MASTER'S THESIS

Gespreide Versus Aaneengesloten Oefening in het Economisch Domein

Sleven, Pauline

Award date:
2018

Gespreide Versus Aaneengesloten Oefening in het Economisch Domein

Spaced Versus Massed Practice in the Economics Domain

Pauline Sleven MA

Master Onderwijswetenschappen

Open Universiteit

Datum: 25-6-2018

Begeleiding: Dr. G. Camp

Table of Contents

Spaced Versus Massed Practice in the Economics Domain

Pauline Sleven MA

Summary

Research into the spacing of study goes back over 100 years. The spacing effect, the enhancement in long-term retention of information caused by spreading restudy sessions over time rather than studying material consecutively, has turned out to be a solid finding in many different memory tasks. However, the effect has been mainly examined in laboratory settings. In the last decade, researchers have begun to realise that even though laboratory findings have high internal validity, the circumstances in these experiments do not much resemble the practices in classrooms or during students' self-study activities at home. As a result, there has been a rise in research conducted in actual classrooms, with educationally relevant materials, using educationally valid time frames.

The aim of this research was to add to this search for more ecologically valid evidence to support the spacing effect and to generate highly practical guidelines on how to apply spacing, for educators and students, thereby increasing the chance of both parties incorporating them. The central question was whether spaced study of educationally relevant material, in a real classroom setting of first year Bachelor students in the economics domain, yields better results on long term retention than does massed study. Based on the evidence for the spacing effect so far, the prediction was that spaced study indeed increases long-term retention.

The design of this study resembled a classic spacing effect study. It consisted of three phases; a study phase in the form of a lecture, followed by a practice phase in the form of either three consecutive restudy sessions of 20 minutes (massed condition) or three restudy sessions of 20 minutes spaced at least 24 hours apart (spaced condition), and a test phase, in the form of two recall tests. Fourty-four first year Bachelor students, enrolled in a course in the economics domain at a university of applied sciences in The Netherlands, participated in the research. Their mean age was 19.6 (range, 17 – 24 years), 45% were female and 41% were Dutch. During the restudy sessions, 18 flashcards with key concept definitions taken from their economics course book were studied. To measure immediate retention (after three restudy sessions) and long term retention (one week after the last restudy session) of the material, a recall test was designed.

Results showed that, within the subject groups, retention in the immediate test was higher than retention in the delayed test. Between the subject groups, there was only a numeric difference in forgetting rate. A marginally significant interaction-effect between retention interval and condition was found, showing that the forgetting curve of the massed condition was steeper than that of the

spaced condition. Long term retention in the spaced condition was around 6% higher than in the massed condition, however, this difference was not significant.

The results of the present research have not confirmed that spaced study indeed increases long-term retention. Despite the limitations of the study, caused by its lifelike nature, the numeric difference in the score on long-term retention, though not significant, points to the possibility that the spacing effect does exist in a naturalistic setting.

Keywords: spacing effect, testing effect, distributed practice, practice testing, flashcards, higher education

Spaced Versus Massed Practice in the Economics Domain

1. Introduction

What do exercising, healthy eating, not smoking and not drinking have in common? We know it is better for us on the long term but cannot really get ourselves to do it. Spacing study sessions over time can be added to this list. There is massive scientific evidence (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Donovan & Radosevich, 1999) for the spacing effect, i.e. spreading restudy sessions over time rather than studying material consecutively enhances long-term retention of the information. However, educators and students alike do not seem to be all that enthusiastic to incorporate spacing in their daily study and teaching routines (Dempster, 1988; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Son & Simon, 2012).

Students at different levels of education are asked to perform many types of tasks at school every day. They might be asked to: 1) write a summary report of an event, 2) design a market strategy for their product using a known strategy as a model, 3) review a work of art in terms of form, colour and texture, 4) write a TV show, play, puppet show, role play, song or pantomime about a particular topic, 5) conduct a debate about an issue of special interest (Dalton, Smith, & Administration, 1989). Students can only execute these tasks effectively and efficiently if they have formed a foundational knowledge base they can easily access from memory (Kang, 2016). As higher order learning and reasoning depend on this basic knowledge it is clear why teachers aim for long-term retention of the materials they teach. Students, on the other hand, often care more about short term retention of material; knowing enough to pass the test.

Despite the fact that students seem to be aware of the benefits of spacing (Susser & McCabe, 2013), there are several reasons why they do not engage in it on a regular basis. Studying material in shorter, distributed sessions takes more effort, planning wise and discipline wise. It also gives students the impression that learning actually goes slower because they are confronted with forgetting, which in turn affects their confidence in learning (Gerbier & Toppino, 2015; Son & Simon, 2012). Generally, students don't understand that massed practice (for example cramming before an exam) gives a false sense of familiarity with to-be-studied material (Kang, 2016).

When it comes to educators, research suggests that their knowledge about evidence-based approaches to learning is not much more profound than that of their students (Morehead, Rhodes, & DeLozier, 2016). Teachers often use teaching methods they themselves were taught by, or rely on their intuition (Kang, 2016). Moreover, practical issues stand in the way of teachers wanting to apply

spacing; time pressure to cover a huge curriculum and teaching materials that are designed in a modular fashion (Dunlosky et al., 2013).

An overarching reason why spacing activities are not frequently used may be because educators and students have not (yet) fully experienced the practical applications of the scientific findings in their daily routines. Many of the current policy implications, such as spacing study sessions, reviewing previous material during class, offering distributed homework sets and cumulative exams have arisen from laboratory findings (Carpenter, Cepeda, Rohrer, Kang, & Pashler, 2012; Kang, 2016; Son & Simon, 2012). In these lab experiments, however, researchers have generally used materials that are not educationally relevant. A meta-analysis by Janiszewski, Noel, and Sawyer (2003) showed that the materials most often used were: meaningful, familiar, simple words isolated from their context. Neither have researchers used time scales for spacing that are realistic in education. Spacing gaps in the order of seconds or minutes were applied and only brief retention intervals of hardly one day or over (Cepeda et al., 2009). Few practical implications are actually based on applied research. Though there are studies that have tested the spacing effect in a more ecologically valid setting (Carpenter, Pashler, & Cepeda, 2009; Küpper-Tetzel, Erdfelder, & Dickhäuser, 2014; Rawson, Dunlosky, & Sciartelli, 2013), there is a growing need to expand these. Before practitioners make changes in their teaching approach and how they use their (often limited) teaching time, they need to know if the spacing effect holds in real classrooms with actual material from their students' course books using educationally relevant time scales (Dempster, 1988). The aim of the current research is to investigate the spacing effect in such a real classroom setting, so that highly practical guidelines can be generated for educators and students, thereby increasing the chance of both parties incorporating them.

1.1 The Spacing Effect

The best way to ensure long-term retention of to-be-learned material is to distribute practice of such material, rather than massing it together (Cepeda et al., 2006). Research into the spacing of study, meaning leaving a relatively long period of time between restudy sessions in order to ensure long-term retention of information, started over a 130 years ago, with Ebbinghaus (1913). He conducted an experiment on himself, trying to memorize a list of meaningless syllables by studying half of the list using massed practice (repeated study on the same day) and the other half by using spaced practice (repeated study across three days). Ebbinghaus concluded that spacing study over time was more advantageous than massing study at a single time.

Most of the research conducted until this day has followed Ebbinghaus' basic design; two or more study sessions are separated by (an) interstudy interval(s) (ISI). Participants become familiar

with the to-be-learned material in the first learning session and restudy the same material in (a) subsequent learning session(s). Memory is then tested in a final test session after a particular retention interval (RI). The RI is measured from the last restudy session to the test session. During this period no extra review may take place (Küpper-Tetzel, 2015). A distinction is made in the spacing literature between the spacing effect and the lag effect. The spacing effect refers to the comparison between spaced practice (there are interstudy intervals) and massed practice (there are no interstudy intervals). The lag effect involves the comparison of different interstudy intervals between study sessions. Using a research design including more than two study sessions, the different intervals can be equal (constant over time), expanding (increase over time) or contracting (decrease over time) (Küpper-Tetzel, 2015). The term distributed practice is generally used to refer to both effects (Cepeda et al., 2006).

The spacing effect has turned out to be a very solid finding in many different memory tasks such as free recall, recognition and cued-recall (Cepeda et al., 2006; Donovan & Radosevich, 1999). Evidence for the effect has been found in different populations; in animals (Scharf et al., 2002), in children (Ambridge, Theakston, Lieven, & Tomasello, 2006; Vlach, Sandhofer, & Kornell, 2008), in adults (Logan & Balota, 2008) and in memory impaired populations (Green, Weston, Wiseheart, & Rosenbaum, 2014). The spacing effect has mainly been observed in verbal learning like memorizing foreign vocabulary, facts or names of visual objects (Cepeda et al., 2009) but also in motor skill learning like surgery (Spruit, Band, Hamming, & Ridderinkhof, 2014). Moreover, the effect has been detected under various experimental manipulations such as students studying flash cards (Kornell, 2009) and students solving interleaved mathematics problems (Taylor & Rohrer, 2010). Although the majority of studies have been conducted in a laboratory setting for better control of extraneous variables, the spacing effect has also been discovered in (simulated) real-world educational settings using educationally relevant materials (Kapler, Weston, & Wiseheart, 2015; Rawson et al., 2013).

There are several consistent findings in the elaborate spacing effect literature (Maddox, 2016). In light of the current research two of the most well-researched phenomena will be discussed. The first finding is the non-monotonic memory performance function (Maddox, 2016). This means that memory performance initially increases as the length of the study interval (the lag) increases, but then decreases as the lag continues to increase. This inverted-U-shaped relationship has been reconfirmed by Cepeda et al. (2006) in a meta-analysis. The second finding is the lag x retention interval interaction. This means that the ideal length of the study interval(s), the lag(s), between restudy sessions increases as the retention interval increases (Maddox, 2016). This has been researched extensively by Cepeda, Vul, Rohrer, Wixted, and Pashler (2008) who combined RIs ranging from 7 days to 350 days and ISIs ranging from 0 to 105 days in a web-based experiment.

1.2 Underlying Mechanisms of the Spacing Effect

As of yet, no consensus has been reached regarding the underlying mechanisms that are supposedly responsible for producing the spacing effect, as the effect occurs under a wide variety of conditions (Smith & Scarf, 2017). Although different theories have been proposed during the vast history of research on distributed practice, one unifying theory which can account for all the consistent findings around the spacing effect has not yet been accepted by a majority of scholars. In the account below the various theoretical approaches are discussed and an indication is given if they can explain either the non-monotonic memory performance function and/or the lag x retention interval interaction.

One theory that explains the spacing effect is deficient processing. It suggests that when an item is studied for a second time right after it was presented for the first time (massed repetition) it receives less processing than spaced repetitions. This process might be controlled by the learner (Maddox, 2016), meaning that the second presentation is paid less attention to because it feels familiar (Kornell & Bjork, 2008). Deficient processing might also be automatic (Maddox, 2016), meaning that the processing of item one (forming an internal representation and being stored in memory) takes a certain period of time (habituation). During that time, few complete, new traces (of a repeated item) can be stored. Thus, if the second presentation of an item comes too soon after the first one, less will be remembered (Hintzman, 1974).

Another class of theories explaining the distributed practice effect is called encoding variability theory. Glenberg (1979) wrote an elaborate account in which he proposed that when a verbal item is repeated, the encoding process differs every time. These changes in encoding add information to the representation of the item and form different cues for retrieval. When repetitions are further apart in time, there is a higher chance of these informational components being different, leading to more cues that can overlap with available cues at the time of memory test. Glenberg distinguishes three types of informational components: 1) contextual (physical environment of the learner, time of studying, learner's cognitive state), 2) structural (relationship the learner sees between individual events, 3) descriptive (information about meaning).

Study-phase retrieval theory offers a third and last explanation for the distributed practice effect. This theory appears in many forms (Benjamin & Tullis, 2010; Bjork, 1994; Gerbier & Toppino, 2015; Smith & Scarf, 2017; Thios & D'Agostino, 1976; Wickelgren, 1972) but in general comes down to the assumption that a spacing effect can only occur when the second presentation of an item is recognized as a repetition of an earlier item (Küpper-Tetzel, 2015). When this is the case, the second presentation serves as a cue to activate retrieval of old information stored during the first presentation (Thios & D'Agostino, 1976). Retrieving information from memory, which happens in a spaced setting, rather than only noticing/being familiar with information, which happens in a massed setting (Kang, 2016), is a process that greatly enhances memory (Dunlosky et al., 2013).

Many theorists have added to the findings of study phase retrieval or stood at the basis of it. Bjork (1994) developed a desirable difficulty framework which proposes that long-term retention of information can only be achieved when knowledge is encoded in networks of interrelated concepts and ideas and is actively retrieved by producing the knowledge and procedures that need to be remembered. Based upon Bjork's framework, Pyc and Rawson (2009) tested their retrieval effort hypothesis, suggesting that successful but difficult processing will be better for memory than successful but easier processing. This hypothesis was indeed supported.

Benjamin and Tullis (2010) formulated their reminding account, a modern version of the study-phase retrieval theory according to Maddox (2016), meant to offer one simple account for the distributed practice effect. Their reminding model is based on three principles: 1) an item presented at a particular time will be forgotten after a particular time, 2) items in a restudy session (T2) vary in the way they can elicit reminding of items presented in t1, with repetitions having the highest capacity, followed by associated items and unrelated items, 3) the act of retrieval enhances memory and the more difficult retrieval is, the more memory benefits.

Two accounts very similar in nature to the study-phase retrieval theory are those of consolidation (Wickelgren, 1972) and reconsolidation (Gerbier & Toppino, 2015; Smith & Scarf, 2017). Consolidation theory supports the idea that during the second presentation of an item, a new (second) memory trace is formed which acquires the same state of consolidation as the first learning opportunity. Thus, when ISI is long (spaced), more consolidation will have taken place in long-term memory than when ISI is short (massed). Study phase retrieval differs from consolidation in the fact that upon the second presentation of an item, the existing memory trace gets stronger, rather than a second trace being formed (Cepeda et al., 2006).

The reconsolidation account offers an explanation for the spacing effect over long time scales. A longer lag between restudy sessions provides more time for the memory to consolidate. This ensures that the additional consolidation (reconsolidation) is more effective at strengthening that memory (Smith & Scarf, 2017) because the information shared between the original memory and the repetition will tend to be more stable and updated (Gerbier & Toppino, 2015). Part of this reconsolidation mechanism is further processing of the memory during sleep (Smith & Scarf, 2017).

Greene (1989) first proposed a two-factor model which combined study-phase retrieval and encoding variability. He suggested that study-phase retrieval, during which the consecutive presentation of an item reminds a learner of (a) previous item(s) item and subsequently starts a retrieval process, does more than just that. It also stores information about the temporal relationship between the events. This lag information is then stored in memory in the form of contextual change. Since Greene, researchers have expanded this two-factor model and are now using computational memory models to try and predict what the best time is, for a learner, to restudy their materials

(Küpper-Tetzel, Kapler, & Wiseheart, 2014). Which theories can best explain the two most consistent findings in the spacing literature, non-monotonic memory performance function and lag x retention interval interaction, remains a matter of debate. Cepeda et al. (2006) argue that deficient processing should be disqualified but that consolidation/study-phase retrieval and encoding remain candidates to explain the effects. According to Benjamin and Tullis (2010), their 'simple', unified theory of reminding suffices to explain the two findings. Finally, as reported by Maddox (2016), only the two-factor model can justify all of the data obtained so far.

1.3 The Use of Flashcards, a Self-Testing Study Strategy

Students report a wide range of study strategies for learning and preparing for exams. Rereading notes, self-testing (practice recalling material), making outlines, rereading textbooks and doing practice problems are some of them (Susser & McCabe, 2013). However, Dunlosky et al. (2013) rate (massed) rereading as having low utility when compared to more effective techniques like elaborative interrogation, self-explanation and practice testing. Students also report the use of flashcards (Hartwig & Dunlosky, 2012; Wissman, Rawson, & Pyc, 2012) to learn and retain specific detailed info (Miyatsu, Nguyen, & McDaniel, 2018). Flashcard use has shown to be effective in foreign vocabulary learning (Volpe, Mulé, Briesch, Joseph, & Burns, 2011), the retention of medical terms (Schmidmaier et al., 2011), and remembering concept definitions in the psychology domain (Rawson et al., 2013). Their effectiveness has been attributed to the fact that students retrieve information from memory when using the cards (Miyatsu et al., 2018). Therefore, the use of flashcards is considered to be a way of self-testing.

Even though flash card use represents a real-life study strategy which seems to be effective and not too difficult to teach, very little empirical research has been conducted to find out more about when and how students use them (Miyatsu et al., 2018; Wissman et al., 2012). The empirical research that does exist has been executed in experimenter-controlled environments and complemented by survey research (Wissman et al., 2012). It has shown that for self-testing to be effective it is important that students learn to criterion and in a spaced way. The survey research has made clear that there are many individual differences in how students use them (Wissman et al., 2012). These personal preferences mainly regard organization of the cards (size of the pile, dropping cards or not) and pace (how many times to flip a card, how long to study each card) (Sage, Rausch, Quirk, & Halladay, 2016). What is still unclear – because it is difficult to test or observe - is *what* exactly students do in the time they study a flashcard and they can make their own decisions. Do they just read and reread? Do they adopt the 3R method of McDaniel, Howard, and Einstein (2009) to learn the bits of text? 3R meaning read – recite (covertly in their heads?) – review. Do they read and copy a definition on paper

(when paper is provided?). And perhaps in a second phase, do they read and test themselves by trying to write down the correct definition? Or all of the above? Even though these differences are difficult to control for in a real educational setting, the self-testing strategy by means of flashcards has been used in this randomised experiment.

There are some guidelines though that can be formulated to assume some control over flashcard use and standardise it to some extent. These guidelines are based on two aspects specifically related to flashcard use; within-session spacing and metacognitive decisions if an item has been studied to criterion level, meaning being able to correctly recall the information without mistakes. Keeping within-session spacing similar (the order in which the cards are stacked) with all students, practically means that they must study their cards in the same order every round, that they cannot divide the stack in smaller stacks and that they cannot drop cards (Kornell, 2009). Raising metacognitive awareness in students, practically means that they must read external information (back of the flashcard) to check if they have remembered the information accurately and correctly. Students report that they often do not check or only check in their minds. This behaviour can lead to overconfidence which has negative effects on learning as students do not process material attentively enough (Wissman et al., 2012).

As explained before, self-testing is a form of practice testing, which is described as formative testing (assessment for learning) outside of class (Dunlosky et al., 2013). Practice testing also comes in other forms like completing practice problems or questions at the end of a textbook and doing practice tests in electronic supplement material accompanying a textbook. There is robust evidence that engaging in practice testing (retrieving information from memory) increases and enhances long-term retention of studied material; known as the testing effect (Eloisa Eisenkraemer, Jaeger, & Milnitsky Stein, 2013). Although combining spacing and testing can be a very powerful study tool (Delaney, Verkoeijen, & Spirgel, 2010), spaced retrieval practice is *not* the focus of this study. Rather, the use of flashcards will only be adopted to increase the ecological validity of the current research as it is a study strategy that students reportedly use in their actual studies. Moreover, flashcards create a convenient opportunity to study actual course content, in this case key concepts taught in an economics class. The focal point of this research is indeed the spacing effect.

1.4 Research Question & Hypothesis

As stated earlier in this introduction, it is clear from past research that the spacing effect is a robust finding. However, the effect has been mainly examined in laboratory settings. It the last 5 to 10 years researchers have begun to realise that even though laboratory findings have high internal validity, the circumstances in these experiments do not resemble the practices in actual classrooms or during

students' self-study activities at home much. This has led to issues of concern regarding the ecological validity of results found so far. This unease might have been fuelled by Mitchell (2012), who suggested that effect sizes observed in the psychological laboratory do not always correspond with effect sizes found in field studies. As a result of this consideration, a number of studies have been conducted in actual classrooms, with educationally relevant material, using educationally valid time frames (lags).

For example, Küpper-Tetzel, Erdfelder, et al. (2014) who conducted a field experiment in an authentic secondary school classroom setting and had German sixth graders practice and re-practice new German–English vocabulary from advanced chapters of their textbook in two learning sessions separated by a 0-day (massed), 1-day, or 10-day lag. Students were tested either 7 or 35 days later on their memory performance for the vocabulary pairs. Results showed that students in the massed condition correctly recalled 74% per cent of the vocabulary pairs at the retention test 7 days later, whereas the spaced condition correctly recalled 87% of the vocabulary pairs, which was significant.

Factual material in the form of U.S history facts was used in a study by Carpenter et al. (2009), in which 8th graders were taught the facts in class and reviewed them after 1 week, 16 weeks or not at all, with the final test taking place after 9 months. Review took place either by testing + feedback or by rereading the facts. The study showed that spacing effect benefits approached significance with d = 0.5, p = .060.

Finally, Rawson et al. (2013) combined testing and spacing - which they call successive relearning, adopted from Bahrick (1979) - to examine if this enhances performance on authentic outcome measures (students' actual exam results). Students enrolled in an Introductory Psychology class learned key concept definitions from actual content in their course via a virtual flashcard programme, during 14 sessions, spread out over 64 days. The students were assigned to different conditions which represented different ways of studying: successive relearning, self-regulated practice, restudy only and baseline control: learning via "business as usual". The students' scores on their actual (combined) exams showed that successive relearning improved performance by more than a letter grade based on the grading metric used by the instructor (84% versus 72%) for course exam questions tapping practiced concepts versus baseline control concepts, $t(66)=3.92$, $p<0.001$, d=0.65.

The current study adds to the search for more ecologically valid evidence to support the spacing effect. The research contributes to the existing literature because it investigates spacing in a, to my knowledge, new domain: economics. The ecological validity of the experiment is high. Firstly, the to-be-learned material was taken from the students' actual course book and consisted of concept definitions, a step up from simpler word pairs, most commonly used in spacing studies. Secondly, the learning session took place in the actual classroom and the practice sessions were simulated as self-study sessions at home. Thirdly, the method used to practice the material from the learning session is a

study method that students actually use on a frequent basis in their homes: studying flashcards. Lastly, an educationally relevant (but perhaps not for all students realistic!) time frame was used. The experiment simulated students learning for a test one week before it took place.

The central question in the present study was whether spaced study of educationally relevant material in a real classroom setting of first year Bachelor students in the economics domain yields better results on long term retention than does massed study of the to-be learned material. Based on the evidence for the spacing effect so far, the prediction was that spaced study indeed increases long-term retention.

2. Method

2.1 Design

The design of this study resembled a classic spacing effect study. It consisted of an initial learning session (in the form of a lecture), followed by either three consecutive restudy sessions of 20 minutes (massed condition) or three restudy sessions of 20 minutes spaced at least 24 hours apart (spaced condition), and a final test session. Students were randomly assigned to one of the conditions in this between-subjects design. The independent variable was the spacing condition, the dependent variable was the students' performance on the final recall test.

2.2 Participants

Three weeks before the start of the experiment, 224 first year students of Bachelor programmes International Business and Management Studies (IBMS), Finance & Control (FC) and Logistics Management Economics (LME) at a University of Applied Sciences in the Netherlands, making up the total population, were asked to join the experiment. As the researcher is employed at the institute it provided the opportunity to conduct a true experiment. Step one in the recruitment process was sending the students an email, briefly explaining the reason of the experiment and the set up. Attached to the email was an infographic, summarizing the written information. In the email it said that students could sign up for the experiment by filling out a form the week after, in either their economics class or in their study coaching class. Students were not promised credits, but would receive a gift voucher of 7.50 euros, free sets of flash cards and two articles on effective study strategies upon completion of the experiment.

Step two in the recruitment process was having the economics lecturers and student coaches actively ask the students present in their class (in lecture week 1) to join the experiment. They gave students the opportunity to read an information letter about the experiment and sign up for it by filling

out a registration form and consent form. They repeated this action in lecture week 2 for students who were absent in lecture week 1.

Step tree in the recruitment process was sending all the students who had not yet registered for the experiment after lecture week two, another email asking if they wanted to participate or not. If they wanted to, they could fill out the forms digitally and send them to the researcher. After the recruitment period of three weeks ended, 100 students agreed to participate in the experiment.

In the end, 44 students completed the experiment. Their mean age was 19.6 (range, 17 – 24 years), 45% were female and 41% were Dutch. The international students came from Oman (1), Ukraine (5), Great-Britain (1), Romania (1), Germany (4), Vietnam (5), Indonesia (2), China (2), Pakistan (2), Cyprus (1), Poland (1) and Italy (1).

2.3 Materials

In order to measure long term retention (retention after 1 week) of to-be-learned classroom material, 18 key concept definitions from a book used in an economics module were selected as study material. The key terms were taken from chapter 9, *Unemployment and Inflation*, of the book *Macroeconomics* by Hubbard, O'Brien and Rafferty (Pearson, 5th international edition). The selection of the definitions from the text, which had already been done in the book, was adopted by the researcher. In the book the terms were highlighted and printed again separately alongside the main text. Chapter 9 was selected for both practical and substantive reasons. The practical reason was that choosing another chapter, which would be discussed later in the lecture period, would not allow students to complete the experiment before the exam period. The substantive reason was that the definitions in this chapter were most diverse regarding content and most similar regarding length compared to other chapters.

In order to learn the terms in context, students read chapter 9 of the book before attending a lecture in which a powerpoint presentation was discussed. The 18 terms and definitions were printed on cardboard flashcards, ten by seven cm in size. Terms were printed on the front side and definitions on the back side and these flashcards were studied during three restudy sessions. A list of the 18 key concept definitions can be found in Table 1.

Table 1

The to-be-studied material: 18 key concept definitions

KEY CONCEPT	DEFINITION
1. Unemployed	In the government statistics, someone who is not currently at work but who is available for work, and who has actively looked for work during the previous month.
2. Labor Force	The sum of employed and unemployed workers in the economy.
3. Unemployment rate	The percentage of the labor force that is unemployed.
4. Discouraged workers	People who are available for work, but have not looked for a job during the previous four weeks, because they believe no jobs are available to them
5. Labor force participation rate	The percentage of the working-age population in the labor force.
6. Frictional unemployment	Short-term unemployment that arises from the process of matching workers with jobs.
7. Structural unemployment	Unemployment that arises from a persistent mismatch between the skills or attributes of workers and the requirements of jobs.
8. Cyclical unemployment	Unemployment caused by a business cycle recession.
9. Natural rate of unemployment	The normal rate of unemployment, consisting of frictional unemployment and structural unemployment.
10. Efficiency wage	An above-market wage that a firm pays to increase workers' productivity
11. Price level	A measure of the average prices of goods and services in the economy.
12. Inflation rate	The percentage increase in the price level from one year to the next.
13. Consumer price index (CPI)	A measure of the average change over time in the prices a typical urban family of four pays for the goods and services they purchase.
14. Producer price index (PPI)	An average of the prices received by producers of goods and services at all stages of the production process.
15. Nominal interest rate	The stated interest rate on a loan.
16. Real interest rate	The nominal interest rate minus the inflation rate.
17. Deflation	A decline in the price-level
18. Menu costs	The costs to firms of changing prices.

To measure immediate retention (after three restudy sessions) and long term retention (one week after the last restudy session) a recall test was designed (see Appendix A). Students were asked to write down the definition after each term in the test. The order of the terms listed in the test was different from the order the students had studied them during the restudy sessions, to prevent the order serving as an extra cue for recall. To score the recall tests an assessment tool was created (see

Appendix B). All the to-be-learned definitions were broken down into (different numbers of) idea units. An idea unit consists of (a combination of) words that make up the essence of the definition. When scoring the recall test, certain synonyms of words were also allowed.

2.4 Procedure

The experiment consisted of three phases. First, a study phase, in which the key terms were learned, in the form of a lecture. Second, a practice phase, in which the key terms were re-studied by using flashcards, in the form of three restudy sessions. Third, a test phase, in which the key terms were recalled, in the form of an immediate recall test after the last restudy session and a delayed recall test one week after the immediate recall test. For practical reasons (suitable material, big pool of participants, convenient planning) an economics module of seven weeks was chosen to host the experiment. In weeks two to five of the seven week module, the two conditions undertook several activities, whose sequence in time has been specified in Table 2.

2.4.1 The Study Phase

In lecture week 3 of their 7 week module, the students attended a lecture (learning session). The lecture started with a recap of the content of the previous week (15 min). Then the students were presented with the new content of chapter 9 by a powerpoint presentation that was discussed, which included the main topics of the chapter. The lecturers explained the new concepts and elaborated on important themes (45 min). During their presentation, the lecturers stimulated interaction with the students by asking them questions and making statements to provoke discussion. At the end of the class, the students were given time to process the new material by doing exercises from the book (30 min). The students were expected to come to class prepared, meaning having read chapter 9. The students who attended this lecture and who declared that they had read chapter 9, (82 out of 100 that had been recruited) were randomly assigned to one of two groups: the massed condition (40) and the spaced condition (42).

Table 2

Overview of all experimental phases in the massed and spaced condition, over time

Condition	Study phase	Study phase	Practice phase + test phase			Test phase
Massed	Read chapter 9 before going to class (week 2).	Join the lecture in lecture week 3	Study flashcards for 3 x 20 = 60 minutes (on Monday of lecture week 4). Afterwards, immediate recall test.			Delayed recall test (on Monday of lecture week 5).
Condition	Study phase	Study phase	Practice phase	Practice phase	Practice phase + test phase	Test phase
Spaced	Read chapter 9 before going to class (week 2).	Join the lecture in lecture week 3.	Study flashcards for 20 minutes (on Monday of lecture week 4).	Study flashcards for 20 minutes (on Wednesday of lecture week 4).	Study flashcards for 20 minutes (on Friday of lecture week 4). Afterwards, immediate recall test.	Delayed recall test (on Friday of lecture week 5).

2.4.2 The Practice Phase

Before the start of the practice phase, five of the researcher's teacher colleagues were instructed how to supervise the restudy sessions by means of a protocol. On the Monday of lecture week 4, both conditions started (and in the case of the massed condition also completed) their first restudy session. At the start, students in each condition were instructed on what to do by the supervisors, with the help of a powerpoint presentation. The presentation also included a demonstration of how to, specifically, study the flashcards.

To summarize the information from the powerpoint; the students were asked to study the 18 terms and definitions on the flashcards with the aim of being able to write down the definitions of the terms (in their own words) at test time. The massed condition would study the flashcards on the same day, for 3 x 20 minutes in a row, with only a short break (of two minutes) in between the sessions. The spaced condition would study the flashcards for 3 x 20 minutes on separate days, with a long break (44 – 48 hours) in between sessions. To make sure (as far as possible) that the students would study the flashcards in the same way, some restrictions were imposed and students were told (1) to study the flashcards in the same order (card 1 to 18) every time they started a new round, (2) not to divide the stack of flashcards into smaller stacks, (3) not to put flashcards (that they knew or did not know) apart but keep them in the stack, (4) to make sure to study each term at least once in every session of 20 minutes, (5) to check themselves by reading the definition on the back of the card if they thought they knew a definition, (6) to use the total time of 20 minutes. Students were freely allowed to (1) flip the cards as many times as they liked and (2) decide how long to study each flashcard. At the start of the sessions, pen and paper were handed out for the students to use as they saw fit. No further instructions on how to study the flashcards were given in order to maintain a naturalistic setting.

2.4.3 The Test Phase

After their three consecutive restudy sessions of 60 minutes in total, the massed condition took an immediate recall test to measure their short term retention of the to-be-learned material and went home. Exactly one week later they came back to fill out the same recall test to measure their long term retention of the to-be-learned material (delayed test). They had been told not to study the definitions again in between the two tests. In the recall test the items were put in a different order than they had previously been studied in the restudy sessions.

After the first restudy session on the Monday, the spaced condition, on the other hand, repeated the restudy sessions two more times and then took the immediate recall test to measure short term retention of the material. Exactly one week later they also returned to take the same recall test to measure their long term retention of the material. They had also been instructed not to study the definitions again in between the two tests. Unfortunately, at several stages of the experiment, students pulled out, leaving the total number of participants that completed the experiment at 44 (44% of the students initially recruited). Twenty students participated in the massed condition, whereas 24 students participated in the spaced condition.

2.5 Analysis

An assessment tool was created to score the recall tests. All the definitions were broken down into (different numbers of) idea units. An idea unit consists of (a combination of) words that make up the essence of the definition. For each correct idea unit one point was awarded. For each idea unit that was omitted or incorrect, no points were awarded. The maximum score to be obtained was 56 points. During assessment, mistakes in spelling and/or grammar were ignored. The students' total number of points were transformed into percentages for the analyses.

The short-term and long-term retention of both conditions was compared by using a 2 (learning condition) x 2 (retention interval) mixed ANOVA which included three covariates: subject economics, nationality and language level. Part of the students in the sample might have had particular prior knowledge about the to-be-studied material, due to their Dutch educational background. As an entry requirement to the Bachelor courses of IBMS, IFC and LM(E), Dutch students must have completed the subject economics in their secondary school curriculum. This entry requirement does not apply to the international students, who might not have studied economics at all in secondary school. Secondly, it was considered that nationality might influence the dependent variable as international students tend to be more accustomed to having to reproduce information than Dutch students, as a result of different education systems. Third, it was taken into account that the different levels of English of the students might influence the dependent variable, as their listening and reading skills were put to the test during the preparation for class, during the lecture itself and during the restudy sessions.

To compare long-term retention between the conditions an ANCOVA was used, thus controlling again for extraneous variables. The dependent variable being the score on the delayed recall test and the independent variable being the condition (spaced or massed). In all analyses a significance level α of 0.05 was applied.

In Excel, item analyses were performed on the immediate and delayed recall test to check if the materials used in the experiment (the key terms) were neither too easy nor too difficult. Moreover, to check whether fatigue or boredom might have influenced the performance of the massed condition on the immediate test, the test scores of both conditions were compared by using an independent t-test. The dependent variable being the score in the immediate recall test and the independent variable being the condition (spaced or massed).

3. Results

The researcher first scored all recall-tests that measured immediate retention (44 tests), individually. After that, in order to guarantee interrater reliability, the researcher and a colleague independently scored the recall-tests that measured long term retention of the spaced condition (24 tests, 27% of the total number of tests). There were some discrepancies in the scores. With a few subitems (the 56 idea units in the 18 definitions) the researcher wanted a specific word to be written in the idea unit, while the colleague judged it was clear the student understood the key term without mentioning the specific term. After discussing the matter, scores either remained the same or were changed (and also those of the recall-tests measuring short-term retention) accordingly. As 99% of all answers given at subitem level were scored similarly, it was concluded that both the scoring instrument and list of synonyms used were reliable. Therefore, only the researcher scored the recall-tests measuring the long term retention of the massed condition (20 tests).

3.1 Comparison of Immediate and Long-Term Retention

The mean percentage of correct recall of the to-be-learned material after 3 x 20 minutes (immediate) and after 1 week (delayed) is shown in Table 3. Results were analysed with a 2 (learning condition) x 2 (retention interval) mixed ANOVA which included three covariates: nationality, language level and subject economics. The analysis revealed a significant main effect of retention interval, $F(1, 39) = 6.022$, $p = .019$, $\eta p^2 = .132$. Within the subject groups, retention in the immediate test was higher than retention in the delayed test. Secondly, one of the covariates, nationality, was significantly related to learning condition, $F(1, 39) = 5.045$, $p = .030$, $\eta p^2 = .115$. There was no main effect of learning condition, $F(1, 39) = .089$, $p = .767$, $\eta p^2 = .002$, after controlling for the effect of nationality. Between the subject groups, there was only a numeric difference in forgetting rate. Finally, there was a marginally significant interaction effect between retention interval and condition, $F(1, 39) = 3.654$, $p = .063$, $\eta p^2 = .086$. The profile plot showed that the forgetting curve in the massed condition was steeper than the forgetting curve of the spaced condition.

Table 3

Mean percentage of correct recall of massed and spaced key concept definitions with SD in brackets

Learning condition	Immediate recall test	Delayed recall test
Spaced	77.00% (16.60)	68.90% (18.53)
Massed	76.70% (16.44)	63.20% (19.39)

3.2 Long Term Retention

Long term retention of both conditions was compared by using ANCOVA. The three covariates were not significantly related to learning condition. No effect was found, $F(1, 39) = .591$, $p = .447$, $\eta p^2 = .015$. Long term retention in the spaced condition was around 6% higher than in the massed condition (see Table 3), however, this difference was not significant.

3.3 Item-analyses

Item analyses were performed on the immediate and delayed test to check if the materials used in the experiment were neither too easy nor too difficult. It was first checked whether the key terms varied in difficulty. Table 4 shows the results of this analysis. Per key term, the test score of all students was divided by the maximum score. For the immediate test the scores ranged from 96.97% to 62.88%. For the delayed test the scores ranged from 88.64% to 40.15%. These results show that the 18 key terms varied in difficulty. Secondly, it was checked whether in the top three key terms (the easiest ones) any forgetting had occurred, or whether the definitions had already been learnt by heart (automated) before. This could have been the case due to general prior knowledge of the students, regardless of them having studied economics as a separate subject in secondary education or not. The scores in Table 5 show that forgetting did indeed occur.

3.4 Immediate Retention

The massed condition studied the flashcards for 3 x 20 minutes in a row. To check whether fatigue or boredom might have influenced their performance on the immediate test, the test scores of both conditions were compared. After 3 x 20 minutes of studying flashcards, students from the spaced and massed condition scored equally well on the recall-test (spaced: $M = 77.00\%$, SD = 16.59; massed: $M = 76.70\%$), SD = 16.44), $t(42) = .062$, $p = .950$, $d = 0.02$. Thus, there were no signs of fatigue or boredom in the massed condition after three consecutive practice sessions.

Table 4

Ranking of relative average score per item of immediate and delayed recall tests

Immediate recall test		Delayed recall test	
Relative average score per item (total scores divided by maximum score)	Number of key concept definition		Relative average score per item (total scores divided by maximum score)
96.97%	6	16	88.64%
96.59%	18	6	87.88%
92.05%	16	18	87.50%
89.39%	10	10	85.61%
87.88%	13	13	84.09%
85.23%	8	9	80.30%
82.58%	9	12	77.27%
82.58%	15	4	72.27%
81.82%	5	11	71.21%
81.82%	7	8	68.18%
80.68%	12	3	60.23%
79.09%	4	15	58.33%
78.41%	3	2	57.58%
76.52%	11	5	56.82%
74.24%	2	7	56.82%
65.34%	1	1	54.55%
63.64%	17	17	54.55%
62.88%	14	14	40.15%

Table 5

Relative average scores of top three ranked concept definitions on immediate and delayed recall test compared

Number of key concept definition	Relative average score (total score divided by maximum score) in immediate test	Relative average score (total score divided by maximum score) in delayed test
6	96.97%	87.88%
18	96.59%	87.50%
16	92.05%	88.64%

4. Discussion

The central question in the current study was whether spaced study of educationally relevant material in a real classroom setting of first year Bachelor students in the economics domain yields better results on long term retention than does massed study of the to-be learned material. Based on the evidence for the spacing effect so far, the prediction was that spaced study indeed increases long-term retention. The results of the present research have not confirmed this hypothesis. Numerically, long-term retention of the to-be-studied material was higher in the spaced condition than in the massed condition. However, this difference was not significant. Due to students dropping out at different stages of the experiment, the total number of participants was relatively low for a between-subjects design. More participants might have increased the statistical power needed to find a significant effect.

Returning to the literature on spacing and testing, there are other plausible explanations as to why a spacing effect was not found. The first matter to consider is the special nature of the sample used in the experiment: a mix of Dutch (18) and international students (26). The covariance analysis in the mixed ANOVA showed that nationality was significantly related to learning condition. This suggests that besides the manipulation, cultural differences between students have also influenced immediate and delayed retention. In fact, nationality explained much more variance than did the manipulation. In the context of the present research I specify cultural differences as (1) education system and (2) motivation. To be more specific, the majority of the international students in the sample come from education systems that place more emphasis on the lower levels of Bloom's taxonomy; remembering and understanding (Asia, Eastern Europe) than on the higher levels such as applying and analysing (the Netherlands). This difference in focus means that the international students are probably more experienced in and thus better at memorisation and rote learning. When it comes to motivation (in this respect the drive to get high grades) there is also a difference between Dutch and international

students. In Asia getting high grades is of utmost importance because of the fierce competition students face when applying at university. However, in Dutch egalitarian culture this notion does not play any part in the performance of Dutch students who are considered to be content with mediocre grades. In German culture, to give another example, obtaining high grades is crucial as well since employers scrutinize applicants'grade lists trying to find the best employees. Perhaps the international students who participated in the experiment could be considered to be more driven (on average) than the Dutch students. On both recall tests, the international students significantly outperformed the Dutch students.

The idea that individual differences between students is a factor to take into account in research on spacing, is shared by others. In their exploration of the effectiveness of 10 different learning techniques, Dunlosky et al. (2013) conclude that distributed practice yields positive results in different age groups. However, they identify an open issue: to what extent the spacing effect may be moderated by other individual characteristics, such as prior knowledge and motivation. They suggest future research should examine possible individual differences beyond age. Delaney et al. (2010) also note that numerous variables which might optimize the spacing effect have been investigated over the years, but that not many studies have assessed individual differences. They warn that there is no such thing as one size fits all: a learning schedule that improves retention for one student might have neutral of even damaging effects for another. Delaney et al. (2010) have singled out baseline memory ability (how well can a person retrieve previously seen items from memory) as a source of individual differences. Research by Verkoeijen and Bouwmeester (2008) shows that the spacing effect is smaller for college students with an overall lower memory-performance level than for students with an overall higher memory-performance level. Relating these results to the current study, it can be said that the Dutch students – possibly because of their lack of training in memorisation – had a lower memory performance than the international students. Had the current study been conducted with either only Dutch students or only international students, the outcome might have been different.

Another explanation why a spacing effect was not found is related to the manner in which the to-be-learned material was studied during the restudy sessions. Students in the massed condition might have benefited from more and different retrieval practice than students in the spaced condition. Students in the massed condition studied for 3 x 20 minutes in a row, without a break. At some point during the session they might have felt they had studied the material to an acceptable criterion level (being able to reproduce the definitions correctly) because the definitions seemed familiar. As a consequence, some students may have moved on to self-testing earlier that students in the spaced condition. They may have, for example, written down the definitions on paper (which had been provided) or made little drawings. During the restudy sessions the supervisors indeed saw students in the massed condition doing this. For the massed condition, writing the definitions down on paper

might have served as an extra visual encoding cue to remember them. More so than for the spaced condition who probably self-tested in this way less because they needed more time to restudy to criterion level after each lag of 24 hours. Kornell (2009) found in his study (in which participants could choose how long to study a flashcard) that spaced and massed conditions did not differ significantly in terms of the amount of time participants chose to spend studying. He had expected that the massed condition would spend less time studying, based on one of the explanations of the spacing effect that suggests people pay less attention to repetitions under massed conditions because they perceive the items as more fluent and well learned. However, this was not the case in his experiment and Kornell proposed this was because the massed condition used the (extra) time to test themselves. It is possible that the massed condition in the current study did the same as they had the opportunity (pen and paper) and time (they had to study for the full 60 minutes).

Research by Cull (2000) confirms that different types of restudy lead to different scores in final recall. Cull tried to untangle the benefits that distributed study and testing have on learning. His aims were to compare the effect of testing and expanding test spacing when tests with and without feedback were provided within a single learning session (Experiment 1), immediately after initial learning (Experiment 2), and days after initial learning (Experiments 3 and 4). Experiment 4 resembles the current research most and will thus be focused on. In an educationally relevant learning situation, using actual vocabulary words, participants were assigned to one of three spacing conditions (expanding, uniform and massed) in which they were asked to complete four different review sessions (study-only, test-only, test-study and no review). The study-only review meant that participants received one card with a word and definition for 10 items, the test-only review meant one card with just a word for 10 items, the test-study review meant one card with just a word followed by a second card with the same word and definition for 10 items, and no review for a remaining 10 items. A retention test followed 8 days later. Results showed that in the uniform spacing condition, the proportion of correct recall in the final test was significantly higher than the proportion in the massed spacing condition. A finding which has not been replicated in the present study. However, Cull's results do make clear that recall was highest in the test-study review session, followed by test-only review, study-only review and no review. These sessions resemble the approaches the students in the current experiment are likely to have used. The participants were only instructed to 'study' the flashcards and then check themselves when they thought they knew a definition. Since it was impossible to monitor what students actually did, it is not clear what the final balance was in their approach. Did they study (reread) more or did they self-test more? And when they self-tested, did they check their answers or not? In other words, was it test-study or test-ony? If the massed condition tested more than they studied (for reasons mentioned before) this might be an explanation why their scores on the tests more or less equalled those of the spaced condition.

Pashler, Rohrer, Cepeda, and Carpenter (2007) have also tried to shed more light on the consequences different choices regarding spacing and testing have for learning and forgetting. They have discussed the matter of overlearning of to-be-studied material, an issue that might also have played a role in the current research. Overlearning: continuing to practice material after error-free performance is attained has been shown to increase retention of the material at a later stage (Pashler et al., 2007). A meta-analysis by Driskell, Willis, and Copper (1992) suggests that the more someone overlearns, the better later retention will be. They also found that, for cognitive tasks, longer retention intervals weakened the overlearning effect. Driskell et al. established, from the research conducted up to that point, that after 19 days the overlearning effect diminished by half and was gone completely after 38 days. Rohrer, Taylor, Pashler, Wixted, and Cepeda (2005) had participants study a word-pair list containing 10 items either 5 times or 20 times. Results showed that retention of students who had overlearned exceeded the retention of students who had underlearned for the first three weeks. After that, retention in both groups remained equal until week 9. Relating the results above to the current experiment, especially in the massed condition, the to-be-studied material might have been overlearned. In order to keep the spacing between the flashcards equal, students were instructed not to take cards out and to always study them in the same order. This means they were forced to read and check themselves on all key terms, even though they already knew them. Moreover, students were required to keep studying the cards until the 60 minutes had passed. For some students in the massed condition the 60 minutes might have been longer than they needed to learn the material. This was actually confirmed by some students in the massed condition who told the supervisor after round two that they already knew most of the terms. For the spaced condition the 60 minutes were not necessarily too long, as it took them longer to retrieve the items from memory. So, the overlearning in the massed condition might have been a reason why both conditions did not differ much in the immediate and delayed retention test. Especially seeing that the retention interval was only a week. This is a time scale on which the overlearning effect is assumed to be still effective, as concluded by Driskell et al. (1992).

Interestingly, Driskell et al. (1992) also raise the issue of individual differences in ability among learners. They state that the ability level of learners affects the level of initial skill acquisition but has little effect on the rate of retention. In other words, learners with a high memory performance may retain more information, but the rate of decay in retention is similar for both high-and low ability learners. The strategy of overlearning may ensure that all learners acquire initial skill mastery and retain the information for an optimal period. In the present experiment, as made clear before, an overlearning strategy was no deliberately employed. However, extending the prior assumption that Dutch students might be lower-ability learners than international students, the Dutch students in the

massed condition may have done better than they would otherwise have done because of this unintended opportunity to overlearn.

The aim of this study was to provide more ecologically valid evidence to support the spacing effect. Its strength, the naturalistic approach, has also shown to be its most important limitation. More specifically, the choice not to prescribe how exacty the students should study the flashcards has caused ambiguity. Even though some guidelines were laid down, the students' differences in approach unintentionally caused subconditions within the two main conditions which made them more difficult to compare. For example, subconditions like not self-testing at all, self-testing a lot, self-testing by writing things down or drawing and overlearning. This way other factors than the spacing manipulation alone played a part. Factors that seem to have been more influential.

Despite the limitations of the study, caused by its lifelike nature, the numeric difference in the score on long-term retention between the conditions, though not significant, points to the possibility that the spacing effect does exist in a naturalistic setting, outside the laboratory.As for the practical guidelines for educators and students, which I set out to collect, the results of the study seem to suggest they should take into account there is no one size fits all approach to adopt when it comes to spacing study sessions. Individual differences between students, such as their memory performance level, imply that different students benefit from different learning schedules.

Based on the present study there are four suggestions for future research I would like to highlight. First, in order to provide more tailor-made guidelines to educators and students worldwide, future research could focus on investigating the spacing effect in more culturally diverse student populations, thus taking into account different education systems. Then, for educators and students to take 'the trouble' and actually implement spacing strategies in their daily routines, they would probably want to see long term effects of their endeavours. Therefore, future research should measure long-term retention at multiple, educationally relevant points in time, for example after a week (to mimic an exam), after eight weeks (to mimic the end of a module) and after 6 months (to mimic the start of a new semester). Third, the current research has also given new insights into the different approaches students have when they study flashcards. From observing the students the supervisors could tell that their strategies included: read-recite-review, writing down the definitions on paper (as self-testing and as a way to become familiar with the content) and making little drawings or symbols on paper. Perhaps future research could investigate if one approach is more effective than others. Finally, it might be time for researchers and practitioners to do some soul-searching and ask themselves why most (of their) students are focused on short term retention of to-be-studied material in the first place? Perhaps we (being both a researcher and practitioner I include myself) should acknowledge that the education system we are upholding does not stimulate long-term retention of information because of its primary focus on testing and assessment. It is no surprise that in an educational culture that is

fixated on assessment *of* learning (summative), most students are only interested in learning for the test. Perhaps, in an educational culture that also concentrates on assessment *for* learning (formative), students will feel stimulated to improve in exactly those areas in which they lack (foundational) knowledge and skills. Hopefully such self-motivated improvements will have the long lasting effects practitioners are looking for. As long as students feel the pressure of 'passing the test', they are more likely to continue their (successful) habit of cramming. In this respect, more research on spacing seems quite futile. Devoting our efforts to investigating and helping to establish a feedback culture in schools, rather than a test culture, might be a more valuable direction to take.

5. Reference list

Ambridge, B., Theakston, A. L., Lieven, E. V. M., & Tomasello, M. (2006). The distributed learning effect for children's acquisition of an abstract syntactic construction. *Cognitive Development, 21*(2), 174-193. doi:10.1016/j.cogdev.2005.09.003

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General, 108*(3), 296-308. doi:10.1037/0096-3445.108.3.296

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology, 61*(3), 228-247. doi:https://doi.org/10.1016/j.cogpsych.2010.05.004

Bjork, R. (1994). *Memory and Meta-memory Considerations in the Training of Human Beings*.

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, H. K., & Pashler, H. (2012). Using Spacing to Enhance Diverse Forms of Learning: Review of Recent Research and Implications for Instruction. *Educational Psychology Review, 24*(3), 369-378. doi:10.1007/s10648-012-9205-z

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using Tests to Enhance 8th Grade Students' Retention of US History Facts. *Applied Cognitive Psychology, 23*(6), 760-771. doi:10.1002/acp.1507

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing Distributed Practice Theoretical Analysis and Practical Implications. *Experimental Psychology, 56*(4), 236-246. doi:10.1027/1618-3169.56.4.236

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychol Bull, 132*(3), 354-380. doi:10.1037/0033-2909.132.3.354

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing Effects in Learning: A Temporal Ridgeline of Optimal Retention. *Psychological Science, 19*(11), 1095-1102. doi:10.1111/j.1467-9280.2008.02209.x

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14*(3), 215-235.

Dalton, J., Smith, D., & Administration, V. O. o. S. (1989). *Extending Children's Special Abilities: Strategies for Primary Classrooms*: Office of Schools Administration, Ministry of Education, Victoria.

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Chapter 3 - Spacing and Testing Effects: A Deeply Critical, Lengthy, and At Times Discursive Review of the Literature. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 53, pp. 63-147): Academic Press.

Dempster, F. N. (1988). The Spacing Effect: A Case Study in the Failure to Apply the Results of Psychological Research. *American Psychologist, 43*(8), 627-634. doi:10.1037/0003-066X.43.8.627

Donovan, J. J., & Radosevich, D. J. (1999). A Meta-Analytic Review of the Distribution of Practice Effect: Now You See It, Now You Don't. *Journal of Applied Psychology, 84*(5), 795-805. doi:10.1037/0021-9010.84.5.795

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of Overlearning on Retention. *Journal of Applied Psychology, 77*(5), 615-622. doi:10.1037/0021-9010.77.5.615

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology. *Psychological Science in the Public Interest, 14*(1), 4-58. doi:10.1177/1529100612453266

Eloisa Eisenkraemer, R., Jaeger, A., & Milnitsky Stein, L. (2013). *A Systematic Review of the Testing Effect in Learning* (Vol. 23).

Gerbier, E., & Toppino, T. C. (2015). The effect of distributed practice: Neuroscience, cognition, and education. *Trends in Neuroscience and Education, 4*(3), 49-59. doi:10.1016/j.tine.2015.01.001

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7*(2), 95-112.

Green, J. L., Weston, T., Wiseheart, M., & Rosenbaum, R. S. (2014). Long-Term Spacing Effect Benefits in Developmental Amnesia: Case Experiments in Rehabilitation. *Neuropsychology, 28*(5), 685-694. doi:10.1037/neu0000070

Greene, R. L. (1989). Spacing Effects in Memory: Evidence for a Two-Process Account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(3), 371-377. doi:10.1037/0278-7393.15.3.371

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review, 19*(1), 126-134. doi:10.3758/s13423-011-0181-y

Hintzman, D. (1974). *Theoretical implications of the spacing effect*.

Janiszewski, C., Noel, H., & Sawyer, Alan G. (2003). A Meta-analysis of the Spacing Effect in Verbal Learning: Implications for Research on Advertising Repetition and Consumer Memory. *Journal of Consumer Research, 30*(1), 138-149. doi:10.1086/374692

Kang, S. H. (2016). Spaced Repetition Promotes Efficient and Effective Learning: Policy Implications for Instruction. *Policy Insights from the Behavioral and Brain Sciences, 3*(1), 12-19.

Kapler, I. V., Weston, T., & Wiseheart, M. (2015). Spacing in a simulated undergraduate classroom: Long-term benefits for factual and higher-level learning. *Learning and Instruction, 36*, 38-45. doi:10.1016/learninstruc.2014.11.001

Kornell, N. (2009). Optimising Learning Using Flashcards: Spacing Is More Effective Than Cramming. *Applied Cognitive Psychology, 23*(9), 1297-1317. doi:10.1002/acp.1537

Kornell, N., & Bjork, R. A. (2008). Learning Concepts and Categories: Is Spacing the "Enemy of Induction"? *Psychological Science, 19*(6), 585-592. doi:10.1111/j.1467-9280.2008.02127.x

Küpper-Tetzel, C. (2015). Understanding the distributed practice effect. *Zeitschrift für Psychologie*.

Küpper-Tetzel, C., Erdfelder, E., & Dickhäuser, O. (2014). The lag effect in secondary school classrooms: Enhancing students' memory for vocabulary. *Instructional Science, 42*(3), 373-388. doi:10.1007/s11251-013-9285-2

Küpper-Tetzel, C., Kapler, I. V., & Wiseheart, M. (2014). Contracting, equal, and expanding learning schedules: The optimal distribution of learning sessions depends on retention interval. *Memory & Cognition, 42*(5), 729-741. doi:10.3758/s13421-014-0394-1

Logan, J. M., & Balota, D. A. (2008). Expanded vs. Equal Interval Spaced Retrieval Practice: Exploring Different Schedules of Spacing and Retention Interval in Younger and Older Adults. *Aging, Neuropsychology, and Cognition, 15*(3), 257-280. doi:10.1080/13825580701322171

Maddox, G. B. (2016). Understanding the underlying mechanism of the spacing effect in verbal learning: a case for encoding variability and study-phase retrieval. *Journal of Cognitive Psychology, 28*(6), 684-706.

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The Read-Recite-Review Study Strategy:Effective and Portable. *Psychological Science, 20*(4), 516-522. doi:10.1111/j.1467-9280.2009.02325.x

Mitchell, G. (2012). Revisiting Truth or Triviality: The External Validity of Research in the Psychological Laboratory. *Perspectives on Psychological Science, 7*(2), 109-117. doi: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. doi:10.1177/1745691617710510

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

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review, 14*(2), 187-193. doi:10.3758/bf03194050

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60*(4), 437-447. doi:https://doi.org/10.1016/j.jml.2009.01.004

Rawson, K. A., Dunlosky, J., & Sciartelli, S. M. (2013). The Power of Successive Relearning: Improving Performance on Course Exams and Long-Term Retention. *Educational Psychology Review, 25*(4), 523-548. doi:10.1007/s10648-013-9240-4

Rohrer, D., Taylor, K., Pashler, H., Wixted, J. T., & Cepeda, N. J. (2005). The effect of overlearning on long-term retention. *Applied Cognitive Psychology, 19*(3), 361-374. doi:10.1002/acp.1083

Sage, K., Rausch, J., Quirk, A., & Halladay, L. (2016). Pacing, Pixels, and Paper: Flexibility in Learning Words from Flashcards. *Journal of Information Technology Education: Research, 15*, 431-456.

Scharf, M. T., Woo, N. H., Lattal, K. M., Young, J. Z., Nguyen, P. V., & Abel, T. (2002). Protein Synthesis Is Required for the Enhancement of Long-Term Potentiation and Long-Term Memory by Spaced Training. *Journal of Neurophysiology, 87*(6), 2770-2777. doi:10.1152/jn.00792.2001

Schmidmaier, R., Ebersbach, R., Schiller, M., Hege, I., Holzer, M., & Fischer, M. R. (2011). Using electronic flashcards to promote learning in medical students: retesting versus restudying. *Medical Education, 45*(11), 1101-1110. doi:10.1111/j.1365-2923.2011.04043.x

Smith, C. D., & Scarf, D. (2017). Spacing Repetitions Over Long Timescales: A Review and a Reconsolidation Explanation. *Frontiers in Psychology, 8*. doi:10.3389/fpsyg.2017.00962

Son, L. K., & Simon, D. A. (2012). Distributed Learning: Data, Metacognition, and Educational Implications. *Educational Psychology Review, 24*(3), 379-399. doi:10.1007/s10648-012-9206-y

Spruit, E. N., Band, G. P. H., Hamming, J. F., & Ridderinkhof, K. R. (2014). Optimal training design for procedural motor skills: a review and application to laparoscopic surgery. *Psychological Research, 78*(6), 878-891. doi:10.1007/s00426-013-0525-5

Susser, J. A., & McCabe, J. (2013). From the lab to the dorm room: metacognitive awareness and use of spaced study. *Instructional Science, 41*(2), 345-363. doi:10.1007/s11251-012-9231-8

Taylor, K., & Rohrer, D. (2010). The Effects of Interleaved Practice. *Applied Cognitive Psychology, 24*(6), 837-848. doi:10.1002/acp.1598

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior, 15*(5), 529-536. doi:https://doi.org/10.1016/0022-5371(76)90047-5

Verkoeijen, P. P. J. L., & Bouwmeester, S. (2008). Using latent class modeling to detect bimodality in spacing effect data. *Journal of Memory and Language, 59*(4), 545-555. doi:10.1016/j.jml.2007.09.005

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition, 109*(1), 163-167. doi:10.1016/j.cognition.2008.07.013

Volpe, R. J., Mulé, C. M., Briesch, A. M., Joseph, L. M., & Burns, M. K. (2011). A Comparison of Two Flashcard Drill Methods Targeting Word Recognition. *Journal of Behavioral Education, 20*(2), 117-137. doi:10.1007/s10864-011-9124-y

Wickelgren, W. A. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology*, *9*(4), 418-455.

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2012). How and when do students use flashcards? *Memory*, *20*(6), 568-579. doi:10.1080/09658211.2012.687052

Appendix A

The immediate and delayed recall test

Retention test

Write down the definition in the right column. Please write clearly.

1. Consumer price index (CPI)	
2. Inflation rate	
3. Labor force	
4. Unemployed	
5. Price level	
6. Menu costs	
7. Frictional unemployment	
8. Nominal interest rate	
9. Efficiency wage	
10. Real interest rate	
11. Labor force participation rate	
12. Discouraged workers	
13. Unemployment rate	
14. Producer price index (PPI)	
15. Natural rate of unemployment	
16. Deflation	
17. Structural unemployment	
18. Cyclical unemployment	

Appendix B

Measurement Tool

Key term (with number of idea units)	Complete definition with idea units (in green)	Separate idea units
1. Consumer price index (CPI) (4)	A measure of the average change over time in the prices a typical urban family of four pays for the goods and services they purchase.	average change over time / prices urban family of four goods and services
2. Inflation rate (3)	The percentage increase in the price level from one year to the next.	percentage increase price level one year to the next
3. Labor Force (2)	The sum of employed and unemployed workers in the economy.	sum of employed and unemployed workers economy
4. Unemployed (5)	In the government statistics, someone who is not currently at work but who is available for work, and who has actively looked for work during the previous month.	government statistics not at work available for work looked for work previous month
5. Price level (3)	A measure of the average prices of goods and services in the economy.	average prices goods and services economy
6. Menu costs (3)	The costs to firms of changing prices.	costs firms changing prices
7. Frictional unemployment (2)	Short-term unemployment that arises from the process of matching workers with jobs.	short-term unemployment process of matching workers with jobs
8. Nominal interest rate (2)	The stated interest rate on a loan.	stated interest rate loan
9. Efficiency wage (3)	An above-market wage that a firm pays to increase workers' productivity.	above-market wage firm increase workers' productivity

10. Real interest rate (3)	The nominal interest rate minus the inflation rate.	nominal interest rate minus inflation rate
11. Labor force participation rate (3)	The percentage of the working-age population in the labor force.	percentage working-age population labor force
12. Discouraged workers (4)	People who are available for work, but have not looked for a job during the previous four weeks, because they believe no jobs are available to them	available for work not looked for a job previous four weeks believe no jobs are available
13. Unemployment rate (3)	The percentage of the labor force that is unemployed.	percentage labor force unemployed
14. Producer price index (PPI) (3)	An average of the prices received by producers of goods and services at all stages of the production process.	average of the prices producers of goods and services at all stages of the production process
15. Natural rate of unemployment (3)	The normal rate of unemployment, consisting of frictional unemployment and structural unemployment.	normal rate of unemployment frictional unemployment structural unemployment
16. Deflation (2)	A decline in the price-level.	decline price-level
17. Structural unemployment (4)	Unemployment that arises from a persistent mismatch between the skills or attributes of workers and the requirements of jobs.	unemployment persistent mismatch skills or attributes requirements of jobs
18. Cyclical unemployment (2)	Unemployment caused by a business cycle recession.	unemployment business cycle recession