

# The relative impact of in-class closed-book *versus* take-home open-book examination type on academic performance, student knowledge retention and wellbeing

Tali Spiegel & Amy Nivette

To cite this article: Tali Spiegel & Amy Nivette (2023) The relative impact of in-class closed-book *versus* take-home open-book examination type on academic performance, student knowledge retention and wellbeing , Assessment & Evaluation in Higher Education, 48:1, 27-40, DOI: [10.1080/02602938.2021.2016607](https://doi.org/10.1080/02602938.2021.2016607)

To link to this article: <https://doi.org/10.1080/02602938.2021.2016607>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 21 Dec 2021.



Submit your article to this journal [↗](#)



Article views: 5218



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 5 View citing articles [↗](#)

# The relative impact of in-class closed-book *versus* take-home open-book examination type on academic performance, student knowledge retention and wellbeing

Tali Spiegel  and Amy Nivette

Sociology Department, Utrecht University, Utrecht, The Netherlands

## ABSTRACT

This study investigates the relationship between take-home (open-book) examinations (THE) and in-class (closed-book) examinations (ICE) on academic performance and student wellbeing. Two social science courses (one bachelor and one master) were included in the study. In the first cohort (2019), students from both courses performed an ICE, whereas students in the second cohort (2020) performed a THE. Four to six months following course completion, students were approached to fill out a survey pertaining to their academic performance and wellbeing during the course, and to complete a test measuring knowledge retention on the course materials. No significant differences were found between the ICE and THE cohorts in academic performance and knowledge retention for either the bachelor or the master students. Bachelor students who completed a THE in 2020 reported significantly lower wellbeing in comparison to their peers completing the ICE a year earlier. The implications of the results in the context of the Covid-19 pandemic are discussed.

## KEYWORDS

Take-home examination (THE); in-class examination (ICE); knowledge retention; wellbeing

## Introduction

There is a large body of literature aiming to shed light on which test type, in-class (closed-book) examination (ICE) or take-home (open-book) examination (THE), is more beneficial to students' educational performance and wellbeing. Bengtsson (2019) reviewed 35 studies comparing outcomes following ICEs and THEs and found that while there are a few points of debate there are some clear advantages to using an ICE. ICEs appear to be a more effective tool for testing the lower levels of Bloom's taxonomy scale (Forehand 2010) such as remembering (Rich 2011), and are associated with less student cheating (Berrett 2012; Tao and Li 2012). However, Bengtsson (2019) argues that there seem to be two clear advantages to THEs: first, THEs appear to be more productive for developing and testing higher levels of the Bloom's taxonomy scale (e.g. analysing and evaluating), as the student can invest their energy into deeper engagement and understanding rather than memorising the material (Hagström and Scheja 2014). Second, THEs are associated with lower levels of student anxiety (Zoller and Ben-Chaim 1989).

Existing research assessing the advantages of THEs tends to be limited in two ways. First, an important aspect of determining to what extent a test can adequately assess higher levels

of Bloom's taxonomy scale is to evaluate whether the knowledge acquired during the course was retained *over time* (Whitley and Parton 2014). However, there is not yet a clear answer on which examination type is more beneficial for knowledge retention. Some studies suggest THEs are not effective for knowledge retention because students are less likely to study for the examination if they know it is an open-book examination (Marsh 1980, 1984; Moore and Jensen 2007), while others suggest that students engage more deeply with the material when they know in advance that they will be performing a THE (Rich 2011). Some scholars find no difference in knowledge retention between the two examination types (Agarwal et al. 2008), while others suggest that mid-way alternatives such as cheat sheets (a document containing notes made by the student to assist them in recalling material during the examination) are most effective (Nsor-Ambala 2020). The first aim of this study is, therefore, to measure and evaluate to what extent examination type is associated with long-term knowledge retention and overall academic performance.

Second, while most studies suggest that THEs are related to lower anxiety levels, it is also suggested that the way the examination is conducted is important to understanding whether or not students experience reduced anxiety (Nsor-Ambala 2020). This relationship seems to relate to various examination characteristics such as the location (in class or at home), the accessibility of the materials and differences in time constraints (Durning et al. 2016; Bengtsson 2019). Thus, the second aim of this study is to examine the extent to which differences in examination type are associated with wellbeing outcomes. Specifically, we assess differences between students who performed a timed ICE and those who performed a timed THE.

In order to address these aims, we compare academic and wellbeing outcomes among two subsequent cohorts of bachelor and master students: one cohort that completed an ICE, and one cohort the following year that completed a THE for the same course. For each cohort, we measured knowledge retention, overall academic performance, skills development, wellbeing and workload stress 4–6 months after the completion of the course.

### ***Academic performance and long-term knowledge retention***

Research on examination type and academic performance suggests that students who know they will take a THE are less likely to attend lectures and to prepare for the examination as they expect to be able to rely on the materials during the examination (Moore and Jensen 2007). ICEs require students to engage with the material thoroughly in order to be able to perform on the examination, meaning students taking ICEs should score higher on their examination than peers taking a THE (Agarwal and Roediger 2011). In particular, weaker students may score lower on a THE with a time constraint as they will struggle to obtain and process the answers within the time frame of the examination (Boniface 1985). In line with this research, studies examining knowledge retention outcomes typically find that students taking THEs tend to rely on 'surface learning' and will therefore not retain as much information over time as peers who took an ICE (Moore and Jensen 2007).

By contrast, some argue that THEs are more effective for academic performance, measured most frequently by examination grades (Block 2012; Ackerman and Leiser 2014). This research suggests that THEs are more effective at examining higher levels of the taxonomy scale. Roelle and Berthold (2017) show that, provided the student is given more time to engage with the material during a THE, they will outperform peers taking an ICE if the questions are more complicated, but students will generally perform better on an ICE if the questions are of lower complexity. Others suggest that THEs lead to marginally higher examination scores, not just due to the deeper learning it requires but also because they are associated with reduced anxiety amongst students, which is known to impede academic performance (Block 2012). Examinations that test higher levels of the taxonomy scale do not lend themselves to solely scanning the

material during the examination and encourage the student to gain deeper understanding of the materials to be able to score well. In accordance with this line of research, studies that do find a link between THEs and long-term knowledge retention suggest that deeper learning (as opposed to surface learning) takes place when the type of examination requires higher-level taxonomy application and problem-solving skills, which are in turn key to long-term knowledge retention (Smith 1999; Rich 2011; Johanns, Dinkens, and Moore 2017).

Both lines of argumentation make clear that deeper learning is a key factor to the effectiveness of a test for academic performance. Whether or not a THE requires deeper learning depends on the format of the examination. Studies have shown that for instance THEs are not suitable for multiple-choice examinations (Zoller and Ben-Chaim 1989), and that having an open-book condition is important for encouraging students to interact more deeply with the materials as part of the THE (Hagström and Scheja 2014). Further, group THEs may test higher-level taxonomy skills but may be more vulnerable to free riding behaviour (Hall and Buzwell 2013), whereby not all group members engage in deeper learning. Examinations without a relatively strict time constraint may result in students not preparing for the examination and relying on looking up the answers during the examination (Durning et al. 2016).

One methodological concern to consider is the timing of the knowledge retention test. Studies that assess long-term knowledge retention focus on a follow-up time ranging from a few days to a few weeks after the THE or ICE (Haynie 2003; Moore and Jensen 2007; Rich 2011). The retention test typically takes place within the course, either in the form of a popup ICE test a few days after the examination or as an ICE final examination (e.g. Agarwal 2009; Nsor-Amabala 2020). As far as we know only the study by Rummer et al. (2019) tested knowledge retention with a lengthier time lag of 8 weeks as they claim that a longer delayed retrieval test can more appropriately capture long-term retention. By contrast, similar studies examining the impact of teaching and learning styles on long-term knowledge retention follow up with students from five months up to 2 years after the course (Herzig et al. 2003; Lynse and Miller 2017; Taglieri et al. 2017). This study, therefore, extends the follow-up time frame beyond previous studies, which will help discern whether advantages found in the literature so far can be attributed to a short-term advantage or to a long-term retention.

### ***Student wellbeing***

Different examination conditions may relate differently to student wellbeing. This can be measured in different ways including student satisfaction, happiness and in particular stress and anxiety (Durning et al. 2016). Examinations which take place at home and allow for consultation with the materials may be considered more relaxing and less anxiety-provoking for the student (Zoller and Ben-Chaim 1989). Thus, when preparing for an open-book examination (Phillips 2006) or for a THE students often experience less anxiety than with an ICE (Weber, McBee, and Krebs 1983; Fernald and Webster 1991; Tao and Li 2012; Dave, Dixon, and Patel 2020; Akulwar-Tajane et al. 2021). Studies also show that THEs relate to more positive learning experiences (Tao and Li 2012; Senel and Senel 2021). Some studies find that students experience both ICEs and THEs as difficult and anxiety-provoking when they take the same form as the final course examination, as the examination is then seen by the student as more relevant to their successful completion of the course (Krarup, Naeraa, and Olsen 1974). Durning et al. (2016) reviewed studies comparing academic and wellbeing outcomes between open-book and closed-book examinations (without a clear focus on the examination location), and concluded that the impact of open-book examination on reduced anxiety might be overstated. The authors found that while students associate open-book examination with lower anxiety they do not necessarily score lower on anxiety measures.

## ***This study***

Given the research and limitations outlined, this study aims to evaluate to what extent THEs are advantageous for academic performance and wellbeing outcomes among students. Specifically, this study advances knowledge on the relative performance of examination types by assessing both academic and wellbeing outcomes among two cohorts of university students: one cohort that completed an ICE in 2019 and one that completed a THE in 2020. Long-term knowledge retention was assessed among students roughly 4–6 months following the completion of the course. In addition to knowledge retention, several academic and wellbeing outcomes are measured: perceived wellbeing and stress during the course, perceived skills development and overall grades in the course. The format of the THE consisted of open-ended essay questions which were announced at the beginning of the course. Based on previous research, these conditions should be more conducive to deeper learning and testing higher levels of the taxonomy scale, leading to longer-term knowledge retention (Johanns, Dinkens, and Moore 2017). Given the location (home) and format of the (open-book) THE, students may experience higher wellbeing and reduced stress during the course. However, given the disagreement within the literature, it is also possible that students report higher levels of stress and lower wellbeing when preparing for and taking the THE or that no difference will be found. Furthermore, we explore to what extent there are differences in effects between bachelor and master-level students.

## **Methods**

The data for this study were drawn from surveys conducted among students following a bachelor or master course in a social science programme at Utrecht University in the Netherlands. The bachelor course was an upper-level course open to students in the faculty as well as international exchange students. The master course was situated within a 1-year master programme, and was also open to students from other master programmes in the university. All students who passed these courses were approached at the end of April 2020 and at the end of April 2021 to fill out a survey 4 (master course) to 6 months (bachelor course) following the completion of the course. In both courses, the final grade was composed of the average between the examination grade and additional graded activities (e.g. paper and/or presentation). Compensation of the grade between the examination and the additional assessment forms was not possible. That means that a student would have to pass their examination to pass the course, irrespective to how they performed on the written assignments. As such, each student who was approached who had passed the course had also passed the examination.

Data were collected from two cohorts of students during the academic years 2019–2020 and 2020–2021 (here on referred to 2019 and 2020). If students did not pass the final examination, they may have been eligible to participate in a retake examination if they met certain requirements outlined within the course (e.g. participation). In both courses, students must have scored a 5.5 on their examination in order to pass (on a scale from 0 to 10). In the academic year 2019, students completed an ICE. In the bachelor course, students completed a midterm and final examination. The mid-term examination covered the first 4 weeks of material, and consisted of lower-level taxonomy questions in the form of multiple choice and short open answer questions. The second examination covered the remaining weeks' materials, and consisted of essay questions requiring higher-level taxonomy skills. In the master course, the examination was taken at the end of the course and covered all the literature and (guest)lecture materials. The examination consisted of two parts. Part one consisted of three essay questions on three-course themes covered in the literature and the lectures. Students had to select two of these three questions to answer. Part two contained one essay question on the course guest lectures. The choice element was announced two weeks prior to the examination. The examination contained

questions testing both lower and higher taxonomy level skills. Students completed the examinations using pencil and paper in both courses.

In the academic year 2020, students completed a THE. In the bachelor course, the format of the examinations was changed and students only completed one final examination covering all course materials. The examination had four versions each linked to a different intervention programme, with four identical essay questions asking students to apply the knowledge from the course to explaining and evaluating the programme. While the format changed, the content of the course and materials remained relatively the same across cohorts. In the master course students had an examination which covered all of the reading materials and course lectures. The examination was composed of five essay questions, of which students had to choose four to complete. The questions primarily required the application of higher-order taxonomy skills.

In the 2019 academic year, 166 students took part in the bachelor course of which 148 successfully passed the course. Of the 148 students who were approached *via* email, 50 students completed the survey (37.7% response rate). It is important to note that the bachelor course had roughly 40 international exchange students who may have already returned to their home countries and thus have had limited access to their university email at the time of the survey. In the master course, 60 students took part, of which 58 passed the course. Out of the 58 students who were approached, 34 completed the survey (58.6% response rate).

In the 2020 academic year, 147 students participated in the bachelor course, of which 138 passed. Two of the email addresses were no longer in use. Of the 136 students approached, 48 students completed the survey (35.3% response rate). There were no international students in the bachelor course due to COVID-19 restrictions. In the master course, 55 students were enrolled in the course, of which 53 passed. One email was returned as undeliverable. Thirty students completed the survey (response rate of 57.7%).

For each cohort, students who participated were offered the opportunity to enter a raffle where they could win a voucher of 25 euros.

## **Survey design**

Each survey contained two parts. In the first part, students were asked to indicate to what extent they developed academic skills as a result of the course, and to report on their wellbeing and perceived workload stress at the time of the course. In the second part, students were asked to complete a 10-item multiple-choice test encompassing the in-class material from their respective course examinations.

The bachelor course test included questions on both reading and lecture material, covering topics from both the mid-term and final examination in academic year 2019 and the final examination in year 2020. The knowledge test for the bachelor course remained the same across cohorts. The master course test contained questions on the lectures, the reading material and the guest lectures. In the academic year 2019, students answered questions related to the two themes they had selected for their ICE. Depending on the combination of themes they chose they would be directed to one of three tests. In the academic year 2020, the test was equal amongst all students as they were not informed of the choice component and, therefore, had studied the entirety of the course material for the examination.

## **Measures**

### **Academic performance outcomes**

**Knowledge retention.** For each course, 10 knowledge questions were asked that covered the main arguments, concepts and debates in the course. Each multiple-choice question

had one correct answer. The sum of correct answers is computed to determine the student score. This grade can range between 0 and 10.

**Self-reported grades.** Students were asked to report their highest examination grade (including the retake examination if applicable) and their overall course grade, which was a weighted combination of the examination grade and any additional assignments.

**Skills development.** The original Generic Skills Scale (GSS; Byrne and Flood 2003) contained six items such as 'the course developed my problem-solving skills' with the answer categories: 'strongly disagree', 'disagree', 'neutral', 'agree' and 'strongly agree'. Two items were added to this scale that were important for these courses, which included policy evaluation elements: 'the course enhanced my ability to critically evaluate policy' and 'the course enhanced my knowledge in the field.' The Cronbach's alpha in 2019 was 0.75 (bachelor) and 0.71 (master), and 0.77 (bachelor) and 0.64 (master) in 2020.

### **Wellbeing outcomes**

**Student wellbeing.** The wellbeing scale was drawn from the World Health Organisation, and contained five items including, e.g.: 'I felt cheerful and in good spirits' (WHO [Five] Well-Being Index (1998 version)). There were six answer categories ranging from 'all of the time' to 'at no time'. The Cronbach's alpha for this scale was 0.84 (bachelor) and 0.83 (master) in 2019 and 0.86 (bachelor) and 0.78 (master) in 2020. The scale was reversed so higher scores indicate lower perceived wellbeing.

**Study workload stress.** Perceived workload stress was measured using eight items adapted from the original Job Stress Scale (Shukla and Srivastava 2016). For example, one item states 'I had a high study load and feared that I had very little time to do it' with the following answer categories: 'strongly disagree', 'disagree', 'neutral', 'agree' and 'strongly agree'. For the bachelor course, Cronbach's alphas were 0.86 in academic year 2019, and 0.90 in 2020. Cronbach's alphas for the master course were 0.91 and 0.88 for the academic years 2019 and 2020, respectively.

### **Demographic characteristics**

We included two variables to reflect demographic characteristics. Sex at birth included the answer categories 'male' (coded 0) and 'female' (coded 1). Students also indicated whether or not they were a Dutch (coded 0) or international student (coded 1).

### **Ethics**

This study was approved by the ethics review board of the Faculty of Social and Behavioural Sciences at Utrecht University. At the beginning of the survey, students reviewed the consent form and had to indicate their consent and participation by selecting 'I consent' prior to filling out the rest of the survey. All responses remained anonymous and students were informed that they could quit the survey at any time.

### **Analytical approach**

The goal of this study is to assess to what extent ICEs *versus* THEs are associated with wellbeing and academic outcomes. In order to do so, we utilised independent samples t-tests to estimate

the differences in mean scores on wellbeing and academic variables between academic years 2019 and 2020. Where relevant, we adjusted for unequal variance. The data and statistical code used in this study are available from the corresponding author on request.

## Results

### Descriptive results

Tables 1 and 2 present the descriptive results for the sample of bachelor and master students for each academic year. Sample participants for both courses and years were largely female and identified as Dutch. Self-reported examination scores for the bachelor course remained relatively similar across academic cohorts ( $M_{2019}=6.67$ ,  $SD_{2019}=0.61$ ;  $M_{2020}=6.64$ ,  $SD_{2020}=0.68$ ). Students in the bachelor course scored relatively moderately on the knowledge test, with an average of 5.58 questions out of 10 correct in 2019 and 5.44 out of 10 in 2020. In the master course, students also reported similar examination scores across cohorts ( $M_{2019}=7.37$ ,  $SD_{2019}=0.85$ ,  $M_{2020}=7.22$ ,  $SD_{2020}=0.70$ ). Similar to the bachelor course, students in the master course scored on average 5.41 on the knowledge test in 2019 and 4.6 in 2020.

Bivariate correlations for the bachelor and master course are presented in Tables 3 and 4. Given that international students were not present in the bachelor course during the 2020 academic year, we excluded this variable from the correlation table. The results show that low wellbeing is correlated with perceived higher workload ( $r_{\text{bachelor}}=0.48$ ,  $p < .001$ ;  $r_{\text{master}}=0.61$ ,  $p < .001$ ) and lower skills development ( $r_{\text{bachelor}}=-0.37$ ,  $p < .001$ ;  $r_{\text{master}}=-0.24$ ,  $p=.06$ ) in both the bachelor and master courses, although the relationship between workload and skills is not significant for the master course. Higher knowledge retention, as measured by scores on the knowledge test, is associated with higher scores on the examination itself for both the bachelor and master course ( $r_{\text{bachelor}}=0.31$ ,  $p < .01$ ;  $r_{\text{master}}=0.31$ ,  $p < .05$ ).

**Table 1.** Descriptive statistics for survey participants from the bachelor course.

Variable	2019 (ICE)					2020 (THE)				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Female	49	0.76	0.43	0	1	48	0.7	0.41	0	1
International student	50	0.06	0.24	0	1	–	–	–	–	–
High workload scale	50	2.66	0.63	1.3	4	48	2.81	0.91	1.13	4.38
Low wellbeing scale	49	3.07	0.83	1.8	5.4	47	3.58	0.96	2	5.8
Skills development	50	3.55	0.48	2.4	4.5	48	3.44	0.5	1.13	4
Examination grade	50	6.67	0.76	5	8.5	48	6.64	0.68	5.5	8.3
Overall grade	50	6.98	0.61	6	8.5	47	6.74	0.57	5.7	8
Knowledge test score	50	5.58	1.59	3	10	48	5.44	1.74	1	9

SD: standard deviation; M: mean; ICE: in class, closed book exam; THE: take-home open book exam

**Table 2.** Descriptive statistics for survey participants from the master course.

Variable	2019 (ICE)					2020 (THE)				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Female	34	0.88	0.33	0	1	30	0.87	0.35	0	1
International student	34	0.09	0.29	0	1	30	0.07	0.25	0	1
High workload scale	34	2.84	0.86	1	4.4	30	2.9	0.77	1	4
Low wellbeing scale	34	3.23	0.83	1.4	4.8	30	3.38	0.74	2	5.2
Skills development	34	3.75	0.44	2.9	4.5	30	3.63	0.41	3	4.75
Examination grade	34	7.37	1.07	5.5	9.6	30	7.22	0.7	6	8.6
Overall grade	34	7.33	0.85	6	9.1	30	7.47	0.54	6.5	9
Knowledge test score	34	5.41	1.73	0	9	30	4.6	1.59	1	7

SD: standard deviation; M: mean; ICE: in class, closed-book exam; THE: take-home open-book exam

**Table 3.** Bivariate correlations between demographic characteristics, academic outcomes, and wellbeing outcomes for the bachelor course.

	1	2	3	4	5	6	7	8
1 Female	1							
2 High workload	0.12	1						
3 Skills development	-0.01	-0.03	1					
4 Low well-being	0.11	0.48***	-0.37***	1				
5 Examination grade	-0.02	-0.1	0.22*	-0.05	1			
6 Overall grade	-0.05	-0.07	0.17	-0.17	0.73***	1		
7 Knowledge test score	-0.07	-0.12	0.03	0.01	0.31**	0.37***	1	
8 Academic year 2020 (THE)	0.04	0.1	-0.11	0.28**	-0.03	-0.21*	-0.04	1

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

THE: take-home open-book exam

**Table 4.** Bivariate correlations between demographic characteristics, academic outcomes, and wellbeing outcomes for the master course.

	1	2	3	4	5	6	7	8	9
1 Female	1								
2 International student	0.11	1.00							
3 High workload	0.20	0.07	1.00						
4 Skills development	0.15	0.19	-0.08	1.00					
5 Low well-being	0.07	-0.10	0.61***	-0.24	1.00				
6 Examination grade	-0.07	0.14	-0.23	0.27*	-0.11	1.00			
7 Overall grade	-0.07	0.19	-0.18	0.47***	-0.27*	0.68***	1.00		
8 Knowledge test score	0.12	-0.14	-0.25*	0.12	-0.12	0.31*	0.08	1.00	
9 Academic year 2020 (THE)	-0.02	-0.04	0.04	-0.14	0.10	-0.08	0.10	-0.24	1.00

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

THE: take-home open-book exam

### Differences between ICEs and THEs

Table 5 presents the results for the independent samples t-tests examining differences in academic and wellbeing outcomes between 2019 (ICE) and 2020 (THE) cohorts. The results show that, for the most part there were no significant differences in academic and wellbeing outcomes for both the bachelor and master courses between academic years. In the bachelor course, students reported lower wellbeing ( $M_{2019}=3.07, M_{2020}=3.58, t=-2.78, p=.01$ ) and lower overall grades ( $M_{2019}=6.98, M_{2020}=6.74, t=2.05, p=.04$ ) in the academic year 2020 compared to the

**Table 5.** Independent samples t-test results examining mean differences between academic and wellbeing outcomes.

		Bachelor course				Master course			
		N	Mean	t Value	p Value	N	Mean	t Value	p Value
High workload	ICE	50	2.66	-0.96 <sup>a</sup>	.34	34	2.84	-0.28	.78
	THE	48	2.81			30	2.90		
Low well-being	ICE	49	3.07	-2.78*	.01	34	3.23	-0.77	.45
	THE	47	3.58			30	3.38		
Skills development	ICE	50	3.55	1.10	.27	34	3.75	1.13	.26
	THE	48	3.44			30	3.63		
Examination grade	ICE	50	6.67	0.26	.79	34	7.37	0.66 <sup>a</sup>	.51
	THE	48	6.64			30	7.22		
Overall grade	ICE	53	6.98	2.05*	.04	34	7.33	-0.78	.44
	THE	48	6.74			30	7.47		
Knowledge test score	ICE	50	5.58	0.42	.67	34	5.41	1.95	.06
	THE	48	5.44			30	4.60		

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ .

ICE: in class, closed-book exam; THE: take-home open-book exam

<sup>a</sup>Adjusted for unequal variance.

**Table A1.** Ordinary least squares regression of knowledge retention and well-being on type of examination and covariates for the bachelor and master courses.

Variables	Bachelor course		Master course	
	Knowledge test score	Low well-being	Knowledge test score	Low well-being
Academic year 2020–2021 (THE)	–0.05 (0.35)	0.21* (0.15)	–0.21 (0.40)	0.06 (0.16)
Low well-being	0.10 (0.23)	–	0.07 (0.34)	–
Knowledge test score	–	0.07 (0.05)	–	0.05 (0.05)
High workload	–0.11 (0.25)	0.45*** (0.10)	–0.26 (0.33)	0.63*** (0.10)
Skills development	–0.01 (0.38)	–0.35*** (0.16)	–0.01 (0.51)	–0.2 (0.20)
Examination grade	0.30** (0.24)	0.06 (0.11)	0.25 (0.24)	0.08 (0.10)
Female	–0.06 (0.41)	0.07 (0.18)	0.18 (0.62)	–0.02 (0.25)
Constant	1.38 (2.09)	3.12** (0.89)	2.27 (2.58)	2.34* (0.98)
<i>N</i>	95	95	64	64
F-value	1.78	10.22	2.48	6.86
Adjusted <i>R</i> <sup>2</sup>	0.05	0.37	0.12	0.36

Standardised beta coefficients; Standard errors in parentheses; THE: Take home exams

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ .

previous year. However, neither examination grades nor knowledge retention significantly differed between academic years. As additional analyses, we estimated ordinary least squares regressions for the two main outcomes (knowledge retention and wellbeing) on examination type and relevant covariates. The results are substantively similar to the bivariate correlations and t-tests, whereby examination type, indicated by cohort year, is not significantly associated with differences in knowledge retention. In the academic year 2020 (THE), bachelor students (but not master students) reported significantly lower wellbeing compared to the previous year. The full results are reported in the Appendix.

## Discussion

This study aimed to shed light on to what extent ICEs compared to THEs are associated with academic performance and long-term knowledge retention, and to discern which examination type is associated with overall higher wellbeing.

No clear advantage related to academic performance or knowledge retention was found for either examination type. The examination grades were not statistically different between examination types and students' performance on the knowledge retention test was moderate, answering on average around 5–5.5 questions correctly out of 10. Only the overall course grade for the bachelor students in the THE cohort was significantly lower than that of their ICE peers, however, the overall grade includes additional course assignments. These findings are in line with other studies which do not find clear differences in performance between the examination types (Ioannidou 1997; Agarwal et al. 2008), but are contrary to most of the literature which expects that students perform better on closed-book examinations (Durning et al. 2016). Studies favouring closed-book examinations (and by implication ICE) suggest that students simply study more rigorously when preparing for a closed-book examination compared to students who tend to engage in surface learning when they expect they can consult the class materials during the examination (Moore and Jensen 2007; Agarwal and Roediger 2011). Scholars suggest that communication about the difficulty of the examination and expected learning behaviours are key to student success in a THE (Durning et al. 2016). In this study students taking the THEs

were informed at the beginning of the course that the time frame they are allocated for completing the examination and the degree of complexity of the examination questions will not allow them to thoroughly consult the materials during the examination and will require them to study well in advance. These set expectations may have encouraged students to prepare for the examination as if it were an ICE, and explain the lack of differences in student outcomes between the two examination types.

In addition, the findings suggest that there is no difference in long-term knowledge retention between students who have taken a THE or ICE. However, students who scored well on their initial examination also performed well on the retention test. This suggests that deeper learning is key to long-term knowledge retention (Beers and Bouwden 2005) which may outweigh the role of examination type. These findings do not align with the study by Rummer, Scheweppe, and Schwede (2019), which also tested a longer time lag and found that students who had closed-book examinations throughout the course performed better on a knowledge test 8 weeks later in comparison to peers who took open-book examinations throughout the course. Studies have shown that the nature and format of teaching and learning during the course can play a role in knowledge retention (Lynse and Miller 2017; Taglieri et al. 2017), meaning that these results may be accounted for by differences in the structure of the courses between studies. More frequent assessments may increase student preparation for a closed-book examination, but may decrease study behaviour in an open-book examination as students may rely on consulting the materials during the examination.

In the bachelor course, students who took a THE in 2020 reported significantly lower well-being during the course compared to students who took the ICE a year prior. This is contrary to previous findings suggesting that THEs are associated with lower student anxiety and stress compared to ICEs (Weber, McBee, and Krebs 1983; Fernald and Webster 1991; Tao and Li 2012; Dave, Dixon, and Patel 2020; Akulwar-Tajane et al. 2021). It is possible that the examination type may have been perceived as too challenging and hence stressful for bachelor students at this phase of their academic careers. Studies show that when students anticipate that the THE will be composed of high complexity questions they will experience it as highly stressful (Eilertsen and Valdermo 2000). However, this result needs to be interpreted with caution given that the bachelor and master students took the THE in the context of the Covid-19 pandemic. As a result, the courses were taught fully online (rather than on location as the previous year), and students were subject to stressors related to the Covid-19 pandemic measures which may have negatively impacted the wellbeing and study behaviours of students (Meo et al. 2020; Tam 2021). These conditions may have enhanced the stress experienced by students more so than the testing type itself. Recent studies show that the conditions related to the pandemic had a relatively greater negative effect on undergraduate and younger students' wellbeing compared to postgraduate and older students (Dodd et al. 2021). It is plausible that younger students who are less experienced with academic learning would experience more stress and impaired wellbeing in adjusting to an online learning environment than their older peers.

Importantly, some studies have raised concerns that students may be more likely to cheat during the THE (Berrett 2012). Following each examination, it was possible to identify and address plagiarism cases, however, other cheating behaviours such as collaboration and help-seeking among students, which are known concerns with THEs (Hellas, Leinonen, and Ihtantola 2017), may have been missed. This can be seen as a disadvantage of THEs but can be partially mitigated with plagiarism and collusion detection tools (Cleophas et al. 2021) and a strict time limit on executing the examination (Ng 2020; Tam 2021), as used in this study.

While this study provides insights into the differences between THEs and ICEs, there are important limitations to consider. First, the context of the Covid-19 pandemic and restrictions influenced the structure and form of learning for the 2020 cohort. Since performance and wellbeing in the 2020 cohort may in part be due to the pandemic context, we hesitate to generalise our results beyond the current context. Recent research shows that the Covid-19 pandemic has negatively influenced students' learning experiences. They find online learning more challenging,

and they feel it impairs their ability to connect with peers and teachers (Dodd et al. 2021). However, studies on examination during the Covid-19 pandemic illustrate that students hold favourable attitudes towards THEs and would like to continue having open-book assessments in the future (Chadha, Maraj, and Kogelbauer 2020). Future research should not only examine the varying impact of the Covid-19 pandemic on student learning and academic outcomes, but also replicate this study to examine differences under non-pandemic conditions.

Second, the multiple-choice knowledge retention test used in this study focused primarily on basic knowledge retention. The test focused on lower-level taxonomy skills, namely, remembering and understanding. Future studies could design a knowledge retention test that is targeted at specifically identifying differences in high- and low-level taxonomy skills over time using both ICE and THE testing forms.

Overall, this study provides insights into the extent to which examination format is associated with student academic performance and wellbeing. Our findings show that the differences between students who took the THE compared to the ICE were marginal, suggesting that there is not yet any clear advantage or disadvantage to implementing THEs. It is important to note that student performance and wellbeing are dependent on a range of individual factors, such as motivation and self-regulation, teaching and learning characteristics, such as teacher feedback, learning styles and the design of the course, and technical connectivity (Fluck, Adebayo, and Abdulhamid 2017; Tam 2021). Future research should assess to what extent these factors may interact to influence student performance and knowledge retention following different forms of assessment and examination. Finally, this study contributes to cross-national research on outcomes related to examination type (Durning et al. 2016). However, the study was conducted at a single Dutch university in the social science faculty with a limited sample size. Differences in study culture, structure of the academic year, and teaching style between countries and faculties may play a role in the link between examination type, knowledge retention and wellbeing. It would be valuable to replicate this study design in other national contexts and other faculties to reach a more generalizable conclusion.

## Ethics statement

The current research project has been approved by the Social Science Ethical Review Board of Utrecht University (20-237). Before filling out the survey participants provided informed consent. The data are available to download at the Dutch Data Archiving and Networked Services [DANS]: <https://doi.org/10.17026/dans-2xg-26aq>.

## Disclosure statement

There are no competing interests. Both authors contributed to the conceptualisation of the research project and to the writing of the manuscript.

## Funding

Funding was acquired through an internal Utrecht University grant for teaching innovation.

## Notes on contributors

*Tali Spiegel* is an Assistant Professor at Utrecht University. Her research interests include marginalised groups, stigma and health, with a particular focus on individuals with a disability and the LGBT community.

*Amy Nivette* is Associate Professor at Utrecht University. Her research interests include the determinants of institutional legitimacy and attitudes toward the law over the life course, with a particular focus on what social, experiential, and individual factors shape attitudes towards the police.

**ORCID**Tali Spiegel  <http://orcid.org/0000-0002-9932-1781>**References**

- Ackerman, R., and D. Leiser. 2014. "The Effect of Concrete Supplements on Metacognitive Regulation during Learning and Open-Book Test Taking." *British Journal of Educational Psychology* 84 (2): 329–348. doi:10.1111/bjep.12021.
- Agarwal, P. K. 2009. "Test Expectancy and Transfer of Knowledge with Open-Book and Closed-Book Tests." All theses and diss. doi:10.7936/K7Z899J0.
- Agarwal, P. K., J. D. Karpicke, S. H. K. Kang, H. L. Roediger, III., and K. B. McDermott. 2008. "Examining the Testing Effect with Open-and Closed-Book Tests." *Applied Cognitive Psychology* 22 (7): 861–876. doi:10.1002/acp.1391.
- Agarwal, P. K., and H. L. Roediger, III. 2011. "Expectancy of an Open-Book Test Decreases Performance on a Delayed Closed-Book Test." *Memory* 19 (8): 836–852. doi:10.1080/09658211.2011.613840..
- Akulwar-Tajane, I., H. Raikundlia, R. Gohil, and S. Shinde. 2021. "Academic Stress in Physiotherapy Students: Are Open Book Examinations the Solution in the Face of COVID-19 Pandemic?" *ITS* 5 (2): 1–28. doi:10.31058/j.hr.2021.52001.
- Beers, G. W., and S. Bowden. 2005. "The Effect of Teaching Method on Long-Term Knowledge Retention." *Journal of Nursing Education* 44 (11): 511–514. doi:10.3928/01484834-20051101-07.
- Bengtsson, L. 2019. "Take-Home Exams in Higher Education: A Systematic Review." *Education Sciences* 9 (4): 267. doi:10.3390/educsci9040267.
- Berrett, D. 2012. "Harvard Cheating Scandal Points out the Ambiguities of Collaboration." *Chronicle of Higher Education* 59 (3): 7. <http://www2.econ.iastate.edu/classes/econ362/hallam/NewspaperArticles/HarvardCheating.pdf>.
- Block, R. M. 2012. "A Discussion of the Effect of Open-Book and Closed-Book Exams on Student Achievement in an Introductory Statistics Course." *Primus* 22 (3): 228–238. doi:10.1080/10511970.2011.565402.
- Boniface, D. 1985. "Candidates' Use of Notes and Textbooks during an Open-Book Examination." *Educational Research* 27 (3): 201–209. doi:10.1080/0013188850270307.
- Byrne, M., and B. Flood. 2003. "Assessing the Teaching Quality of Accounting Programmes: An Evaluation of the Course Experience Questionnaire." *Assessment & Evaluation in Higher Education* 28 (2): 135–145. doi:10.1080/02602930301668.
- Chadha, D., M. Maraj, and A. Kogelbauer. 2020. "Opening up Assessment in the Age of COVID-19: Exploring the Utility of Online Open-Book Exams." *Advances in Engineering Education* 8 (4): 1–5. <https://advances.asee.org/wpcontent/uploads/Covid%2019%20Issue/Text/AEE-COVID-19-Chadha.pdf>.
- Cleophas, C., C. Hoennige, F. Meisel, and P. Meyer. 2021. Who's Cheating? Mining Patterns of Collusion from Text and Events in Online Exams. *Mining Patterns of Collusion from Text and Events in Online Exams*. doi:10.2139/ssrn.3824821.
- Dave, M., C. Dixon, and N. Patel. 2020. "Open-Book Exams." *British Dental Journal* 229 (12): 759–759. doi:10.1038/s41415-020-2512-8.
- Dodd, R. H., K. Dadaczynski, O. Okan, K. J. McCaffery, and K. Pickles. 2021. "Psychological Wellbeing and Academic Experience of University Students in Australia during COVID-19." *International Journal of Environmental Research and Public Health* 18 (3): 866. doi:10.3390/ijerph18030866.
- Durning, S. J., T. Dong, T. Ratcliffe, L. Schuwirth, A. R. Artino, J. R. Boulet, and K. Eva. 2016. "Comparing Open-Book and Closed-Book Examinations: A Systematic Review." *Academic Medicine: Journal of the Association of American Medical Colleges* 91 (4): 583–599. doi:10.1097/ACM.0000000000000977.
- Eilertsen, T. V., and O. Valdermo. 2000. "Open-Book Assessment: A Contribution to Improved Learning?" *Studies in Educational Evaluation* 26 (2): 91–103. doi:10.1016/S0191-491X(00)00010-9.
- Fernald, P. S., and S. Webster. 1991. "The Merits of the Take-Home, Closed Book Exam." *The Journal of Humanistic Education and Development* 29 (4): 130–142. doi:10.1002/j.2164-4683.1991.tb00017.x.
- Fluck, A., O. S. Adebayo, and S. M. Abdulhamid. 2017. "Secure e-Examination Systems Compared: Case Studies from Two Countries." *Journal of Information Technology Education: Innovations in Practice* 16: 107–125. doi:10.28945/3705.

- Forehand, M. 2010. "Bloom's Taxonomy." *Emerging Perspectives on Learning, Teaching, and Technology* 41 (4): 47–56. <https://www.d41.org/cms/lib/IL01904672/Centricity/Domain/422/BloomsTaxonomy.pdf>
- Hagström, L., and M. Scheja. 2014. "Using Meta-Reflection to Improve Learning and Throughput: Redesigning Assessment Procedures in a Political Science Course on Power." *Assessment & Evaluation in Higher Education* 39 (2): 242–252. doi:10.1080/02602938.2013.820822.
- Hall, D., and S. Buzwell. 2013. "The Problem of Free-Riding in Group Projects: Looking beyond Social Loafing as Reason for Non-Contribution." *Active Learning in Higher Education* 14 (1): 37–49. doi:10.1177/1469787412467123.
- Haynie, W. J.III, 2003. "Effects of Take-Home Tests and Study Questions on Retention Learning in Technology Education." *Journal of Technology Education* 14 (2): 6–18. doi:10.21061/jte.v14i2.a.1.
- Hellas, A., J. Leinonen, and P. Ihanntola. 2017. "Plagiarism in Take-Home Exams: Help-Seeking, Collaboration, and Systematic Cheating." In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education*, 238–243.
- Herzig, S., R. M. Linke, B. Marxen, U. Börner, and W. Antepohl. 2003. "Long-Term Follow up of Factual Knowledge after a Single, Randomised Problem-Based Learning Course." *BMC Medical Education* 3 (3): 3. doi:10.1186/1472-6920-3-3.
- Ioannidou, M. K. 1997. "Testing and Life-Long Learning: Open-Book and Closed-Book Examination in a University Course." *Studies in Educational Evaluation* 23 (2): 131–139. <https://eric.ed.gov/?id=EJ550018>
- Johanns, B., A. Dinkens, and J. Moore. 2017. "A Systematic Review Comparing Open-Book and Closed-Book Examinations: Evaluating Effects on Development of Critical Thinking Skills." *Nurse Education in Practice* 27: 89–94. doi:10.1016/j.nepr.2017.08.018.
- Krarp, N., N. Naeraa, and C. Olsen. 1974. "Open-Book Tests in a University Course." *Higher Education* 3 (2): 157–164. doi:10.1007/BF00143789.
- Lynse, S. J., and B. G. Miller. 2017. "A Comparison of Long-Term Knowledge Retention between Two Teaching Approaches." *Journal of College Science Teaching* 46 (6): 100–107. doi:10.2505/4/jcst17\_046\_06\_100..
- Marsh, R. 1980. "Should we Discontinue Classroom Tests? An Experimental Study." *The High School Journal* 63 (7): 288–292. <https://www.jstor.org/stable/40365604>.
- Marsh, R. 1984. "A Comparison of Take-Home versus in-Class Exams." *The Journal of Educational Research* 78 (2): 111–113. doi:10.1080/00220671.1984.10885583.
- Meo, S. A., A. A. Abukhalaf, A. A. Alomar, K. Sattar, and D. C. Klonoff. 2020. "COVID-19 Pandemic: Impact of Quarantine on Medical Students' Mental Wellbeing and Learning Behaviors." *Pakistan Journal of Medical Sciences* 36: S43–S48. doi:10.12669/pjms.36.COVID19-S4.2809.
- Moore, R., and P. H. Jensen. 2007. "Do Open-Book Exams Impede Long-Term Learning in Introductory Biology Courses?" *Journal of College Science Teaching* 36 (7): 46–49. <https://eric.ed.gov/?id=EJ769010>
- Ng, C. K. C. 2020. "Evaluation of Academic Integrity of Online Open Book Assessments Implemented in an Undergraduate Medical Radiation Science Course during COVID-19 Pandemic." *Journal of Medical Imaging and Radiation Sciences* 51 (4): 610–616. doi:10.1016/j.jmir.2020.09.009.
- Nsor-Ambala, R. 2020. "Impact of Exam Type on Exam Scores, Anxiety, and Knowledge Retention in a Cost and Management Accounting Course." *Accounting Education* 29 (1): 32–56. doi:10.1080/09639284.2019.1683871.
- Phillips, G. 2006. "Using Open-Book Tests to Strengthen the Study Skills of Community-College Biology Students." *Journal of Adolescent & Adult Literacy* 49 (7): 574–582. doi:10.1598/JAAL.49.7.3.
- Rich, J. D. 2011. "An Experimental Study of Differences in Study Habits and Long-Term Retention Rates between Take-Home and in-Class Examinations." *International Journal of University Teaching and Faculty Development* 2 (2): 121–129. <https://www.proquest.com/openview/31d0dc030106da9950e6ee4699f5379e/1?pq-origsite=gscholar&cbl=2034856>
- Roelle, J., and K. Berthold. 2017. "Effects of Incorporating Retrieval into Learning Tasks: The Complexity of the Tasks Matters." *Learning and Instruction* 49: 142–156. doi:10.1016/j.learninstruc.2017.01.008.
- Rummer, R., J. Schweppe, and A. Schwede. 2019. "Open-Book versus Closed-Book Tests in University Classes: A Field Experiment." *Frontiers in Psychology* 10: 463–471. doi:10.3389/fpsyg.2019.00463.
- Şenel, S., and H. C. C. Senel. 2021. "Use of Take-Home Exam for Remote Assessment: A Case Study from Turkey." *Journal of Educational Technology and Online Learning* 4 (2): 236–255. doi:10.31681/jetol.912965.

- Shukla, A., and R. Srivastava. 2016. "Development of Short Questionnaire to Measure an Extended Set of Role Expectation Conflict, Coworker Support and Work-Life Balance: The New Job Stress Scale." *Cogent Business & Management* 3 (1): 1–20. doi:10.1080/23311975.2015.1134034.
- Smith, S. R. 1999. "Is It Time to Close the Book on Closed-Book Examinations?" *Medicine and Health, Rhode Island* 82 (8): 285–288. <https://archive.org/details/medicinehealthrh8211rhod/page/284/mode/2up>
- Taglieri, C., D. Schnee, L. D. Camiel, K. Zaiken, A. Mistry, S. Nigro, G. Tataronis, D. Patel, S. Jacobson, and J. Goldman. 2017. "Comparison of Long-Arm Knowledge Retention in Lecture-Based versus Flipped Team-Based Learning Course Delivery." *Currents in Pharmacy Teaching and Learning* 9 (3): 391–397. doi:10.1016/j.cptl.2017.01.007.
- Tam, A. C. F. 2021. "Students' Perceptions of and Learning Practices in Online Timed Take-Home Examinations during Covid-19." *Assessment & Evaluation in Higher Education* 1–16. doi:10.1080/02602938.2021.1928599.
- Tao, J., and Z. Li. 2012. "A Case Study on Computerized Take-Home Testing: Benefits and Pitfalls." *International Journal of Technology in Teaching & Learning* 8 (1): 33–43. [https://sicet.org/main/wp-content/uploads/2016/11/ijttl-12-01-3\\_Tao.pdf](https://sicet.org/main/wp-content/uploads/2016/11/ijttl-12-01-3_Tao.pdf)
- Weber, L. J., J. K. McBee, and J. E. Krebs. 1983. "Take Home Tests: An Experimental Study." *Research in Higher Education* 18 (4): 473–483. doi:10.1007/BF00974810.
- Whitley, H. P., and J. M. Parton. 2014. "Improved Knowledge Retention among Clinical Pharmacy Students Using an Anthropology Classroom Assessment Technique." *American Journal of Pharmaceutical Education* 78 (7): 140. doi:10.5688/ajpe787140.
- World Health Organization. 1998. "Wellbeing Measures in Primary Health Care/the DEPCARE Project: Report on a WHO Meeting, Stockholm, Sweden 12–13 February 1998." In *Wellbeing Measures in Primary Health Care/the DEPCARE Project: Report on a WHO Meeting, Stockholm, Sweden 12–13 February 1998*. Geneva, Switzerland: World Health Organization.
- Zoller, U., and D. Ben-Chaim. 1989. "Interaction between Examination Type, Anxiety State, and Academic Achievement in College Science; an Action-Oriented Research." *Journal of Research in Science Teaching* 26 (1): 65–77. doi:10.1002/tea.3660260107.