Contemporary Educational Psychology 64 (2021) 101939

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

Supporting the self-regulated use of retrieval practice: A higher education classroom experiment

Marloes Broeren [a,*], Anita Heijltjes [b], Peter Verkoeijen [b,c], Guus Smeets [c], Lidia Arends [c,d]

[a] Academy of Communication and User Experience, Avans University of Applied Sciences, the Netherlands
[b] Learning and Innovation Centre, Avans University of Applied Sciences, Breda, the Netherlands
[c] Erasmus University Rotterdam, Department of Psychology, Education and Child Studies, the Netherlands
[d] Erasmus University Rotterdam, Department of Biostatistics, the Netherlands

ARTICLE INFO

Keywords:
Self-regulated learning
Retrieval practice
Cognitive learning strategies
Higher education

ABSTRACT

This study investigated whether an instructional intervention can increase students' self-regulated use of retrieval practice in a higher education classroom environment. A lab-experiment by Ariel and Karpicke (2017) revealed that strategy instructions improved students' self-regulated use of retrieval practice and subsequent test performance. Our goal was to determine whether these effects generalize to a classroom environment using key concepts from marketing communication. We compared two groups on their self-regulated use of retrieval practice using an online environment. An experimental group ($n = 58$) received strategy instructions on retrieval practice and a control group ($n = 58$) received neutral instructions. Instructions were provided during sessions 1 and 2; no instructions were provided in a third, transfer session, measuring self-regulated use of retrieval practice. In sessions 1 and 2, no significant differences between groups were found. In the transfer session, the experimental group tested themselves more (Hypothesis 1) and displayed a larger number of (correct) retrieval attempts per key concept (Hypothesis 2) than the control group. No correlations were found with performance (Hypothesis 3). With our experiment, we took a first step in supporting students in their self-regulated use of retrieval practice in a classroom environment with complex materials.

1. Introduction

A large amount of research has demonstrated that being able to self-regulate one's learning is an important characteristic of successful and effective learners (e.g., Bjork, Dunlosky, & Kornell, 2013). Self-regulated learners are active learners who set goals for themselves, attempt to monitor and regulate their cognition, motivation and behavior and who are able to adapt to the contextual features of the environment at hand (Pintrich, 2000; Zimmerman, 1990). Applying self-regulated learning strategies is found to be positively correlated with achievement (e.g., Zimmerman & Martinez-Pons, 1986; Zepeda, Richey, Ronevich, & Nokes-Malach, 2015; Chen, Chavez, Ong, & Gunderson, 2017; Fernandez & Jamet, 2017) and becomes increasingly important when student autonomy increases, i.e., in higher education environments where there is limited classroom and/or teacher support (Bjork et al., 2013; Jansen, van Leeuwen, Janssen, Jak, & Kester, 2019).

An important aspect of self-regulated learning (SRL) is the use of effective cognitive learning strategies which support longer retention of learned materials, successful retrieval of knowledge and the possibility of transferring knowledge to new situations (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Fiorella & Mayer, 2016). Retrieval practice, attempting to bring learned information to mind by deliberately recalling this information, is a cognitive learning strategy that has proven to be highly effective (Dunlosky et al., 2013; Morehead, Rhodes, & DeLozier, 2016). Students who engage in retrieval practice generally perform better on a delayed test than students who re-studied the materials. This effect on performance has been established for a large variety of participants, for various learning tasks and materials with different complexity levels, in different study contexts and on a substantial amount of outcome measures (e.g., McDaniel, Roediger, & Mcdermott, 2007; Roediger & Karpicke, 2006b; Dunlosky et al., 2013; Rowland, 2014; Karpicke & Aue, 2015).

Although retrieval practice is a highly effective learning strategy, research shows that learners do not always use it as a learning strategy (Dunlosky et al., 2013; Yan, Thai, & Bjork, 2014). During independent study, students tend to use non-optimal strategies like re-studying or

* Corresponding author at: Avans University of Applied Sciences, PO Box 90116, 4800 RA Breda, the Netherlands.
E-mail address: mmdhj.broeren@avans.nl (M. Broeren).

https://doi.org/10.1016/j.cedpsych.2020.101939

Available online 27 December 2020

underlining (Kornell & Bjork, 2007; Karpicke, Butler, & Roediger, 2009; Karpicke, 2009; Dirkx, Camp, Kester, & Kirschner, 2019). When students do practice retrieval, they use it to assess whether a certain level of learning has been reached, while they remain unaware of the learning benefits of retrieval practice and its effects on performance (Karpicke et al., 2009; Kornell & Bjork, 2007; Hartwig & Dunlosky, 2012; Tullis, Finley, & Benjamin, 2013; Kornell & Son, 2009; Karpicke, 2009; Lee & Ahn, 2017). They even believe that re-studying is a more beneficial strategy than retrieval practice (e.g., Roediger & Karpicke, 2006a; Karpicke & Roediger, 2008). Finally, when students do have accurate knowledge of effective cognitive learning strategies like retrieval practice, they often still fail to put knowledge into action. Students feel they lack the time to use a strategy like retrieval practice, they do not see the benefits, they feel uncapable of applying the strategy effectively and they may find the whole process too arduous (Foerst, Klug, Jöstl, Spiel, & Schober, 2017).

Students' unawareness of the benefits of retrieval practice and their reluctance to engage in retrieval practice as a learning strategy might be related to the fact that the act of retrieval makes learning more effortful in the short-term. Retrieving an item from memory may be hard and even unsuccessful at first (Toppino, LaVan, & Iaconelli, 2018). This desirable difficulty, which is necessary for learning to take place in the long run is not seen as advantageous by the learner himself (Bjork & Bjork, 2011; Vaughn & Kornell, 2019). More specifically, when learners are asked to predict their future performance, i.e., make a Judgment of Learning (JOL), they often predict that they will recall restudied materials better than tested materials while the reverse is true. Students base their JOLs on how difficult it is to retrieve an item - retrieval fluency - and to remember the item in the first place - encoding fluency, while these retrieval-based cues do not reflect their actual memory performance (Kornell & Son, 2009; Tullis et al., 2013). Based on these JOLs, students may end up with incorrect learning beliefs about retrieval practice, which could in turn influence the self-study choices they make.

In sum, students tend not to use retrieval practice as much as they use non-optimal strategies like re-reading. Also, students view retrieval practice as a means of assessment, not as a learning strategy, and have incorrect beliefs about the mnemonic effects of retrieval practice based on their JOLs. Therefore, for students to become aware of and appreciate the benefits of retrieval practice and apply it by themselves is unlikely to happen due to experience alone. Instead, it may require extensive additional support that gives information about the benefits of retrieval practice and directly addresses incorrect beliefs students may have. This type of theory-based support may need to be combined with an actual opportunity to practice retrieval, experiencing the strategy and its potential benefits. (Koriat & Bjork, 2006; Yan, Bjork, & Bjork, 2016)

A recent study by Ariel and Karpicke (2017) showed that a brief strategy intervention can train students to use a (repeated) retrieval practice strategy to regulate their study more effectively with a positive effect on their performance. In a laboratory experiment, the authors compared a group of students who received retrieval practice strategy instructions and a control group who experienced no strategy instructions. Both groups regulated their learning of foreign language word pairs (Lithuanian-English) using a flashcard-like computer-based learning environment. The retrieval practice strategy instructions informed students about (1) how retrieval practice maximizes retention and is more beneficial for learning than repeated studying and (2) how to use a repeated retrieval practice during learning (recalling each word-pair three times). The experiment consisted of two sessions; in the first session students received strategy instructions while in the second transfer session (one week later) no instructions were provided. The results showed that in session 1, students in the strategy instructions group practiced retrieval significantly more than the control group and were less likely to stop learning items after their initial correct retrieval; they also outperformed the control group on a cued-recall test. "Perhaps most important from an applied standpoint, students who were given retrieval practice instructions in an initial session continued to use a repeated retrieval practice strategy when learning new material in a second session one week later, without receiving any strategy instructions or reminders in the second session." (2017, p. 10). Again, the strategy instructions group also outperformed the control group on a cued recall test. These results are hopeful for educators and students as they suggest that a relatively small intervention can have large effects on students' behavior and that students can be trained to regulated their study more effectively (Ariel & Karpicke, 2017).

Inspired by Ariel and Karpicke, we were interested in finding out whether a similar intervention can have an effect on the self-regulated use of retrieval practice in an authentic education environment. A first reason to assume that this might be possible is nurtured by research that shows that large effects in the psychological laboratory replicate fairly well in the field with differences across subfields and research topics. Lab-effects from domains such as industrial-organizational psychology, personality psychology and to a lesser extent educational psychology tend to show a high correlation with field effects (Anderson, Lindsay, & Busman, 1999; Mitchell, 2012). Second, the extensive literature on retrieval practice suggests that it is a technique with a very high utility value which can be easily implemented in classroom practice (Dunlosky et al., 2013; Rowland, 2014). Third, providing students with information and instructions about an effective learning strategy and giving them the opportunity to practice with the strategy is likely to be an important factor in students actually using the strategy (McDaniel & Einstein, 2020). Finally, two other experimental studies have shown large effects on strategy use and performance produced with a relatively small intervention embedded in an authentic education context (e.g., Zepeda et al., 2015; Chen et al., 2017).

Nevertheless, there are several characteristics of an authentic educational context that may influence the effects of the strategy intervention as found by Ariel and Karpicke (2017). First of all, the learning environment is considerably less controlled than in a laboratory. A classroom intervention embedded into study seminars involves students participating in their year groups at different times during the day while other study-related activities on that same day may modify the effects of the intervention. Second, students are – to a certain extent - free to make choices about how they approach their learning, as long as their study activities are relevant for the tasks at hand; a clear contrast with a laboratory setting in which participants might expect a type of treatment and may also be aware of the fact that this treatment can induce a change (Orne, 1962). Third, students' behavior inside a classroom may be largely determined by established study habits that have developed during years of experience with formal education (Miyatsu, Nguyen, & McDaniel, 2018). Fourth, the complexity of the learning materials, i.e., key concepts instead of word pairs, may vary in authentic educational contexts. While this may require other types of learning processes (Fernandez & Jamet, 2017), as well as different types of goals and motivations when compared to the lab, the vast literature on the retrieval practice effect convincingly demonstrates that the effect applies to a broad range of (more complex) learning materials (see for example Dunlosky et al., 2013; Karpicke & Aue, 2015; Rawson, 2015). The aim of the current study was to investigate whether a strategy intervention similar to the one developed by Ariel and Karpicke (2017) can support students in the self-regulated use of retrieval practice with more complex learning materials in a higher education classroom environment.

In the present study, we attempted to increase students' self-regulated use of retrieval practice by letting them regulate their learning of key concepts from their Marketing Communication course using an online learning environment (OLE). We provided an experimental group with explicit instructions on the benefits of retrieval practice and how to use it as a learning strategy, while a control group received neutral instructions. Our goal was to compare the self-regulated study choices the two groups made in the OLE across three study sessions and to compare their performance on the final course exam. Based on Ariel and Karpicke (2017) results, we predicted that providing students with instructions on the benefits of retrieval practice

in a higher education classroom environment would lead to an increased use of retrieval practice, using testing as a study choice in the OLE (Hypothesis 1) and to an increased number of correct retrieval attempts, retrieval attempts and correct retrievals per attempt for the key concepts in the OLE (Hypothesis 2). Also, we predicted that instructions on retrieval practice were correlated to an increased performance on a final delayed test (Hypothesis 3).

2. Method

2.1. Subjects

First-year students ($N = 116$) of the Communication Studies Program of a Dutch University of Applied Sciences participated in the experiment. Our sample included 80 women and 36 men with a mean age of 19.2 years (age range: 16–25 years); participants' type of prior education consisted of 66% higher general secondary education, 25% lower vocational education and 9% other.

Participants were randomly assigned to either the experimental ($n = 58$) or the control group ($n = 58$). The experimental group consisted of 44 women and 14 men with a mean age of 19.3 years (age range: 17–25 years); type of prior education consisted of 62% higher general secondary education, 31% lower vocational education and 7% other. The control group consisted of 36 women and 22 men with a mean age of 19.2 years (age range: 16–24 years); type of prior education consisted of 69% higher general secondary education, 19% lower vocational education and 12% other. A pre-test administered in week 1 of the semester showed there were no differences between participants in prior knowledge of the key concepts used in the experiment. The experiment took place as an integral part of a first-year marketing communication course, which was one of the major courses in the semester. Informed consent was obtained from participants at the end of the first study session. Participants were presented with an informed consent statement in the online learning environment. The integration of the experiment in the curriculum was approved by both the curriculum committee and the exam committee of the department.

Considering the sample size and a priori power analysis from Ariel and Karpicke's experiment (2017) which indicated that a total number of 60 subjects (experimental $n = 30$, control $n = 30$) was required to acquire sufficient power (0.95) based on the effect size ($d = 0.87$) of a previous experiment (Grimaldi & Karpicke, 2014), we concluded that our present sample size was considerably larger than in Ariel and Karpicke (2017). For any effect size the power level of the present study would be higher. However, as the present experiment examined the effect of retrieval practice strategy instructions on study choices in a classroom environment, it remained difficult to predict the exact effect size and power level beforehand.

The actual sample sizes in our experiment resulted in $N = 98$ in Study Session 1 (experimental group, $n = 51$; control group, $n = 47$), $N = 104$ in Study Session 2 (experimental group, $n = 52$; control group, $n = 52$) and $N = 93$ in the Transfer Session (experimental group, $n = 49$; control group, $n = 44$). A small number of students left the university during the experiment. Some students did not attend all three sessions due to illness or other unforeseen circumstances. Students were only included in the analyses of the transfer session when they had taken part in both study sessions 1 *and* 2 in which the strategy instructions on the use of retrieval practice were provided. Planned statistical analyses for our experiment were pre-registered on the Open Science Framework and are part of the Online Supplemental Materials.

2.2. Materials

Sixty key concepts from the field of marketing communication were used for the experiment. All key concepts were selected by an expert panel of two marketing communication lecturers and were part of the prescribed literature of the marketing communication course. Key concepts were categorized into three levels of complexity (1 = low; 2 = medium, 3 = high) which were based on the cognitive structure of the concept, the specificity in relation to the field of marketing communication and the theoretical level of a concept. Of the 60 concepts, 25 concepts were at complexity level 1, 22 were at level 2 and 13 were at level 3. All key concepts and definitions are available upon request. As all materials are in Dutch, a number of translated examples can be found in Table A1 in Appendix A.

To test for prior knowledge, a pre-test on 54 of the 60 key concepts was administered before the start of the semester for which the same question type was used as in the study sessions. After the pre-test, 6 concepts were added due to minor last-minute adaptations to course materials. These 6 concepts were not part of the final delayed test. With 60 key concepts in total, an even number of 20 concepts was used for each study session.

The materials used for the delayed test - the final course exam - consisted of a selection of 14 key concepts (four concepts were at complexity level 1, 8 concepts were at level 2 and two concepts were at level 3). The 14 concepts were tested with open-ended questions in the exact same way as students practiced them in the study sessions (i.e., *Please provide a description of the concept* X). Students were required to provide a clear and complete description of the key concepts. The second part of the final course exam consisted of 10 multiple-choice questions that tested recognition of marketing communication concepts in specific examples and the professional context of marketing. Two sample multiple-choice questions can be found in Appendix E. A majority of the multiple-choice questions contained one or more of the key concepts presented to the students during the experiment's study sessions. The third part of the exam was a case study with four open-ended questions that focused on applying marketing communication knowledge to a real-life case. Each of these case-study questions contained one or more key concepts from the study sessions. The case study questions required high-order thinking skills as students were asked to analyze and evaluate the advertising campaign used in the case-study as well as create elements of the campaign themselves. Two sample case-study questions can be found in Appendix E.

The strategy instructions we provided to students in the experimental group were based on the written instructions from Ariel and Karpicke's experiment (2017). In their experiment, the intervention was a short text describing what retrieval practice is and how to use it accompanied by a figure reporting research on the positive effects on performance of retrieval practice that was presented to students before the learning task. Due to the fact that our experiment took place embedded in students' study seminars and that we used an online learning environment (OLE) to administer the intervention, we decided to use an instruction video instead of a short written text. The video consisted of four main elements: (a) a brief reference to the goal of the OLE, i.e., a helpful tool to learn key concepts and definitions in preparation for the course exam; (b) instructions on the self-study choices in the OLE, i.e., study, test and/or mark key concepts as done; (c) practical information regarding time, learning individually and practicing concepts as often as preferred; (d) instructions on retrieval practice as a learning strategy containing information on the active nature of retrieval practice, its relation to training one's memory, the beneficial effects on performance compared to restudy and the importance of recalling concepts three times. In the control group, elements 1 through 3 were the same as in the experimental group, but participants received neutral instructions instead of retrieval practice instructions. The neutral instructions provided information on how the OLE related to the concept of blended learning, i.e., a combination of learning during seminars and individually using the OLE and how the OLE related to the marketing communication course. The original versions of the Dutch-spoken videos are available in the Online Supplemental Materials.

2.3. Procedure

The experiment was administered during three weekly study

seminars with both the experimenter and the students' lecturer present. A schematic overview of the procedure can be found in Fig. B1 in Appendix B. Students took part in the experiment in a computer classroom in groups of equal size. Each student was assigned to a single computer and the intervention was supported by an online learning environment (OLE). The OLE was an easy-to-use website designed especially for our experiment. The interface and functionalities of the website were based on the software used in Ariel and Karpicke's experiment. As the experiment was embedded in students' seminars, it was instrumental to give students plenary instructions (8–10 min) on the intervention, which were delivered by the experimenter and lead author of this study. Students were told how the intervention would help them in their self-study and their preparations for the exam, that the intervention was part of a larger investigation at the university and that they should learn as much as possible during the sessions. Lastly, instructions were given on how to log on to the OLE, the importance of watching the instruction video before starting to learn, how long they were going to study (30 min) and restrictions regarding no talking and no use of materials like a writing pad and pen.

After the plenary instructions, students logged on to the OLE and watched an instruction video on their computer using headphones. Due to the fact that students were presented with the instruction individually on a computer, students remained unaware of the different versions for the two conditions. The instruction video in the OLE either contained instructions on the benefits of retrieval practice and how to use it (experimental group) or non-strategy related information on the importance of blended learning (control group). Instructions on retrieval practice had an informative nature and did not contain any requirements, directives or conditions. In both groups, the same video model (a university lecturer on educational psychology and c-author of this study) delivered the instructions. Both versions of the instruction video were approximately 4 min long. In total, administering the instructions (oral and video) took 15 min. The oral instructions and the instruction video were administered during the first study session in week 3 and were repeated in the second study session in week 5 of the semester. In the transfer session in week 6 of the semester, no instructions were provided. Students were merely told that today was another self-study session using the OLE, that they had to log-on to the OLE and start learning. In all sessions, students were given the opportunity to apply the instructions on retrieval practice in the online learning environment, i.e., to practice the key concepts on marketing communication for 30 min. Thus, the total intervention comprised of two sessions of 45 min and another 30-minute session over a period of four weeks. As the study sessions of our experiment were integrated in students' weekly Marketing study seminars, they were required to take no longer than 45 min in total. By using Respondus® LockDown Browser software (version 2.0.2.09), we prevented students from using other applications on the computer (once inside the LockDown Browser, trying to exit the OLE-screen would result in a warning and a lockdown of the current browser).

Students used the OLE to learn the key concepts of their marketing communication course. Per study session, the OLE contained 20 unique key concepts for which students had to make self-study choices. A key concept was presented to them in the shape of a flash card; the screen showed 20 flash cards at the same time. Please see Appendix C for print screens of the OLE. For each key concept, students had three self-study choices; they could (a) study the concept; a pop-up window appeared that allowed students to read the definition of the concept; they could (b) test themselves on the concept; a pop-up window appeared showing an open-ended question for which they could provide an answer (e.g., *Please provide a description of the concept Unique Selling Point*); they could (c) mark a concept as done; the concept disappeared from the OLE showing a green check mark. The study choices in the OLE were based on the set-up of Ariel and Karpicke (2017). Students could test themselves as often as they preferred; the system showed how many times they had tested a concept when they moved the cursor over it. After entering an answer in the text box in the pop-up window, students confirmed their answer by clicking a 'Submit'-button. After submitting their answer, students were able to compare their own answer with the correct answer by clicking on a 'Feedback'-button. The system did not automatically correct students' answers and students could skip this option if preferred.

2.4. Measures

Per participant per study session, all actions that were performed in the OLE were logged and saved to an external server. Students' mouse clicks were saved as time stamps; answers students provided to test questions were saved in a database accessible through the content management system of the OLE.

2.4.1. Self-study choices

The first outcome measure of this experiment was students' self-study choices in the OLE during the study sessions. Students' self-study choices during the transfer session were central to the goal of this study, as these would reflect the extent to which students used retrieval practice in a self-regulated way without any instructions. Self-study choices were operationalized as the mean number of *study*, *test*, *answer*, *feedback* and *done actions* per item. The difference between test and answer actions consisted of whether students chose the option 'testing' or whether they submitted an actual (partial) answer. Both test and answer actions could reflect the application of (covert) retrieval practice, while answer actions would reflect the actual retrieval (attempt). Feedback actions were the number of times students consulted the correct answer by using the feedback-button.

We compared self-study choices for the experimental group and the control group per study session by using independent-samples *t*-tests with a Bonferroni corrected alpha level of $0.05/5 = 0.01$ to correct for multiple testing. Since we expected a positive effect of instructions on study choices related to self-testing, our analyses for test and answer actions were one-tailed while for study, feedback and done actions we used two-tailed tests. If assumptions for the independent samples *t*-test were not met, we would perform a Mann-Whitney *U* test as an alternative.

2.4.2. (Correct) retrieval attempts

The second outcome measure of this experiment were students' correct retrieval attempts during the three study sessions, and more specifically during the transfer session. Per session, we compared conditions on (1) the number of *correct* retrieval attempts per item, (2) the number of retrieval attempts per item and (3) the number of correct recalls per attempted item (e.g., a person attempted 16 items in total and obtained 13 correct recalls for these 16 items). Planned analyses were independent-samples *t*-tests with a Bonferroni corrected alpha level of $0.05/3 = 0.016$ to correct for multiple testing. As we expected a positive effect of instructions on (correct) retrieval attempts, we used one-tailed hypotheses tests for these measures. If assumptions for the independent samples *t*-test were not met, we would perform a Mann-Whitney *U* test as an alternative.

2.4.3. Performance

The last outcome measure was students' performance on a final delayed test, their marketing communication course exam. Performance was measured with two types of scores: (a) *key concept scores* on a subset of 14 items practiced during study sessions which were part of the delayed test and (b) *transfer scores* for the final course exam (scores on all final delayed test items minus scores on the subset). Important to note is that the way performance was measured in our classroom experiment deviated from the study by Ariel and Karpicke in which cued recall tests were administered immediately after each study session. In our study, performance was measured two and a halve weeks after the transfer session. This delayed test - students' final exam - consisted of more than a cued recall test for the 14 selected key concepts. The results of the pre-

Table 1
Mean number and mean proportion of action types in study sessions.

Session 1				
	Experimental		Control	
Action type	*M* (*SD*)	*M* Proportions	*M* (*SD*)	*M* Proportions
Study	14.82 (4.80)	0.30	15.87 (3.96)	0.29
Test	9.55 (4.25)	0.19	10.51 (3.33)	0.19
Answer	9.06 (4.21)	0.18	10.15 (3.40)	0.19
Feedback	9.10 (4.19)	0.19	9.81 (3.66)	0.18
Done	6.45 (5.43)	0.13	8.40 (4.39)	0.15
Session 2				
	Experimental		Control	
Action type	*M* (*SD*)	*M* Proportions	*M* (*SD*)	*M* Proportions
Study	16.98 (3.77)	0.25	17.73 (3.96)	0.24
Test	13.44 (4.49)	0.20	14.52 (4.10)	0.20
Answer	12.96 (4.72)	0.19	14.12 (4.51)	0.19
Feedback	12.81 (4.58)	0.19	14.15 (4.18)	0.19
Done	10.87 (6.20)	0.16	13.67 (5.08)	0.18
Transfer Session				
	Experimental		Control	
Action type	*M* (*SD*)	*M* Proportions	*M* (*SD*)	*M* Proportions
Study	17.73 (3.13)	0.27	18.16 (2.84)	0.26
Test	12.31 (4.78)	0.19	13.46 (4.69)	0.19
Answer	11.76 (5.01)	0.18	12.89 (5.24)	0.18
Feedback	11.57 (4.89)	0.18	12.81 (5.18)	0.18
Done	11.18 (5.43)	0.17	13.52 (4.91)	0.19

Note: only descriptive analyses are provided.

Table 2
Mean and median number of study choices per item in study sessions.

Session 1				
	Experimental		Control	
Action type	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
Study	2.03 (0.70)	1.90	2.11 (0.73)	1.95
Test	2.21 (1.04)	1.90	1.92 (0.65)	1.75
Answer	1.95 (0.95)	1.69	1.67 (0.55)	1.60
Feedback	1.94 (0.94)	1.69	1.64 (0.59)	1.60
Done	0.73 (0.45)	1.00	0.91 (0.28)	1.00
Session 2				
	Experimental		Control	
Action type	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
Study	2.09 (0.71)	1.94	2.21 (0.79)	2.03
Test	2.02 (0.80)	1.83	1.84 (0.86)	1.56
Answer	1.86 (0.82)	1.65	1.62 (0.82)	1.35
Feedback	1.85 (0.83)	1.65	1.63 (0.76)	1.35
Done	0.87 (0.34)	1.00	0.94 (0.24)	1.00
Transfer Session				
	Experimental		Control	
Action type	*M (SD)*	*Mdn*	*M (SD)*	*Mdn*
Study	2.00 (0.70)	1.85	1.98 (0.61)	1.87
Test*	1.70 (0.63)	1.57	1.42 (0.48)	1.31
Answer*	1.54 (0.62)	1.40	1.28 (0.48)	1.16
Feedback*	1.55 (0.62)	1.40	1.31 (0.45)	1.16
Done	0.96 (0.20)	1.00	1.00 (0.00)	1.00

* $p < .01$.

test administered in week 1 of the semester were used as a covariate to control for the effect of prior knowledge on students' performance on the final exam.

2.4.4. Scoring of open-ended questions

To measure students' correct retrieval attempts, retrieval attempts, pre-test scores and delayed test scores, answers to open-ended questions on key concepts needed to be scored. For this, we developed a system with three possible scores (0, 1 or 2 or missing for deviant entries). *Correct* retrieval attempts were answers that received a 2-point score; answers that received a 1-point or 0-point score were marked as retrieval attempts. A score of 2 points was given when essential elements of the concept were provided (or synonyms) and when the answer contained at least 70% of the correct description of the concept. A 1-point score was given when an answer was provided or when synonyms or a description of the concept was provided but the answer contained a 40–60% correct description of the concept. An answer received a 0-point score when no answer was present, when the answer consisted of single, irrelevant words or when relevant words were given but these did not provide a coherent description of the concept. The scoring system was validated by three raters (all experienced university lecturers and researchers) who checked students' answers for 9 key concepts independently of each other. The Intra Class Correlation coefficient score of 0.62 showed a good interrater reliability between raters. The pre-test answers and the (correct) retrieval attempts were scored by one single rater using the scoring system; the answers on the delayed test were scored by three different raters (the experimenter and two marketing lecturers); inter-rater reliability was very high (over 0.90).

3. Results

3.1. Self-study choices

We predicted that providing students with instructions on the benefits of retrieval practice in an higher education classroom environment would lead to an increased use of retrieval practice, using testing as a self-study choice in the OLE (Hypothesis 1). Self-study choices were operationalized as five different types of actions in the OLE: study actions, test actions, answer actions, feedback actions and done actions.

As students self-paced their study, students did not perform actions on all 20 items presented to them. In Study Session 1, of the 20 items, the experimental group performed at least one or more actions on a mean number of 15.06 items ($SD = 4.81$) and the control group on 16.02 items ($SD = 3.94$). For Study Session 2, the experimental group performed at least one or more actions on a mean number of 17.31 items ($SD = 3.73$) and the control group on 18.06 items ($SD = 3.30$). In the Transfer Session, the experimental group performed at least one or more actions on a mean number of 18.00 items ($SD = 3.04$) and the control group on 18.59 items ($SD = 2.19$). Table 1 presents descriptive statistics for the five types of actions performed on items during the three study sessions.

Visual inspection of the data demonstrated that the distribution of scores on the dependent variable was heavily skewed due to outliers both in the experimental group and in the control group. Due to the non-normality of the distributions, independent samples Mann-Whitney U tests were used as an alternative to independent samples *t*-tests. Additionally, descriptive analyses showed that Levene's tests for equality of variances was significant for 4 out of 5 action types in study session 1 and in one action type in study session 2; in each comparison, the variance in the experimental group was significantly larger than in the control group. Pearson correlation coefficients were computed based on *z*-scores and for the interpretation of the effect sizes we use the following guidelines: effect size is low if *r* varies around 0.1, medium if *r* varies around 0.3 and large if *r* varies around 0.5 (Cohen, 1992).

The mean and median numbers for all five self-study choices per item for Study Sessions 1 and 2 can be found in Table 2. Our analyses showed that in Study Session 1, there were no significant differences between conditions on the number of study actions, $U = 1262.00$, $z = -0.45$, $p = .652$, $r = -0.05$ (2-tailed), the number of test actions $U = 1024.50$, $z = 1.24$, $p = .108$, $r = 0.13$ (1-tailed), the number of answer actions, $U = 1026.50$, $z = 1.22$, $p = .110$, $r = 0.12$ (1-tailed), the number of feedback actions, $U = 1031.00$, $z = 1.19$, $p = .233$, $r = 0.12$ (2-tailed) and the number of done actions, $U = 1425.50$, $z = -2.41$, $p = .02$, $r = -0.24$ (2-tailed).

For Study Session 2, analyses showed there were no significant differences between conditions on the number of study actions, $U = 1515.00$, $z = -1.06$, $p = .289$, r = −0.10 (2-tailed), the number of test

Table 3
Mean number of items correctly recalled and attempted and mean proportions of items correctly recalled and attempted in study sessions.

Session 1				
	Experimental		Control	
Variable	*M* (*SD*)	*M* Proportions (*SD*)	*M* (*SD*)	*M* Proportions (*SD*)
Correct	6.22 (3.13)	0.31 (0.16)	7.83 (2.62)	0.39 (0.13)
Attempted	9.06 (4.21)	0.45 (0.21)	10.15 (3.40)	0.51 (0.17)
Session 2				
	Experimental		Control	
Variable	*M* (*SD*)	*M* Proportions (*SD*)	*M* (*SD*)	*M* Proportions (*SD*)
Correct	10.60 (4.27)	0.53 (0.21)	11.88 (3.90)	0.59 (0.19)
Attempted	12.96 (4.72)	0.65 (0.24)	14.12 (4.51)	0.71 (0.23)
Transfer Session				
	Experimental		Control	
Variable	*M* (*SD*)	*M* Proportions (*SD*)	*M* (*SD*)	*M* Proportions (*SD*)
Correct	8.29 (3.68)	0.41 (0.18)	9.66 (4.12)	0.48 (0.21)
Attempted	11.78 (5.01)	0.59 (0.25)	12.89 (5.24)	0.64 (0.26)

Note: only descriptive analyses are provided.

actions, $U = 1092.00$, $z = 1.69$, $p = .046$, $r = 0.17$ (1-tailed), the number of answer actions, $U = 1099.50$, $z = 1.64$, $p = .05$, $r = 0.16$ (1-tailed), the number of feedback actions, $U = 1111.00$, $z = 1.57$, $p = .117$, $r = 0.15$ (2-tailed) and the number of done actions: $U = 1456.00$, $z = -1.32$, $p = .185$, $r = -0.13$ (2-tailed). Contrary to our expectations and the results found in the original study by Ariel and Karpicke, strategy instructions on the use of retrieval practice did not lead to students using the testing option significantly more in study sessions 1 and 2.

In the crucial Transfer Session, no instructions were given on retrieval practice and students were merely presented with the last batch of 20 key concepts in the OLE. Visual inspection of the data again showed heavily skewed distributions. Due to the non-normality of the distributions, independent samples Mann-Whitney U tests were used with a Bonferroni corrected alpha level of 0.05/5 = 0.01 as an alternative to independent samples *t*-tests. These tests indicated that there were no significant differences between conditions for the number of study actions, $U = 1103.50$, $z = 0.20$, $p = .844$, $r = 0.02$ (2-tailed). However, the experimental group displayed a significantly larger number of test actions than the control group, $U = 769.00$, $z = 2.38$, $p = .009$, $r = 0.25$ (1-tailed). The experimental group also displayed a significantly larger number of answer actions than the control group, $U = 766.00$, $z = 2.41$, $p = .008$, $r = 0.30$ (1-tailed) and a larger number of feedback actions than the control group (difference was marginally significant), $U = 760.50$, $z = 2.46$, $p = .014$, $r = 0.25$ (2-tailed). Finally, the experimental group did not differ significantly from the control group for the number of done actions: $U = 1103.50$, $z = -0.20$, $p = .844$, $r = -0.02$ (2-tailed). Mean and median scores per self-study choice are presented in Table 2.

To summarize, in the crucial transfer session, students who were presented with strategy instructions on retrieval practice in previous study sessions used testing as a study choice significantly more often than students in the control group. The strategy instructions had a moderate effect on students' use of retrieval practice as a study choice.

3.2. Correct retrieval attempts

The outcome measure correct retrieval attempts was operationalized as the number of *correct* retrieval attempts per item, the number of retrieval attempts per item and the number of correct recalls per attempted item. We predicted that providing students with instructions on the benefits of retrieval practice in an higher education classroom environment would lead to an increased number of correct retrieval attempts, retrieval attempts and correct retrievals per attempt for the key concepts in the OLE (Hypothesis 2). Table 3 presents descriptive statistics for the mean number of items (correctly) recalled during the three study sessions.

Table 4
Mean and median number of correct retrieval attempts, retrieval attempts and correct recalls per item for study sessions.

Session 1				
	Experimental		Control	
Variable	*M* (*SD*)	*Mdn*	*M* (*SD*)	*Mdn*
Correct attempts	1.75 (0.95)	1.46	1.45 (0.46)	1.30
Retrieval attempts	1.95 (0.95)	1.69	1.67 (0.55)	1.60
Correct per attempted	1.33 (0.83)	1.08	1.16 (0.51)	1.08
Session 2				
	Experimental		Control	
Variable	*M* (*SD*)	*Mdn*	*M* (*SD*)	*Mdn*
Correct attempts	1.65 (0.73)	1.43	1.40 (0.64)	1.20
Retrieval attempts	1.86 (0.82)	1.65	1.62 (0.82)	1.35
Correct per attempted	1.43 (0.83)	1.17	1.23 (0.69)	1.00
Transfer Session				
	Experimental		Control	
Variable	*M* (*SD*)	*Mdn*	*M* (*SD*)	*Mdn*
Correct attempts*	1.40 (0.61)	1.22	1.17 (0.41)	1.00
Retrieval attempts*	1.54 (0.62)	1.40	1.28 (0.48)	1.16
Correct per attempted	1.09 (0.68)	0.93	0.92 (0.47)	0.83

* $p < .01$.

Visual inspection of the data demonstrated that the distribution of scores on the dependent variable was heavily skewed due to outliers both in the experimental group and in the control group. Due to the non-normality of the distributions, independent samples Mann-Whitney U tests were used as an alternative to independent samples *t*-tests. Furthermore, in Study Session 1 and the Transfer Session, Levene's tests for equality of variances was significant for all three measures: the variance in the experimental group was significantly larger than in the control group. Mean and median scores for (correct) retrieval attempts for all sessions are presented in Table 4.

In Study Session 1, Mann-Whitney U tests indicated that there were no significant differences between conditions for the number of correct retrieval attempts per item, $U = 1001.00$, $z = 1.41$, $p = .079$, $r = 0.14$ (one-tailed), for the number of retrieval attempts per item, $U = 1026.50$, $z = 1.22$, $p = .111$, $r = 0.12$ (one-tailed) and for the number of correct recalls per attempted item, $U = 1101.50$, $z = 0.69$ $p = .245$, $r = 0.07$ (one-tailed). For Study Session 2, analyses showed that there were no significant differences between conditions for the number of correct retrieval attempts per item, $U = 1085.00$, $z = 1.75$, $p = .04$, $r = 0.17$ (one-tailed), for the number of retrieval attempts per item, $U = 1099.50$, $z = 1.644$, $p = .05$, $r = 0.16$ (one-tailed) and for the number of correct recalls per attempted item, $U = 1158.50$, $z = 1.259$, $p = .104$, $r = 0.12$ (one-tailed). Consistent with the results from the outcome measure Study Choices, strategy instructions on retrieval practice did not lead to students (correctly) recalling items more in these study sessions.

In the Transfer Session, our analyses indicated that the experimental group displayed a significantly larger number of correct retrieval attempts per item than the control group, $U = 744.00$, $z = 2.65$, $p = .004$, $r = 0.27$ (one-tailed). Also, the experimental group showed a significantly larger number of retrieval attempts per item than the control group, $U = 766.00$, $z = 2.41$, $p = .008$, $r = 0.25$ (one-tailed). There were no significant differences between conditions for the number of correct recalls per attempted item, $U = 933.00$, $z = 1.12$, $p = .132$, $r = 0.12$ (one-tailed).

In Ariel & Karpicke's experiment the effects of the instructions on students' use of a repeated retrieval practice strategy (i.e. recalling a key concept at least three times) was provided by a comparison of the mean proportions of items correctly recalled 0, 1, 2, or 3 times. Generally speaking, in our experiment neither the experimental nor the control group correctly recalled the majority of the items presented to them in

the OLE once or more than once. With an exploratory analysis, we took a closer look at the proportions of items recalled, once, twice or three times or more and the median number of items correctly recalled once, twice or three or more as part of the total number of attempted items.

In Study Sessions 1 and 2, the median proportions of items correctly recalled once was higher in the control condition than in the experimental condition (Stuy Session 1, $U = 1541.00$, $p = .01$; session 2, $U = 1878.00$, $p = .03$). For the median number of items correctly recalled once as part of the total number of attempted items, the control condition displays a significantly higher number in all sessions (Study Session 1, $U = 1484.00$, $p = .04$; Study Session 2, $U = 1850.00$, $p = .04$; Transfer Session, $U = 1836.50$, $p = .03$).

In the Transfer Session, the median proportion of items correctly recalled once is the same for both conditions. The median proportion correctly recalled twice was higher in the experimental condition than in the control condition ($U = 1168.00$, $p = .04$). The same goes for the median proportion correctly recalled three times or more ($U = 1248.50$, $p = .04$). We see the same pattern for the median number of items correctly recalled twice as part of the total number of attempted items ($U = 1111.50$, $p = .02$) and the median number of items recalled three times or more as part of the total number of attempted items ($U = 1236.50$, $p = .04$). Instructions provided to students in Study Sessions 1 and 2 stated that for retrieval practice to be effective, all items should be correctly recalled at least three times. Looking at the pattern in the transfer session, this might indicate the experimental groups' willingness to adhere to the instructions.

3.3. Performance

Not all students who participated in the study sessions took part in the final course exam, our final delayed test (due to illness or other unforeseen circumstances). This led to a sample size of $N = 96$ (experimental group, $n = 50$; control group, $n = 46$) which deviated from sample sizes in the study sessions. Performance was measured by comparing the experimental group with the control group on (a) *key concept scores*, i.e., mean proportion scores for the subset of 14 items of the final course exam which were practiced during the study sessions and (b) *transfer scores*, i.e., mean proportion scores for the final course exam (scores on all final course exam items minus scores of the subset of 14 items). To assess performance progress for the group as a whole, mean proportion key concept scores for the pre-test and mean proportion key concept scores for the final exam (delayed test) were compared. We predicted that providing students with instructions on the benefits of retrieval practice in an higher education classroom environment would be correlated to an increased performance on the final exam (delayed test) (Hypothesis 3).

To compare the two types of mean proportion scores we planned two analyses of covariance (ANCOVA) while controlling for the effect of pre-test score by using the mean pre-test proportion scores as a covariate. Analyses however showed that the pre-test key concept scores (mean proportion score for subset of 14 items) were not significantly correlated to the delayed test key concept scores, $r = 0.11$, $p = .231$. The same goes for the correlation between the pre-test transfer scores (mean proportion scores for all items on pre-test minus subset of 14) and the delayed test transfer scores (mean proportion scores for all items on delayed test minus subset of 14), $r = 0.11$, $p = .231$. Due to the lack of a correlation we decided to perform a one-tailed independent samples *t*-test to compare the conditions on the two mean final delayed test scores using a Bonferroni corrected alpha level of $0.05 / 2 = 0.025$. Mean proportion key concept scores were 0.48 ($SD = 0.22$) for the experimental group and 0.45 ($SD = 0.20$) for the control group. Our analyses showed that there were no significant differences between conditions, $t(94) = 0.74$, $p = .231$, $d = 0.15$. Mean proportion transfer scores were 0.55 ($SD = 0.20$) for the experimental group and 0.53 ($SD = 0.17$) for the control group. Again, there were no significant differences between conditions, $t(94) = 0.68$, $p = .498$, $d = 0.14$. In addition, to assess performance progress for the group as a whole, a paired samples *t*-test was used to compare mean proportion key concept scores. The mean proportion key concept score on the delayed test was 0.47 ($SD = 0.21$) and the mean proportion key concept score on the pre-test was 0.07 ($SD = 0.06$). The key concept scores on the delayed test were significantly larger than on the pre-test, $t(94) = 18.50$, $p < .001$, $d = 2.71$.

Based on our hypotheses, we expected that retrieval practice instructions would lead to students using testing as a study choice more, which would in turn be positively correlated with performance on the final exam (delayed test). As the data showed, a considerable part of the items presented to students during the study sessions was not (correctly) recalled. To check for possible relations between test actions, answer actions, correct recall attempts and recall attempts for the 14 items and their delayed test sub scores, Pearson's *r* correlation coefficients were computed (see Table D1 in Appendix D). As the data shows, the majority of correlations vary around $r = 0.1$ and $r = 0.2$ while only a small number of correlations vary around $r = 0.3$. In sum, there were no strong correlations between the number of times that students tested themselves on the 14 items during study sessions and their sub scores for these items on the delayed test. Important to note is that these additional analyses have an exploratory nature; reported *p*-values are not interpreted in relation to our hypotheses.

A second exploratory analysis was conducted on a possible correlation between use of retrieval practice and final grade for all participants *regardless* of condition. In session 3, we found a correlation between the performance on the final course exam, i.e., the final grade, and the total number of recall attempts ($r = 0.23$, $p = .02$), the total number of correct recalls ($r = 0.21$, $p = .03$) and the mean number of correct recalls per item ($r = 0.20$, $p = .04$). Regardless of condition, students' number of (correct) recall attempts and correct recalls per item in session 3 positively correlated with their final course grade. The modest application of retrieval practice in our authentic setting, especially when compared to the large effects found in Ariel and Karpicke's study, does seem to be positively related to performance on a delayed test.

4. Discussion

The current experiment investigated the effects of strategy instructions on the self-regulated use of retrieval practice in a higher education classroom environment. An experimental group received instructions on the benefits of retrieval practice and how to use it; a control group received neutral instructions. Both the experimental and the control group used retrieval practice: in study sessions 1 and 2, the experimental group used retrieval practice more than the control group, but differences were not statistically significant. During the transfer session, students in the experimental group significantly tested themselves more than the control group (Hypothesis 1) and displayed a significantly larger number of (correct) retrieval attempts per key concept than the control group (Hypothesis 2). No correlations were found between the use of retrieval practice and performance (Hypothesis 3). We did not find differences in final performance between conditions based on increased retrieval practice use.

In our experiment, the strategy intervention aimed at increasing the self-regulated use of retrieval practice had a moderate effect as compared to the effect found by Ariel and Karpicke (2017). Furthermore, in our study the intervention group only outperformed the control group on retrieval practice in the transfer session, whereas Ariel and Karpicke found a consistent retrieval practice advantage of the intervention group in all sessions of their experiment. This discrepancy between our results and those of Ariel and Karpicke may be due to the circumstances under which the experiments took place. Our intervention was part of students' actual study sessions in which we presented instructions to them on how they *could* approach their learning. This classroom setting differed considerably from the laboratory-setting in Ariel and Karpicke's experiment. Here, participants' behavior may not only have been influenced through the experimental variable(s), but also through often clearly perceived demand characteristics (Orne,

1962; Nichols & Maner, 2008). Hence, participants in the intervention condition in Ariel and Karpicke's experiment might have interpreted the information in the retrieval practice video as a literal instruction, that is, as a behavioral demand they had to meet in all sessions. This would explain why Ariel and Karpicke found a retrieval practice advantage in all of three sessions in their experiment.

By contrast, due to the real life settings, our participants may have interpreted the information in the retrieval practice video more as an advice rather than a strict behavioral demand. After all, in real- life practice, students in higher education are allowed to select their own strategies for independent study and have strong, ingrained habits when it comes to study strategies (Miyatsu et al., 2018). Furthermore, our students were unaware of the manipulation involved and did not expect to receive a treatment. Hence, students may have behaved as they would normally have done in a real-life classroom setting, with their behavior being guided by pragmatics, such as upcoming exams or assignments rather than systematic decision making (e.g., Hartwig & Dunlosky, 2012; Kornell & Bjork, 2007; Rovers, Stalmeijer, van Meriënboer, Savelberg, & de Bruin, 2018). An important pragmatic factor was the fact that of the three sessions in our study, the transfer session was closest to the exam, i.e., two-and-a-halve weeks prior to the exam. As the retrieval practice instructions explained that it would lead to better performance on tests than re-study, this may have motivated students in the experimental group to use retrieval practice particularly in this session. So, the increased relevance of retrieval practice for the final exam might explain why we only found a benefit of the retrieval practice intervention in the transfer session.

In the current study, we did not find differences in final performance between conditions based on increased retrieval practice use, while Ariel and Karpicke (2017) found a moderate effect on performance in final cued recall tests during the transfer session. The absence of an effect on performance can be explained by the different nature of the performance tests used. In our study, performance was not measured by final cued recall tests immediately after the study sessions, but by a final course exam measuring key concept and transfer scores. In this exam, a modest part of 14 items was dedicated to cued recall questions which were taken directly from the study sessions (e.g., *Please describe the concept 'customer journey'*). In addition, the experimental group's retrieval practice use in the transfer session, although significantly higher than the control group, was much smaller than in the study by Ariel and Karpicke (2017). As stated above, this might have been due to the fact that participants in the laboratory study of Ariel and Karpicke may have taken the instruction as a strong demand characteristic, whereas our participants may have taken it more as an advice they could or could not follow. In any case, the small differences in the use of retrieval practice between conditions may have contributed to a lack of an effect on the delayed test.

One could argue that providing students with instructions and the opportunity to practice retrieval or restudy materials during study sessions *at the university* is only partly self-regulated. A completely self-regulated study session would be planned, executed and monitored by the student himself. While integrating the strategy intervention into students' study seminars maintained a certain degree of teacher control, within the study sessions students were required to self-regulate their learning. They were told to learn as much as possible during the session and could make their own self-study choices. They were not required to meet a certain criterion level while practicing retrieval. Moreover, they could decide not to practice retrieval and merely restudy items instead or to mark all concepts as done. In addition, during the crucial transfer session no instructions were provided and students in the experimental group practiced retrieval significantly more than those in the control group.

Based on Ariel and Karpicke's results (2017), the strategy instructions we provided in our classroom experiment were aimed at increasing students' knowledge of retrieval practice, which was expected to enable the implementation of retrieval practice as a learning strategy during the study sessions. Considering the more moderate effect we found, successfully supporting students in the self-regulated use of retrieval practice in a classroom environment may require more than addressing students' (lack of) knowledge (McDaniel & Einstein, 2020). Additional support might need to focus on students' metacognitive monitoring during learning, as research tells us that self-monitoring is a difficult skill for students without any support (Baars, Vink, van Gog, de Bruin, & Paas, 2014). Further research could explore how integrating metacognitive support into a retrieval practice intervention may help students in correctly monitoring their progress by providing feedback on their use of retrieval practice. It could also entail support in selecting retrieval practice as a relevant learning strategy, planning its actual use and applying it correctly. Future research may also need to explore the (potentially incorrect) learning beliefs students have and how these influence their use of retrieval practice. Students often have very strong a-priori beliefs about which strategies are effective, which may influence their willingness to use strategies (Yan et al., 2016). If we can gain insight into these beliefs in relation to their current study behavior, we might be able to find out what can motivate them to optimize their strategy use (Miyatsu et al., 2018).

Our study showed that a strategy intervention similar to the one used in the lab-experiment by Ariel and Karpicke (2017) can have a partial effect on the self-regulated use of retrieval practice in a higher education classroom environment. While students tended to use retrieval practice in a more modest way than in Ariel and Karpicke's lab-study, the repeated exposure to instructions and the upcoming course exam may have motivated the continued use of retrieval practice by the experimental group in the transfer session. With our results, we have emphasized the importance of shifting learning strategy interventions from the lab to the classroom and have contributed to the growing body of research on how to successfully implement effective learning strategies like retrieval practice in authentic educational settings.

Appendix A

See Table A1

Table A1
Examples of key concepts and definitions.

Key concept	Definition	Complexity level (1 = low, 2 = medium, 3 = high)
unique selling point	The unique qualities of a product or service with which it can be distinguished from other, similar products or services in the product area.	1
brand awareness	The active familiarity with a brand name or the passive recognition of a brand name from a series of brand names.	1
buying intention	The consumer's intention to purchase the brand, product or service.	2
brand persona	The archetypical consumer that represents the target audience with relevant characteristics that belong to the target audience.	2
mere exposure effect	A repeated exposure to a stimulus that creates a positive attitude towards this stimulus.	3
subjective norm	The beliefs of a person about how relevant, important others perceive his behavior towards a matter.	3

Appendix B

See Fig. B1

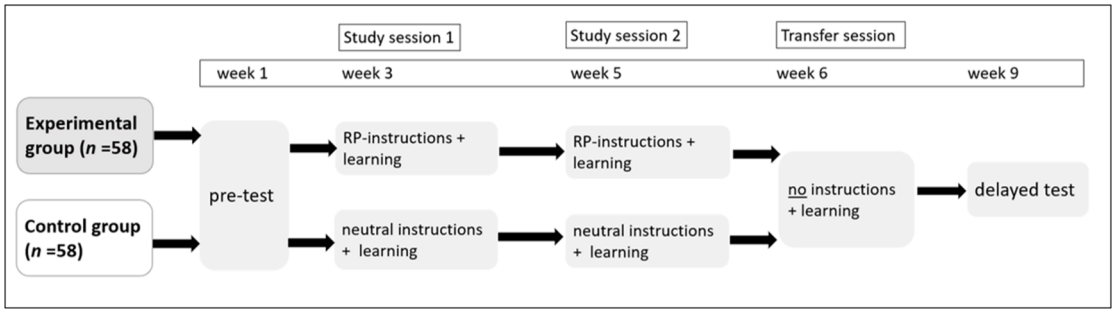

Fig. B1. Schematic overview of the experiment's procedure.

Appendix C

See Figs. C1-C3

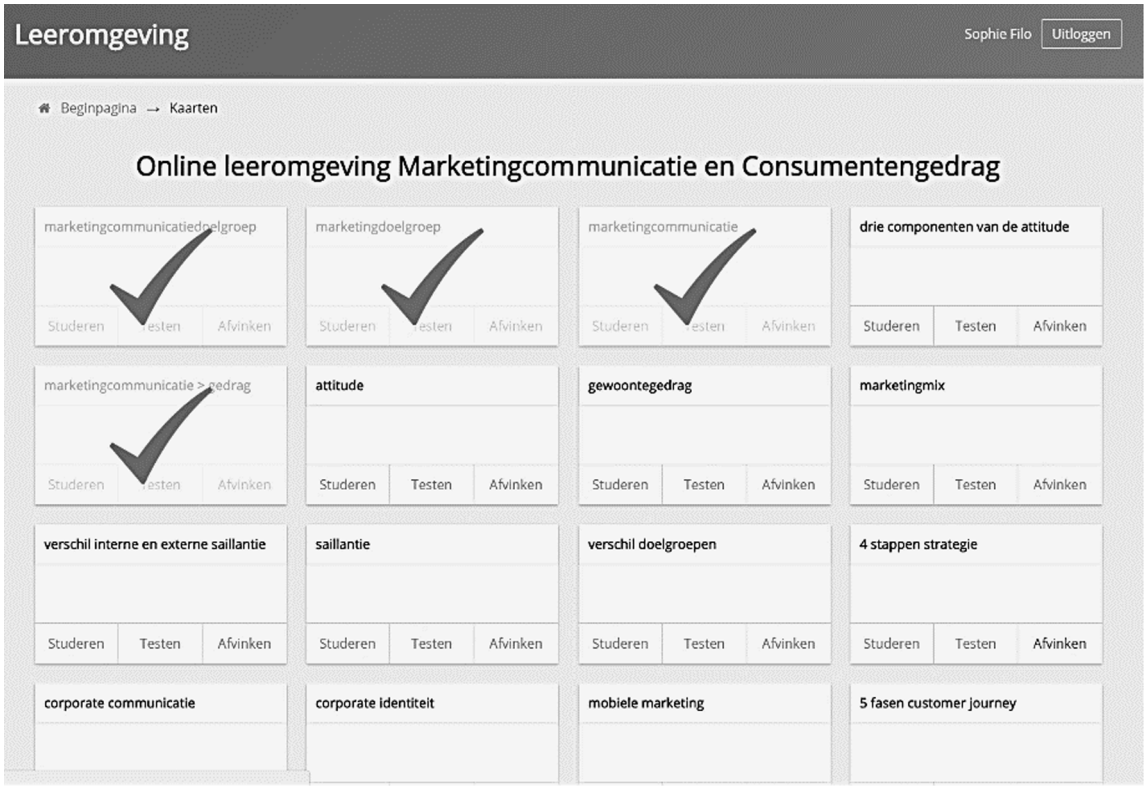

Fig. C1. Print screen of flashcard-like presentation in online learning environment of key concepts of study session 1. Key concepts are in bold (e.g., 'attitude'). Study-choices presented below the key concept are Study ('Studeren'), Test ('Testen') and Mark as done ('Afvinken'). The green check mark indicates concepts that have been marked as done.

Fig. C2. Print screen of pop-up window for testing option in online learning environment for key concept *habitual behavior* ('gewoontegedrag' in Dutch). The question ('Vraag') *Please describe the key concept habitual behavior* ('Beschrijf de term gewoontegedrag.') is presented for which an answer can provided in the text below ('Jouw antwoord'). After entering an answer, students confirmed their answer ('Bevestig je antwoord') and could either compare it to the correct answer - the feedback option ('Bekijk juiste antwoord') or close the window ('Sluiten').

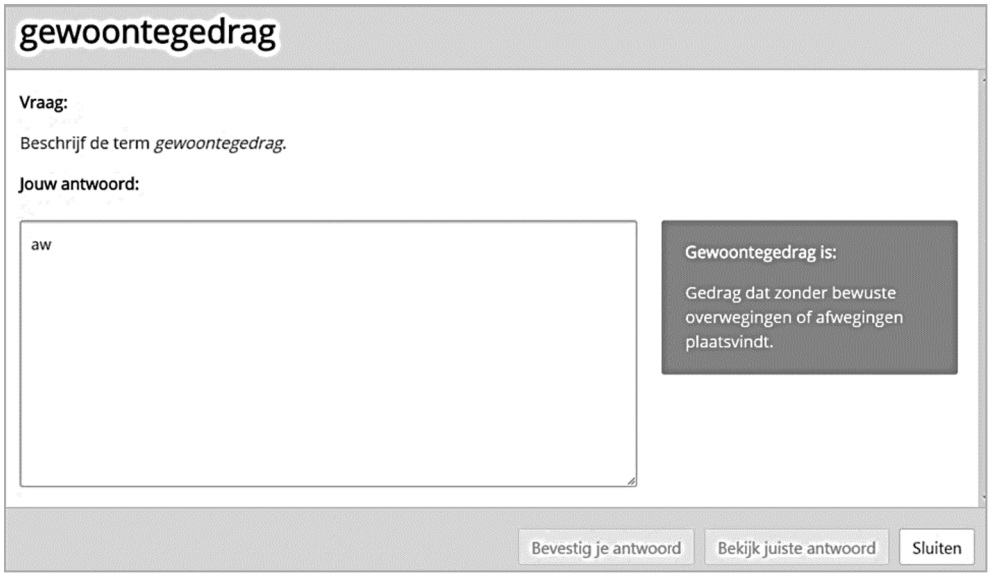

Fig. C3. Print screen of pop-up window for testing *with feedback option* in online learning environment for key concept *habitual behavior* ('gewoontegedrag'). The question ('Vraag') *Please describe the key concept habitual behavior* ('Beschrijf de term gewoontegdrag.') is presented for which an answer can provided in the text below ('Jouw antwoord'). On the right side, the correct answer is provided after clicking 'check correct answer ('Bekijk juiste antwoord').

Appendix D

See Table D1

Table D1
Pearson correlations for items of delayed test and use of retrieval practice during study sessions.

	Test actions		Answer actions		Correct retrieval attempts		Retrieval attemtps	
Test items	Exp	Contr	Exp	Contr	Exp	Contr	Exp	Contr
Item 04	0.265	−0.267	0.239	−0.132	,128	,109	,239	−,132
Item 58	0.193	0.307*	0.222	0.184	,282*	,119	,222	,184
Item 61	0.126	0.074	0.069	0.083	,111	,004	,069	,083
Item 63	0.025	−0.175	0.009	−0.104	−,027	−,019	,000	−,051
Item 74	0.186	0.094	0.204	0.086	,084	,051	,204	,086
Item 75	−0.106	0.173	0.008	0.117	,021	,187	,008	,117
Item 76	−0.010	−0.220	0.035	−0.201	,107	−,113	,082	−,198
Item 77	0.242	0.091	0.149	0.087	,185	,048	,149	,087
Item 78	0.103	0.055	0.069	0.121	,118	,132	,094	,083

(*continued on next page*)

Table D1 (*continued*)

	Test actions		Answer actions		Correct retrieval attempts		Retrieval attemtps	
Test items	Exp	Contr	Exp	Contr	Exp	Contr	Exp	Contr
Item 79	0.112	0.240	0.101	0.270	,060	,241	,117	,238
Item 82	0.268	0.256	0.261	0.318*	,321*	,227	,272*	,315*
Item 89	0.284	0.295	0.155	0.257	,258	,225	,139	,275*
Item 99	−0.088	−0.112	−0.068	−0.134	−,229	−,202	−,072	−,125
Item 102	−0.197	−0.164	−0.217	−0.144	−,189	−,157	−,225	−,128

* $p < .05$.

Appendix E

Sample multiple-choice question 1: The media have reported several messages on the successful night sales of the Apple iPhone 7 (sales started at midnight of the first sales day). Which communication instrument was used by Apple in this example?

A. Sales promotion;
B. Action advertisement;
C. Personal sales;
D. Marketing PR.

Sample multiple-choice question 2: Why is having a strong brand important for *consumers?*

A. Product launches of strong brands are easier;
B. Strong brands are often better and more expensive;
C. With a strong brand, a consumer can show his/her personality;
D. A strong brand ensures continuity.

Sample case study question 1: *The brand Milka has launched an international advertising campaign around the central theme "Dare to be tender". To create local success for this campaign they developed a Facebook application that consumers can use to send compliments to each other.* Please identify which positioning strategy Milka uses in this campaign and explain why you think this strategy is relevant.

Sample case study question 2: *The brand Milka has launched an international advertising campaign around the central theme "Dare to be tender". To create local success for this campaign they developed a Facebook application that consumers can use to send compliments to each other.* Please create the SMART communication objective for this Milka campaign which was aimed at influencing the brand attitude of the target group.

References

Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science, 8*(1), 3–9. https://doi.org/10.1111/1467-8721.00002.

Ariel, R., & Karpicke, J. D. (2017). Improving self-regulated learning with a retrieval practice intervention. *Journal of Experimental Psychology: Applied.* https://doi.org/10.1037/xap0000133.

Baars, M., Vink, S., van Gog, T., de Bruin, A., & Paas, F. (2014). Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem solving. *Learning and Instruction, 33*, 92–107. https://doi.org/10.1016/j.learninstruc.2014.04.004.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (2nd ed., pp. 56–64). New York: Worth.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*(1), 417–444. https://doi.org/10.1146/annurev-psych-113011-143823.

Chen, P., Chavez, O., Ong, D., & Gunderson, B. (2017). Strategic resource use for learning: A self-administered intervention that guides self-reflection on effective resource use enhances academic performance. *Psychological Science, 28*(6), 774–785. https://doi.org/10.1177/0956797617696456.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159. https://doi.org/10.1037/0033-2909.112.1.155.

Dirkx, K. J. H., Camp, G., Kester, L., & Kirschner, P. A. (2019). Do secondary school students make use of effective study strategies when they study on their own? *Applied Cognitive Psychology, 33*, 952–957. https://doi.org/10.1002/acp.3584.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques. *Psychological Science in the Public Interest, 14*(1), 4–58. https://doi.org/10.1177/1529100612453266.

Fernandez, J., & Jamet, E. (2017). Extending the testing effect to self-regulated learning. *Metacognition and Learning, 12*(2), 131–156. https://doi.org/10.1007/s11409-016-9163-9.

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9.

Foerst, N. M., Klug, J., Jöstl, G., Spiel, C., & Schober, B. (2017). Knowledge vs. action: Discrepancies in university students' knowledge about and self-reported use of self-regulated learning strategies. *Frontiers in Psychology, 8*, 1288. https://doi.org/10.3389/fpsyg.2017.01288.

Grimaldi, P. J., & Karpicke, J. D. (2014). Guided retrieval practice of educational materials using automated scoring. *Journal of Educational Psychology, 106*, 58–68. https://doi.org/10.1037/a0033208.

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review, 19*, 126–134. https://doi.org/10.3758/s13423-011-0181-y.

Jansen, R. S., van Leeuwen, A., Janssen, J., Jak, S., & Kester, L. (2019). Self-regulated learning partially mediates the effect of self-regulated learning interventions on achievement in higher education: A meta-analysis. *Educational Research Review, 28* (September), Article 100292. https://doi.org/10.1016/j.edurev.2019.100292.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138*(4), 469–486. https://doi.org/10.1037/a0017341.

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review, 27*(2), 317–326. https://doi.org/10.1007/s10648-015-9309-3.

Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory, 17*(4), 471–479. https://doi.org/10.1080/09658210802647009.

Karpicke, J. D., & Roediger, H. L., III (2008). The critical importance of retrieval for learning. *Science, 319*, 966–968. https://doi.org/10.1126/science.1152408.

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1133–1145.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*(2), 219–224.

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory, 17*, 493–501. https://doi.org/10.1080/09658210902832915.

Lee, H. S., & Ahn, D. (2017). Testing prepares students to learn better: The forward effect of testing in category learning. *Journal of Educational Psychology, 110*(2), 203–217. https://doi.org/10.1037/edu0000211.

McDaniel, M. A., & Einstein, G. O. (2020). Training learning strategies to promote self-regulation and transfer: The knowledge, belief, commitment, and planning framework. *Perspectives on Psychological Science, 15*(6), 1363–1381. https://doi.org/10.1177/1745691620920723.

McDaniel, M. A., Roediger, H. L., & Mcdermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14*(2), 200–206. https://doi.org/10.3758/BF03194052.

Mitchell, G. (2012). Revisiting truth or triviality: The external validity of research in the psychological laboratory. *Perspectives on Psychological Science, 7*(2), 109–117. https://doi.org/10.1177/1745691611432343.

Miyatsu, T., Nguyen, K., & McDaniel, M. A. (2018). Five popular study strategies: Their pitfalls and optimal implementations. *Perspectives on Psychological Science, 13*(3), 390–407. https://doi.org/10.1177/1745691617710510.

Morehead, K., Rhodes, M. G., & DeLozier, S. (2016). Instructor and student knowledge of study strategies. *Memory, 24*(2), 257–271. https://doi.org/10.1080/09658211.2014.1001992.

Nichols, A. L., & Maner, J. K. (2008). The good-subject effect: Investigating participant demand characteristics. *Journal of General Psychology, 135*(2), 151–166. https://doi.org/10.3200/GENP.135.2.151-166.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*(11), 776–783. https://doi.org/10.1037/h0043424.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-012109890-2/50043-3.

Roediger, H. L., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x.

Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x.

Rovers, S., Stalmeijer, R. E., van Merriënboer, J., Savelberg, H., & de Bruin, A. (2018). How and why do students use learning strategies? A mixed methods study on learning strategies and desirable difficulties with effective strategy users. *Frontiers in Psychology, 9*, 2501. https://doi.org/10.3389/fpsyg.2018.02501.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432–1463. https://doi.org/10.1037/a0037559.

Rawson, K. A. (2015). The status of the testing effect for complex materials: Still a winner. *Educational Psychology Review, 27*, 327–331. https://doi.org/10.1007/s10648-015-9308-4.

Toppino, T. C., LaVan, M. H., & Iaconelli, R. T. (2018). Metacognitive control in self-regulated learning: Conditions affecting the choice of restudying versus retrieval practice. *Memory and Cognition, 46*(7), 1164–1177. https://doi.org/10.3758/s13421-018-0828-2.

Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition, 41*(3), 429–442. https://doi.org/10.3758/s13421-012-0274-5.

Vaughn, K. E., & Kornell, N. (2019). How to activate students' natural desire to test themselves. *Cognitive Research: Principles and Implications, 4*(1). https://doi.org/10.1186/s41235-019-0187-y.

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General, 145*(7), 918–933. https://doi.org/10.1037/xge0000177.

Yan, V. X., Thai, K., & Bjork, R. A. (2014). Habits and beliefs that guide self-regulated learning: Do they vary with mindset? *Journal of Applied Research in Memory and Cognition, 3*(3), 140–152. https://doi.org/10.1016/j.jarmac.2014.04.003.

Zepeda, C. D., Richey, J. E., Ronevich, P., & Nokes-Malach, T. J. (2015). Direct instruction of metacognition benefits adolescent science learning, transfer, and motivation: An in vivo study. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000022.

Zimmerman, B., & Martinez Pons, M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal, 23*(4), 614–628.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*, 3–17. https://doi.org/10.1207/s15326985ep2501_2.