University of Groningen

Implementing assessment innovations in higher education

Boevé, Anna Jannetje

Publication date:
2018

Chapter

2

Implementing Computer-Based Exams in Higher Education: Results of a Field Experiment

2.1 Introduction

Computer-based exams (CBE) have a number of important advantages compared to traditional paper-based exams (PBE) such as efficiency, immediate scoring and feedback in the case of multiple-choice question exams. Furthermore CBE allow more innovative and authentic assessments due to more advanced technological capacities (Cantillon, Irish, & Sales, 2004; Csapo, Ainley, Bennett, Latour, & Law, 2012). Examples are the use of video clips and slide shows to assess medical students in surgery (El Shallaly, & Mekki, 2012) or the use of computer-based case simulations to assess social skills (Lievens, 2013). However, there are also drawbacks when administering CBE such as the additional need for adequate facilities, test-security, back-up procedures in case of technological failure, and time for staff and students to get acquainted with new technology (Cantillon, Irish & Sales, 2004). Nevertheless, there have been concerns about equality of test-modes, fairness, and the stress students might experience (Whitelock, 2009)

In order to ensure a smooth transition to computer-based examining 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. As will be discussed below, however, it is unclear from the existing literature whether the different administration modes will result in similar results.

The adoption and integration of computer-based testing in higher education has progressed rather slowly (Deutsch, Herrmann, Frese & Sandholzer, 2012). Besides institutional and organizational barriers, an important implementation consideration is also the acceptance of CBE by the students (Deutsch et al, 2012; Terzis & Economides, 2011). However, as Deutsch et al. (2012) discussed "little is known about how attitudes toward computer based assessment change by participating in such an assessment". Deutsch et al (2012) found a positive change in students' attitudes after a computer-based assessment. As with many studies in prior research (e.g., Deutsch et al., 2012; Terzis & Economides, 2011), this took place in the context of a mock exam that was administered on a voluntary basis. There is little research on student attitudes in the context of high-stakes exams, where students do not take the exam on a voluntary basis.

The aim of the present study was to extend the literature on high-stakes computer-based exam implementation by (1) comparing student performance on CBE with performance on PBE and (2) evaluating students' acceptance of computer-based exams. Before we discuss the design of the present study, however, we first discuss prior research on student performance, and acceptance of computer-based multiple-choice exams. The present study is limited to multiple-choice exams as using computer-based exams in combination with open-question or other format tests, may have different advantages or disadvantages, and the aim of this paper was not to study the validity of various response formats.

2.2 Student performance in computer and paper-based tests

The extent to which different administration modes lead to similar performance in educational tests has been investigated for different levels of education. A meta-analysis on test-administration

mode in K-12 (primary and secondary education) reading education showed that there was no difference in performance between computer-based and paper-based tests (Wang, Jiao, Young, Brooks, & Olson, 2008). A meta-analysis on computer-based and paper-based cognitive test performance in the general population (adults) showed that cognitive ability tests were found to be equivalent in different modes, but that there was a difference in performance on speeded cognitive processing tests, in favor of paper-based tests (Mead & Drasgow, 1993). In the field of higher education, however, as far as we know meta-analyses have not been conducted and results from individual studies seem to vary.

To illustrate the diversity of studies conducted, Table 2.1 shows some characteristics of a number of studies investigating difference in performance between computer-based and paper-based tests with multiple-choice questions in the context of higher education. The studies vary in the number of multiple-choice questions included in the exam, in the extent to which the exam was high-stakes, and in the extent to which a difference in performance was found in favor of a computer-based or paper-based mode of examining. While our aim was not to conduct a meta-analysis, Table 2.1 also shows that many studies do not provide enough statistical information to compute an effect-size. Furthermore, not all studies include a randomized design, implying that a difference cannot be causally attributed to mode of examining. Given these varying findings, establishing that administration mode leads to similar performance remains an important issue to investigate.

Table 2.1. Studies investigating performance differences between paper-based and computer-based tests with multiple-choice questions

	Number of mc-questions	Randomized	High-stakes	Effect size (Cohen's *d*)	Result in favor of
Lee & Weekaron, (2001)	40	no	yes	0.69	paper-based
Clariana & Wallace (2002)	100	yes	yes[a]	0.76	computer-based
Cagiltay & Ozalp-Yaman (2013)	20	yes	yes	0.15	computer-based
Bayazit & Askar (2012)	6	yes	unclear	0.32	paper-based
Nikou & Econo-mides (2013)	30	yes	unclear	0.19	computer-based
Anakwe (2008)	25	no	yes	not possible	
Frein (2011)	3	no	unclear	not possible	
Ricketts & Wilks (2002)	unclear	no	yes	not possible	
Kalogeropoulos et al. (2013)	unclear[b]	yes	unclear	not possible	

[a]the test counted for 15% of the final grade
[b]5 mc-items - but reported means for the mc-test are larger than 5

2.3 Student acceptance of computer-based tests

It is important to understand student acceptance of computer-based testing because the test-taking experience is substantially different from paper-based exams (McDonald, 2002). In paper-based exams with multiple-choice questions, several questions are usually presented per page, and students have the complete exam at their disposal throughout the time allotted to complete the exam. Common test-taking strategies for multiple-choice exams include making notes, marking key words in specific questions, and eliminating answer categories (Towns & Robinson, 1993; Kim & Goetz, 1993). In computer-based multiple-choice exams, however, standard software may not offer these functionalities. For an example where these functionalities were excellently included see McNulty et al. (2007).

Apostolou, Bleu, and Daigle (2009) found mostly negative appraisals of computer-based testing by students in accounting, recommending more research be conducted into what aspects of computer-based exams are important to students. In a mock-exam environment, Wibowo, Grandhi, Chugh, and Sawir (2016) found that most students experienced the exam in a computer-based mode as more stressful compared to the paper-and-pencil mode of examining. While about three quarters of the students who participated in this study were willing to take a digital exam in the future, about half of the students clearly still preferred a paper-based exam. Dermo (2009) investigated student perceptions of the computer-based mode of examining, including both the formative and summative contexts, and found that on average students opinions were rather neutral towards mode of examining. While students were not invited to clearly indicate whether they preferred a particular mode of examining, qualitative comments gave the impression that students in the Dermo (2009) study preferred the CBT mode. A limitation of prior research is that evaluation of computer-based tests has sometimes been confounded with the evaluation of other aspects of testing not directly related to the computer-based testing mode. In the studies of Peterson and Reider (2002), as well as Dermo (2009), the operationalization of student perceptions implies that using computer-based testing means increased testing with multiple choice questions rather than open questions. As a result, the outcomes of these studies may reflect students opinions concerning multiple-choice versus open questions rather than their perceptions of examination mode.

A study by Hochlehnert et al. (2011) in the German higher education context showed that only 37% of students voluntarily chose to take a high-stakes exam via the computer, and that test-taking strategies were a reason why students opted for the paper-based exam. Deutsch et al. (2012) showed that the attitudes of medical students in Germany became more positive towards computer-based assessment after taking a practice exam. The context in which students take a mock exam however, is very different to the actual environment of a formal high-stakes exam. Therefore, it is important to investigate both the test-taking experience and student acceptance of computer-based exams in a high-stakes exam.

Based on focus-group interviews, Escudier, Newton, Cox, Reynolds, and Odell (2011) found that students experienced both advantages and disadvantages in making computer-based multiple choice tests. Among the advantages were, for example, the ease of changing answers, and the prevention of cheating. Advantages of the paper-based mode were, for example, the overview over the whole exam, and making notes and highlighting

in the questions. Although most students found the digital assessment acceptable, almost 25% thought it was not acceptable or less optimal than other methods, and 10% of students thought the computer-based mode was unfair.

As one of the few studies in the context of high-stakes exams, Ling, Ong, Wilder-Smith and Seet (2006) found that students preferred the computer-based mode of examining particularly for multiple-choice exams, and less so for open-question exams. Student response rates however were rather low in this study since students were contacted by e-mail after the exam, which means the results could have been biased if students who did not have a mode preference or a paper-and-pencil preference were less likely to respond to the questionnaire.

2.4 The present study

The present study took place in the last semester of the academic year 2013/2014 with psychology students in the first year of the Bachelor in Psychology program (Dutch track), and was replicated in the academic year of 2014/2015 with a new cohort of students following the International track.

The university opened an exam facility in 2012 to allow proctored high-stakes exams to be administered via the computer. In the academic year 2012/2013 there were 101 computer-based exams, and this number increased to 225 exams in 2013/2014. Of these exams, 102 were multiple-choice exams, 155 were essay question exams, 58 were a mix of both formats, and 11 exams were in a different format. Most computer-based exams were implemented via the university's online learning platform NESTOR, which is embedded in Blackboard (www.blackboard.com), but has extra programming modules developed by the university. Within the broad project to implement computer-based exams, an additional collaboration of faculties started a pilot project to facilitate computer-based exams through the Questionmark Perception (QMP) software (www.questionmark.com). Of the multiple-choice exams administered over the two-year period, 62 were administered via QMP and 40 were administered via Blackboard. Nevertheless, the program of psychology had no previous experience with computer-based examining.

The psychology program is a face-to-face based program (in contrast to distance learning). However, for the course that was included in the present study, attending lectures was not mandatory, and students had the option to complete the course based on self-study alone, given that they showed up for the midterm and final exam.

2.5 Method

To evaluate student performance in different exam modes and acceptance of computer-based exams, computer-based examining was implemented in a biopsychology course, which is part of the undergraduate psychology program. Assessment of the Biopsychology course consisted of two exams receiving equal weight in grading; both were high-stakes proctored exams. Since the computer-based exam facilities could not facilitate the whole group of students, half of the students were randomly assigned to make the midterm exam by computer, and the other half of the students were assigned to make the final exam by computer.

In order to examine whether there were mode differences in student performance

on both exams, we analyzed student performance. Student performance data was collected by the University of Groningen for academic purposes. In line with the university's privacy policy, these data can be used for scientific research when no registered identifiable information will be presented. Since the analysis of student grades presented in this study entailed comparing summary measures of student grades for particular exam mode, no registered identifiable information was presented. Therefore, written informed consent for the use of student grades for scientific research purposes was not obtained.

In order to examine student acceptance of computer-based exams, a questionnaire was placed on the exam desks of students, which they could voluntarily fill out, with the knowledge that their response to the evaluation questionnaire could be used for scientific purposes. Furthermore, students were notified of this procedure at the onset of the course. We did not ask students for written informed consent as to whether they were willing to fill out the questionnaire since they were able to choose to fill out the questionnaire voluntarily and anonymously. Since students were aware that their responses would be used for scientific purposes, informed consent was implied when students chose to fill out the questionnaire. This study, including the procedure for informed consent, was approved by, and adhered to the rules of the Ethical Committee Psychology of the University of Groningen[1].

In the psychology program, this was the first time a computer-based exam was implemented. The total assessment of the course in biopsychology (in both years the study was conducted) consisted of a midterm and final exam, which both contributed equally to the final grade. These exams took place in a proctored exam hall. At the start of the course students were randomly assigned to make the midterm exam either by computer or as a paper-and-pencil test. Subsequently the mode of examining was switched for the final exam, so that everyone was assigned to take either the midterm or the final exam as a computer-based test. After completing the computer-based exam, students were invited to fill-out a paper-and-pencil questionnaire on their experience with the computer-based exam, which they could submit before leaving the exam hall. Students received immediate feedback on their performance on the exam in the computer-based condition (number of questions correct), and thus knew their performance on the exam when completing the questionnaire. Students in the paper-based condition received the exam result within a couple of days after taking the exam.

2.5.1 Participants

At the start of the course in the 2013/2014 study there were 401 students enrolled via the digital learning environment. These students were randomly assigned to make the midterm exam via paper-based mode or computer-based mode. If a student was assigned to complete the midterm on paper, the final exam would be completed by computer and vice versa. In the 2014/2015 study there were 428 students enrolled in the course, and these students were also randomly assigned to take the midterm via paper-based mode or computer based mode as in the 2013/2014 study. All students who completed a computer-based exam were

---

[1] http://www.rug.nl/research/heymans-institute/organization/ecp/

invited to evaluate their experience by responding to a paper-based questionnaire directly after completing the CBE. As can be expected in a field experiment, however, there was both some attrition and non-compliance (Figures 2.1a and 2.1b), which we will discuss below.

There were three sources of attrition in the first study (Dutch cohort 2013/2014): (1) not registering for the exams, (2) registering but not showing up at the midterm, and (3) completing the midterm but not showing up for the final exam. These three sources of attrition led to a 16% overall attrition rate (66 students). There were 16 students who completed both the midterm and final exam on paper. In addition, there was a technical failure at the midterm exam, as a result of which 36 students needed to switch to a paper-based exam in order to be able to complete the exam.

In the international cohort of 2014/2015, there were also 3 sources of attrition: students who did not show up for either exam, students who did not show up to the midterm exam, and students who did not show up to the final exam. Note that in 2013/2014 students were required to both enroll in the course and register for the exam separately, while in 2014/2015 the system had changed so that course enrollment automatically implied exam registration. The overall attrition rate in the 2014/2015 cohort was 10% (44 students). There was no technical failure, and one student completed both midterm and final exam in the paper-based mode.
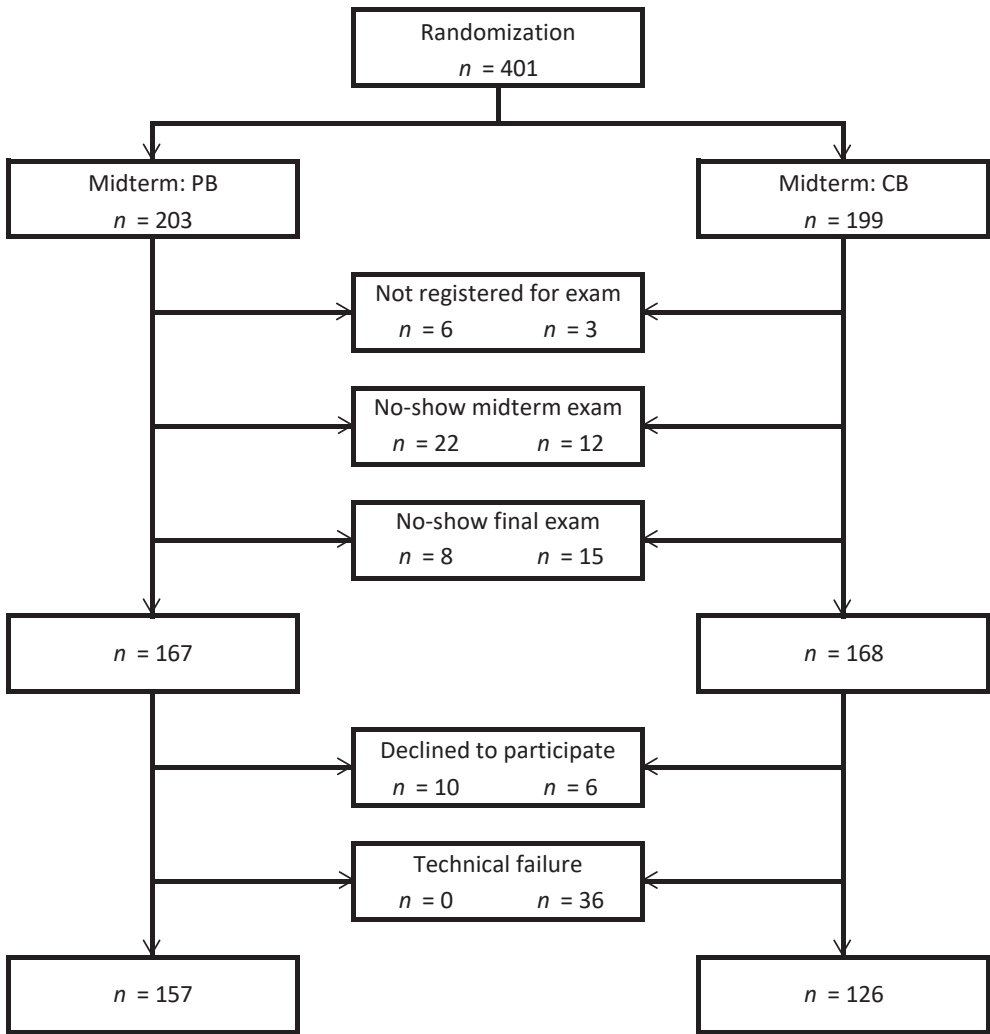
Figure 2.1a. Random assignment of students to different exam modes, subsequent attrition, and non-compliance for the first study (Dutch cohort 2013/2014)
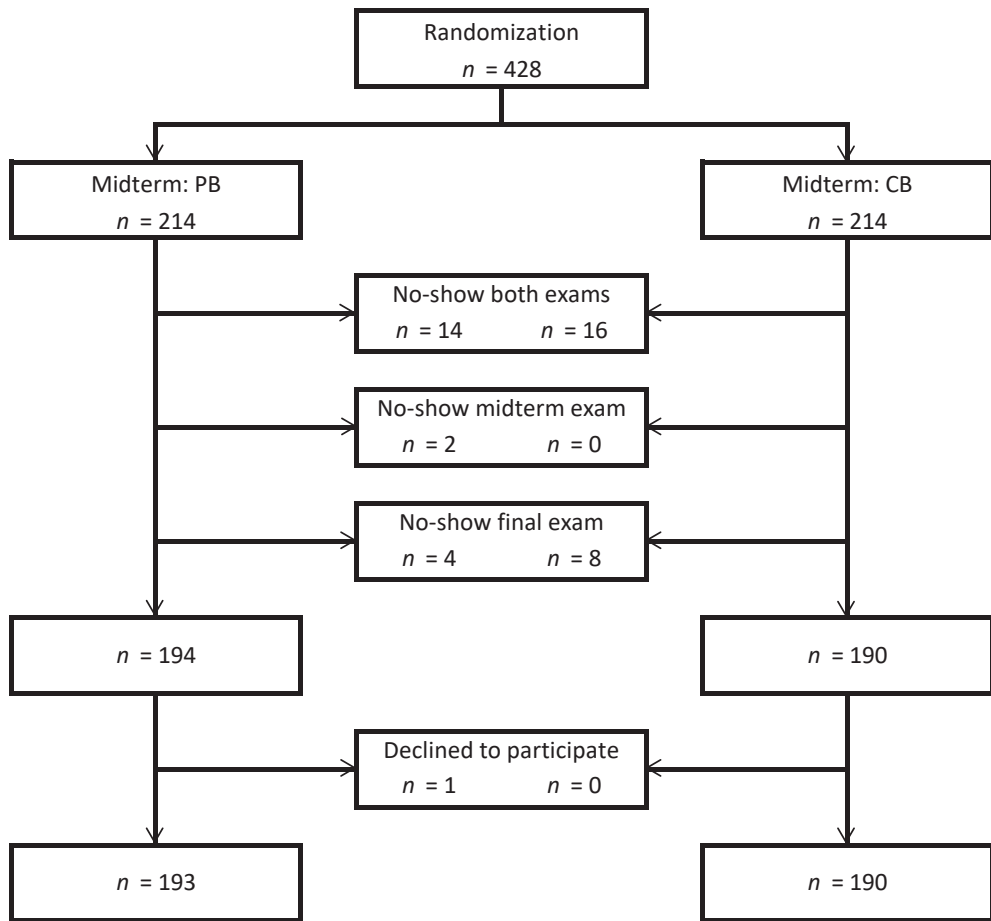
Figure 2.1b. Random assignment of students to different exam modes, subsequent attrition, and non-compliance for the second study (international cohort 2014/2015)

2.5.2 Materials

Student performance. Both the midterm and final exam contained 40 multiple-choice questions with four answer categories. The exams measured knowledge of different topics in biopsychology. The material that was tested on the midterm exam, was not tested again in the final exam. Thus the two exams covered different material included in the course and each exam had an equal weight in determining the final grade. The midterm exam appeared to be somewhat more difficult (mean item proportion correct) in both cohorts (see Table 2.2). Student performance in both modes was investigated by comparing the mean number of questions correct on each exam.

Table 2.2. Exam characteristics of both cohorts, and partial exams, with mean item-total correlations and reliability estimates for the computer-based (CB) and paper-based (PB) mode.

	Proportion correct (Mean)	Item-total correlation (Mean)		Reliability [95% CI]	
2013/2014		CB	PB	CB	PB
Midterm	.72	.32	.29	.78 [.72; .83]	.71 [.66; .76]
Final	.75	.32	.27	.75 [.67; .80]	.66 [.59; .73]
2014/2015					
Midterm	.77	.33	.31	.82 [.79; .86]	.80 [.76; .84]
Final	.80	.33	.35	.82 [.79; .86]	.84 [.81; .87]

Acceptance of computer-based tests. Student acceptance was operationalized in three ways (see Table 2.3). First, students answered questions about their test-taking experience during the computer-based exam and in paper-based exams in general. Second, students were asked whether they preferred a computer-based exam, paper-based exam, or did not have a preference. Third, students were asked whether they changed their opinion about computer-based exams as a result of taking a computer-based exam. Answers to the questions on test-taking experience were given on a five-point Likert response scale ranging from 'completely disagree' to 'completely agree'. The question on whether students' opinions changed had five response options: 'yes, more positive', 'yes, more negative', 'no, still positive', 'no, still negative', and 'no, still indifferent'.

Table 2.3. Evaluations of students test-taking experience and acceptance of computer-based exams

Student acceptance of computer-based exams	
Questions	**Sub-questions**
In this computer-based exam	I was able to work in a structured manner
	I had a good overview of my progress in the exam
	I was able to concentrate well
In paper-based exams in general	I am able to work in a structured manner
	I have a good overview of my progress in the exam
	I am able to concentrate well
I prefer a: paper-based exam, computer-based exam, no preference	
Did your opinion about computer-based ex-ams change as a result of taking this exam?	

2.5.3 Procedure

The midterm computer-based exam in the 2013/2014 cohort was administered through the Questionmark software, but as mentioned above, there was a technical problem. Since the technical issue could not be solved in time, the final exam was administered directly via Nestor (the university's online learning platform). As a result of the change in interface, the design and layout of the computer-based midterm and final exam was slightly different in the 2013/2014 cohort. The midterm exam, administered through QMP, was designed so that all questions were presented simultaneously with a scrolling bar for navigation. In the final exam, administered via Nestor, the questions were presented one at a time and navigation through the exam was done via a separate window with question numbers allowing students to review and change answers given to other questions. In the 2014/2015 cohort, both the midterm and final computer-based exam was directly administered via Nestor, in the same way as the final exam in the 2013/2014 cohort.

For both partial exams in both cohorts students had the opportunity to go back and change their answers at any point and as many times as they liked before submitting their final result. After submitting their final answers to both the midterm and final exam in the computer-based mode, students immediately received an indication of how many questions they answered correctly. For the paper-based mode of examining, students took a list of their recorded answers home, and could calculate an indication of how many questions they answered correctly several days after the exam when the answer key was made available in the digital learning environment.

2.6 Results
2.6.1 Student Performance
Table 2.4 shows that there was no statistically significant difference in the mean-number of questions answered correctly between the computer-based and paper-based mode for both the midterm and final exam in both 2014 and 2015.

Table 2.4. Mean number of questions correct in the different exam conditions for the midterm and final exam

	Computer-based		Paper-based			
2013/2014	*n*	*M(SD)*	*n*	*M(SD)*	*t(df)*	Cohen's *d* [95% CI]
Midterm	126	28.56 (5.31)	157	28.50 (4.61)	0.10 (281)	0.01 [-0.22; 0.25]
Final	157	29.92 (4.60)	126	29.50 (4.32)	0.78 (281)	0.09 [-0.14; 0.33]
2014/2015						
Midterm	190	31.11 (5.69)	193	30.82 (5.47)	0.50 (381)	0.05 [-0.15; 0.25]
Final	193	32.12 (5.35)	190	32.05 (5.72)	0.13 (381)	0.01 [-0.19; 0.21]

2.6.2 Student acceptance of CBE
Test-taking experience. In Figure 2.2 the mean scores on the questions with respect to test taking experiences for the midterm and final exam are provided. A multivariate ANOVA was conducted to examine whether these questions were evaluated differently for the midterm and final exam. Results of the overall model test (α = .05) showed that there was a difference in how the questions were evaluated between the midterm and final exam (2013/2014: $F(6,258) = 7.02$, $p < .001$, partial-η^2 = .14; 2014/2015: $F(6,320) = 3.87$, $p = .001$, partial-η^2 = .07).

Additional (Bonferroni corrected α=.0083) univariate analyses showed that students in the 2013/2014 cohort were less able to concentrate in the midterm computer-based exam compared to the final computer-based exam ($F(1,265) = 22.39$, $p = .00014$). Students in the 2014/2015 cohort on the other hand were less able to monitor their progress ($F(1,325)$ =11.78, $p = .0007$) and concentrate ($F(1,325) = 11.39$, $p = .0008$) in the computer-based final exam compared to the computer-based midterm exam. See Table A2 in the appendix for more details on the means, standard-deviations, and effect sizes.
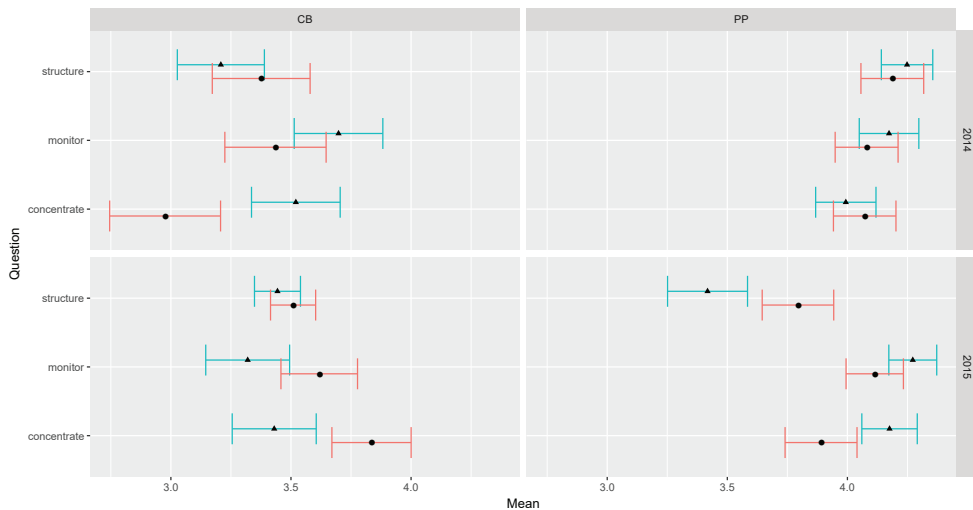
Figure 2.2. Mean scores and 95% confidence intervals for student approaches to completing the computer-based exam, and paper-based exams in general for the midterm (red line) and final exam (blue line).

To examine the difference in test-taking experience between the computer-based exam and paper-based exams in general, Bonferroni corrected (α = .017) dependent-sample *t*-tests were conducted. Table 2.5 shows that students were more positive in terms of their ability to work in a structured manner, monitor their progress, and concentrate during paper-based exams compared to the computer-based exam, with medium (0.33) to large (0.64) effect sizes.

Table 2.5. Mean difference between computer-based and paper-based exam evaluation, with dependent-sample *t*-test results and effect-size

2013/2014	CB – PB mode [95% CI]	*t*(*df*)	Cohen's *d* [95% CI]
Structured approach	-0.94 [-1.12; -0.77]	-10.71 (268)	-0.65 [-0.83; -0.48]
Monitor progress	-0.54 [-0.72; -0.37]	- 6.15 (269)	-0.37 [-0.54; -0.20]
Concentration	-0.74 [-0.92; -0.56]	- 8.13 (269)	-0.46 [-0.67; -0.32]
2014/2015			
Structured approach	-0.72 [-0.87; -0.57]	-9.64 (333)	-0.53 [-0.68; -0.37]
Monitor progress	-0.40 [-0.57; -0.23]	-4.75 (334)	-0.26 [-0.41; -0.11]
Concentration	-0.43 [-0.57; -0.30]	-6.22 (332)	-0.34 [-0.49; -0.19]

Preference for computer-based exams. In the 2013/2014 cohort, 50% of the students preferred a paper-based exam, 28% preferred a computer-based exam, and 22% indicated that they did not have a preference for one mode over another after completing the computer-based exam. There was no difference in preference for a particular exam-mode between students who completed the midterm and final exam via the computer

(Fisher's exact p = .97). In the 2014/2015 cohort, there was a difference in preference for exam mode between students who completed the midterm and final exam via the computer (Fisher's exact p =.007). For students who took the midterm by computer, 38.5% preferred a computer-based exam, 38.5% preferred a paper-based exam, and 23% did not have a preference. For students who took the final exam by computer, 23% preferred a computer-based exam, 51% preferred a paper-based exam, and 26% did not have a preference.

With respect to the change of opinion towards computer-based assessment after taking a computer-based exam, in the 2013/2014 cohort: 43% of students felt more positive, 14% felt more negative, 15% were indifferent, 16% still positive, and 12% still negative towards computer-based exams, with no difference between the midterm and final exam (Fisher's exact p = .12). In the 2014/2015 cohort, of the group who took the midterm by computer 44% were more positive, 9% more negative, 16% indifferent, 24% still positive, and 6% still negative towards taking computer based exams. Of the group who took the final exam by computer however, 44% were more positive, 22% more negative, 11% indifferent, 13% still positive, and 11% still negative towards taking computer-based exams consisting of multiple-choice questions.

2.7 Discussion

2.7.1 Student performance

In line with recent research (Cagiltay & Ozalp-Yaman, 2013; Bayazit & Askar, 2012; Nikou & Economides, 2013), we found no difference in the mean number of questions correct between computer- and paper-based tests for both the midterm and final exam. Earlier findings in the field of higher education in favor of paper-based tests (Lee & Weekaron, 2001), and in favor of computer-based tests (Clariana & Wallace, 2002), were not replicated in this study. Based on these findings, we can conclude that recent findings show that exam-mode may not cause differential student performance in higher education. An important explanation for this finding could be the population of students in this study. Students in this study entered the higher education system largely directly after completing secondary education and represent a generation that has grown up with technology. Earlier studies on the use of computer-based testing may have found a difference in favor of paper-based tests as a result of test takers' unfamiliarity with technology. Therefore, the lack of a difference in performance between modes in the present study may be the result of a generational difference in student population compared to older studies. This also implies that current studies with older populations of students may still find a mode effect, although adults today will have had more technology exposure in daily life than studies conducted with adults twenty years ago.

2.7.2 Student acceptance of CBE

Students indicated that the test-taking experience in PBE in general was more favorable compared to CBE in terms of their ability to work in a structured manner, have a good overview of their progress through the exam, and their ability to concentrate. While there was no difference in performance for computer-based and paper-based exams, these findings suggest that students appear to feel less in control when taking a computer-based exam relative to a paper-based exam. This is in line with previous findings by Hochlehnert et al. (2011) who found that the

absence of functionality to apply test-taking strategies was a reason for students not to choose a computer-based exam. Further research is necessary to see if this difference in approach to taking the exam may be an artefact of the first-time introduction to computer-based exams. Students who regularly take computer-based exams may be more accustomed to this mode, and therefore have developed confidence in their approach to taking computer-based exams. Another avenue that may be pursued in order to better understand the test-taking experience in CBE may be to extend the research of Noyes, Garland and Robbins (2004) who found that students experienced a higher cognitive load in a short computer-based multiple-choice test compared to an equivalent paper-based test. Further research could investigate the extent to which the perceived test-taking experience is related to cognitive load.

In the 2013/2014 cohort, we found that students who took the final exam by computer, were able to concentrate better on average than students who took the midterm exam by computer. A possible explanation for this result may be the technical problem during the midterm. Students in the computer-based exam hall who did not experience the technical problem, may have been affected indirectly by the unrest in the exam hall as the directly affected students were provided with a paper-based exam. If this were the explanation for the difference in concentration between the midterm and final exam, it would seem logical that students who completed the midterm exam were also more negative about computer-based exams compared to the group of students who completed the final exam by computer. We found no difference, however, in the extent to which student opinions became more negative towards CBE after taking the computer-based exam.

Another possible explanation for the difference in the ability to concentrate between the midterm and final exam in the 2013/2014 cohort is the design of the computer-based assessment. According to Ricketts and Wilks (2002) a difference in design from scrolling through all questions to a one-question-at-a-time format explained improved student performance. In the present study all the questions were displayed simultaneously in the midterm file, while in the final exam questions were presented one at a time. In presenting questions one at a time during the final exam students may have been able to focus better on the questions at hand, explaining the greater ability to concentrate reported by students.

The replication in 2014/2015 showed a larger difference in experience between students who took the midterm exam by computer and students who took the final exam by computer, with students who took the midterm exam by computer generally being more favorable about their experience in the CBT compared to students who took the final exam by computer. Furthermore, this difference was also reflected in a difference in student preference for exam mode between students who took the midterm or final by computer. In terms of ability to concentrate in the computer-based exam, results were different from the 2013/2014 findings, namely students in the midterm computer-based exam indicated being able to concentrate better than students in the final computer-based exam. In the 2014/2015 cohort, however, both the midterm and final exam were designed to present one question at a time, and thus no difference in exam experience would be expected. There were no technical failures, or problems with the administration of computer-based exams in 2014/2015, therefore the difference in experience between students who took the midterm and final by computer is difficult to explain.

2.7.3 Practical implications

Based on the above discussion there are several practical implications for Universities seeking to implement CBE. Student performance on multiple choice question exams does not appear to vary across test mode. The benefits of CBE, and the lack of negative consequences, can both be used in the communication towards students prior to the first implementation of CBE in order to maximize acceptance. Furthermore, universities need to invest in good CBE exam facilities. This includes investing in adding more test-taking functionalities so that students test-taking experience may be as optimal as possible. Furthermore, the potential of technical failure is a risk that requires good protocols so that students are able to complete the exam either on a different computer or on paper.

The full potential of computer-based tests can be realized in further developments. One option is to use computer adaptive testing (CAT). The advantage of CAT is that items are chosen from an item pool that best fit the level of the candidate. In many higher education institutes however, this is difficult to realize as a very large item pool with regular refreshment is needed. In combination with the extensive psychometric knowledge necessary for this development, this is generally beyond the scope of many university courses. What may be easier to realize however, is to offer test items to students in random order, which helps prevent cheating.

2.7.4 Limitations

There were several limitations to the present study. First, there were technical problems during the midterm computer-based exam in the 2013/2014 cohort. As a result of this technical failure a number of students had to complete the planned CBE on paper. This remains a risk for computer-based exams in general, and the facilities for computer-based examining need to be organized in such a way that when this occurs unexpectedly in practice, hindrance for students is minimized. In the present study, students were allowed extra time to complete the exam, although no one made use of it. It is important to note, that while students may not have a good test-taking experience, their results are unlikely to suffer as a consequence. Several studies have shown student performance in CBE is not affected by technical issues (Sinharay, Wan, Choi & Kim, 2015; Sinharay, Wan, Whitaker, Kim, Zhang, & Choi, 2014).

An important aspect of introducing computer-based assessment deserves mention as well, namely the teacher of faculty perspective. Since the present study was conducted in a single course, the teacher perspective was outside the scope of the present study. Research into teacher acceptance and willingness to implement computer-based assessment may also provide relevant insight into improving the implementation of computer-based exams in higher education.

Furthermore, our sample consisted of students who were primarily 'traditional' students and started their study soon after completing high school. A population containing more mature aged students may view technology differently. In addition students were studying face to face. Students who study via distance mode may view computer-based testing differently than face-to-face students.

2.7.5 Conclusion

This study found that students performed equally well in computer-based multiple-choice exams compared to paper-based exams. While paper-based exams may be the norm in many universities, investing in computer-based exams may be beneficial for the younger generation who are more and more growing up with computer and digital technologies. Further research is necessary into the optimal design of computer-based exams, such that student-acceptance is maximized and not an irrelevant source of stress during exams in a high-stakes context.