learning regulation and study behaviour were explored in the course evaluations.

Chapter 6 was inspired by the research conducted in chapters 2 through 5 and is more methodologically oriented. To investigate the effect of innovations in the teaching-learning environment, researchers often compare student performance from different cohorts over time, or from similarly designed courses in the same academic year. However, it is important to recognize that variance in student performance can be attributed to both random fluctuation and to various innovations in higher education. Therefore, it is important to take the natural variation in student performance into account. The main question addressed in chapter 6 was: to what extent does student performance in first year courses vary within time, over time, and between courses and how can this information be used to evaluate educational innovations?

Finally, chapter 7 provides an overall discussion of the findings of chapters 2 to 6 and concludes with implications for theory and practice.