UMEÅ UNIVERSITY

RETRIEVAL PRACTICE AND INDIVIDUAL DIFFERENCES

Exploring Factors Relevant to the Benefit and Use of Retrieval Practice

Frida Bertilsson

Department of Psychology
Umeå 2023

Dissertation for PhD
ISBN: 978-91-8070-182-2 (print)
ISBN: 978-91-8070-183-9 (pdf)
Cover photo: Frida Bertilsson
Electronic version available at: http://umu.diva-portal.org/
Printed by: Cityprint i Norr AB
Umeå, Sweden 2023

The capacity to learn is a gift

The ability to learn is a skill

The willingness to learn is a choice

- Brian Herbert

Table of Contents

Abstract

When students are studying outside of the classroom, they tend to use learning strategies that research has found to be relatively ineffective. This is problematic as a relatively large proportion of students do not finish compulsory school and/or upper-secondary school with sufficient grades. Although these problems stem from many different factors, one factor can be assumed to be shortcomings in students' acquisition of knowledge.

Retrieval practice is a highly effective learning strategy that has been shown to enhance retention, compared to other more commonly used strategies such as repeated study. Retrieval practice involves repeated retrieval of information from long-term memory to enhance recall of that information. This learning strategy has often been described as repeated testing of a material, and consequently, the enhanced retention observed after retrieval practice is called the testing effect. While research has found substantial evidence for the testing effect, less is known about the potential impact of individual differences on the testing effect as well as use of the strategy.

Therefore, the overall aim of this thesis was to explore the relationship between retrieval practice and individual differences, both in terms of the testing effect and the optional use of the strategy in an authentic school environment. This was done in a series of studies. Study I and II evaluated the relationship between individual differences in cognitive (i.e., working memory capacity) and non-cognitive factors (i.e., personality characteristics) and the testing effect using quasi-experimental designs. This was studied using a task where participants learned Swedish – Swahili word pairs through retrieval practice or repeated study. The results indicated that retrieval practice led to enhanced recall irrespective of individual differences in cognitive or non-cognitive factors. In Study III, retrieval practice (i.e., repeated online quizzes) was implemented as a learning strategy in an upper-secondary school. Over eight quizzes, students' use of retrieval practice was measured, and they completed tests of their cognitive abilities and self-report measures of non-cognitive factors, such as personality. It was found that retrieval practice was used to a low extent, and that individual differences in non-cognitive factors was related to the amount of quiz use. Quiz use also differed between female and male students.

To conclude, the results indicate that retrieval practice is an effective strategy regardless of individual differences in cognitive and non-

cognitive factors for learning, which makes it a useful strategy in a broad sense. However, regarding the *use of* retrieval practice as an optional learning strategy in a school context, the results showed that students use retrieval practice to a limited extent, and the use seems to vary among students with different characteristics, such as degree of conscientiousness. Based on these findings, it is important to support students in using retrieval practice to ensure that all students benefit.

Abbreviations

ANOVA	Analysis of Variance
AOSPAN	Automated Operation Span Task
ARS	Audience Response System
FFM	Five-Factor Model of Personlity
METQ	The Mental Effort Tollerance Questionnaire
Mini-IPIP	The mini International Personality Item Pool
NFC	Need for Cognition
RAPM	Ravens Advanced Progressive Matrices
WMC	Working Memory Capacity

Sammanfattning på svenska

Att elever använder sig av effektiva studiestrategier/inlärningsstrategier är viktigt för lärandet. Tidigare forskning har visat att elever i stor utsträckning använder sig av inlärningsstrategier som är mindre effektiva när de studerar på egen hand. Detta är problematiskt då ca 15 % av eleverna som går ut grundskolan saknar behörighet till gymnasiet, och av de som faktiskt börjar gymnasiet har ca 17 % inte tagit studenten efter fem år. Att sakna gymnasieexamen kan i sin tur leda till svårigheter att med få ett jobb och försörja sig. Dessa brister beror självklart på många faktorer, men en viktig faktor kan antas vara brister i att tillägna sig kunskap under skoltiden.

Testbaserat lärande är en inlärningsstrategi som har visat sig vara effektiv och leda till bättre minne och lärande över tid i jämförelse med andra strategier som att till exempel läsa en text upprepade gånger. Den viktiga mekanismen bakom strategin är att upprepade gånger försöka minnas, det vill säga testa sig själv på, information som ska läras in. Att upprepade gånger plocka fram informationen från långtidsminnet stärker minnesspåren och gör dels att minnet blir mer lättåtkomligt vid nästa framplockning, dels att det lagras bättre över tid - jämfört med andra strategier såsom att upprepat läsa informationen. Effekten av upprepad framplockning, det vill säga den förbättrade inlärning som sker via testbaserat lärande, kallas testeffekten. Trots att en stor mängd forskning har visat på hur fördelaktigt det är att använda testbaserat lärande saknas fortfarande kunskap om huruvida elever med olika egenskaper och förutsättningar gynnas olika mycket av att använda strategin. En ytterligare viktig och till stor del obeforskad fråga är också *hur* elever använder sig av testbaserat lärande när de studerar på egen hand, utanför lektionstid.

I denna avhandling undersöktes, i en serie studier, sambandet mellan testbaserat lärande och individuella skillnader i kognitiva faktorer (exempelvis arbetsminne) samt icke-kognitiva faktorer (såsom personlighetsdrag), som visat sig vara relaterade till inlärning och skolprestation. Syftet med avhandlingen var att undersöka dels om individuella skillnader är relaterade till skillnader i testeffekten, dels hur studenter använder testbaserat lärande som en frivillig studiestrategi, samt om användningen av strategin är relaterad till individuella skillnader. I studie I och II undersöktes huruvida individuella skillnader i kognitiva och icke-kognitiva faktorer är relaterade till testeffekten. I studierna användes kvasiexperimentell design där deltagarna lärde sig

svenska – swahili-ordpar genom testbaserat lärande eller upprepad läsning. Resultaten visade att testbaserat lärande ledde till förbättrad inlärning jämfört med att enbart läsa ordparen, samt att testeffekten inte var relaterad till individuella skillnader i vare sig kognitiva eller icke-kognitiva faktorer. I Studie III implementerades testbaserat lärande, i form av online quiz, som en studiestrategi på en gymnasieskola. Quizzen fanns tillgängliga för eleverna på deras ordinarie lärplattform och eleverna fick information om hur effektiv testbaserat lärande är. I studien undersökte vi hur eleverna använde sig av quizzen och de fick även genomföra tester av deras kognitiva förmågor samt fylla i enkäter som mäter personlighetsdrag. Resultaten visade att testbaserat lärande användes på frivillig basis i relativt liten utsträckning och att individuella skillnader i icke-kognitiva faktorer, som exempelvis samvetsgrannhet, var relaterade till frivillig användning. Användningen av quiz skilde sig även åt mellan kvinnliga och manliga elever.

Sammantaget visade resultaten från denna avhandling att testbaserat lärande är en effektiv strategi oberoende av individuella skillnader i kognitiva samt icke-kognitiva faktorer, vilket gör den till en användbar strategi i bred bemärkelse. Däremot verkar *användandet* av testbaserat lärande i skolmiljö samvariera med individuella skillnader, vilket innebär att elever med till exempel högre samvetsgrannhet använde quizzen i större utsträckning. För att försäkra sig om att denna studiestrategi kommer alla elever till gagn är det alltså viktigt att stödja elever i användandet av testbaserat lärande.

Baserat på dessa resultat vill jag passa på att ge några praktiska råd till lärare, lärarutbildare, samt andra intresserade.

- Vi vet att testbaserat lärande fungerar och leder till bättre och mer hållbar inlärning – använd det i undervisningen.
- Vi vet också at elever/studenter i alla åldrar är relativt dåliga på att använda effektiva inlärningsstrategier. De vet oftast inte vad de ska studera eller på vilket sätt de bör studera och de bör därför ges stöd till att använda effektiva studiestrategier även utanför klassrummet.
- Lär blivande lärare om kognitionspsykologiska principer för lärande. Att ge alla lärare grundläggande kunskap om lärandets vetenskap, och en medvetenhet om hur man uppmuntrar till självreglerat lärande skulle ge dem värdefulla redskap till att stödja eleverna att bli effektiva i sin inlärning.
- Undervisa om kognitionspsykologiska principer för lärande. Det faktum att elever/studenter har problem med att reglera sitt lärande på gymnasie- och universitetsnivå indikerar att mer behöver göras

tidigare i skolsystemet. Elever bör få lära sig från väldigt unga åldrar varför vissa typer av metoder gynnar inlärningen mer, samt när olika metoder ska appliceras för att inlärningen ska bli effektiv. Om detta ingick som en del av den dagliga undervisningen så skulle *hur* man lär sig inte behöva vara ett hinder för eleverna.

List of papers

I. Bertilsson, F., Wiklund-Hörnqvist, C., Stenlund, T., & Jonsson, B. (2017). The testing effect and its relation to working memory capacity and personality characteristics. *Journal of Cognitive Education and Psychology*, *16*(3), 241–259. https://doi.org/10.1891/1945-8959.16.3.241

II. Bertilsson, F., Stenlund, T., Wiklund-Hörnqvist, C., & Jonsson, B. (2021). Retrieval practice: Beneficial for all students or moderated by individual differences? *Psychology Learning and Teaching*, *20*(1), 21–39. https://doi.org/10.1177/1475725720973494

III. Bertilsson, F., Stenlund, T., Sundström, A., & Jonsson, B (Submitted). Self-regulated use of retrieval practice: Associations with individual differences in non-cognitive and cognitive factors. *Manuscript*

Introduction

Most of us have been there, with our nose in a book or glued to the computer screen way past bedtime, cramming for a test the next day. While there was always a classmate who seemed to have figured it all out and who had already started studying a week or two before the test, many students, including myself, left it until the last minute. What we did not know was that it was unlikely that we would remember much of the information that we had learned while cramming after a few days or weeks. You see, the way in which we learn new information affects how well it will be encoded in our memory and, in turn, for how long and with what ease we will be able to retrieve it (Bjork & Bjork, 2011).

So, using different learning strategies will affect students' achievements. Unfortunately, students often report that they tend to rely on strategies that are known to not be optimally effective (Dunlosky et al., 2013). This is highly problematic considering that, in 2022, 15% of Swedish students who finished compulsory school (grade nine, age 15–16 years) were not qualified to apply for upper-secondary school (Skolverket, 2022b), and of those who start upper-secondary school, about 17% do not graduate within five years. Altogether, 27% of the children in a given age cohort do not graduate from upper-secondary school (Skolverket, 2022a). These numbers are alarming, and it is of course a significant problem because people without a diploma from upper-secondary school are at higher risk of unemployment.

There will of course be many factors contributing to why some students do not reach the required goals, and a lack of effective ways of acquiring knowledge is likely to be one such factor. In fact, it is often reported in the scientific literature that students have little knowledge about how to study effectively (Dunlosky et al., 2013). This suggests that more focus should be put on teaching students effective learning strategies. One step in this direction is the recent (2021-12-22) decision that teacher education in Sweden should incorporate learning and cognitive science (SFS 2021:1335). Ensuring that teachers have in-depth knowledge about the science of learning, in addition to teaching, could perhaps increase both their own and their students' awareness about how to learn effectively.

Over the last decade, a large amount of research has identified one specific learning strategy as superior when compared to others (see Dunlosky et al., 2013, for a review), and it is important to communicate this to students as well as teachers. This learning strategy is called retrieval practice and it

is the focus of this thesis. The term retrieval practice refers to the act of repeatedly retrieving to-be-learned information from long-term memory. This can be accomplished in many ways; for example, by using practice tests or quizzes, drawing mind maps from memory, or asking and answering questions with a friend. Other names used for this technique are test-enhanced learning, practice testing, and self-testing.

Beyond learning strategies, there are several other factors that influence learning and academic success. Some of these factors might also have an impact on how students use effective learning strategies. Cognitive factors, such as working memory capacity and intelligence, will influence how well a student is able to make sense of new information and solve problems, and they are well known to have a relatively large influence on learning and academic achievement (Alloway & Alloway, 2010; Roth et al., 2015). Furthermore, non-cognitive factors, such as personality, could influence how students engage with and use learning strategies. For example, previous research suggests that, if you are a person who can be described as having high grit, you are more likely to achieve academic success (Duckworth et al., 2007).

Specifically, this doctoral thesis examines whether individual differences in cognitive and non-cognitive factors are related to the observed benefits of retrieval practice, and/or to how students use retrieval practice when studying.

In this thesis, as is customary in the research area of psychology, various psychological constructs are examined. Therefore, it is important to explain this notion. Psychological constructs can be defined as labels that we use to describe collections of behaviors or characteristics that are of interest for our studies (AERA, APA & NCME, 2014). For example, memory, intelligence, and grit are some of the constructs of interest that are used in this thesis. Constructs such as these are used to allow us to gain access to attributes that are not directly observable. It is not possible to physically measure someone's intelligence, because intelligence is a mental construction that has been created to label certain behaviors that are commonly considered intelligent behaviors. In order to measure intelligence, the scientific community has operationalized several behaviors or characteristics that are assumed to represent intelligence, and we attempt to measure those behaviors and make inferences about a person's intelligence from the observations made (AERA, APA & NCME, 2014). It is important to keep in mind that these measurements vary in reliability and validity. Thus, while researchers may confidently state that intelligence is highly important for learning, such statements contain

many reservations; for example, a specific intelligence test may only be validated for certain age groups or settings.

The thesis consists of four main parts. Firstly, the background reviews memory and related functions that are central to the area of learning and retention. The concept of retrieval practice is described, starting with an overview of the research field, continuing with suggested theoretical explanations, and ending with a discussion of what we know about student use of retrieval practice. Retrieval practice is also discussed in relation to individual differences in both cognitive and non-cognitive factors. Then, the aims of the thesis are presented. In the next section, the materials and methods used in the studies are described. Following that is a summary of the included studies. And finally, in the last section, the findings and their implications are discussed.

Background

Memory and learning

The human memory is a fascinating and complex phenomenon that has earned an undeservedly bad reputation. It is certainly more common to hear people complain about their memory, rather than praise it. However, while our memory is prone to more or less momentary failures, it is also quite remarkable in its capacity. A single hint of a particular scent can instantly transport us back to an event that took place many years ago and that was completely "forgotten" before that instant. But, of course, it was not really forgotten, just stored away and not brought to mind for a long time.

When memory is mentioned in a scientific context, a few different mechanisms may be intended. For example, one aspect concerns how memory is structured in the brain, another concerns different types of memories (i.e., explicit/implicit memories), and a final aspect concerns different memory processes (i.e., encoding, storage, retrieval). These are explained in turn below.

There are different theories about how our memory is structured, but one of the most classic theories is the Multi-Store Memory Model by Atkinson and Shiffrin (1968). This model claims that the human memory system is comprised of three separate stores with different characteristics. The first module in the chain is sensory memory, which keeps information from the senses stored for a few milliseconds. If this information is attended to, it can be transferred to the next module, which is short-term memory. Short-term memory is usually said to store 7 ± 2 items of information for a few seconds (Miller, 1956). If the information is rehearsed, for example by repeating a phone number until you find a pen, the duration of such a short-term memory can increase. Then, information that is rehearsed and understood can be transferred to long-term memory, which has an unlimited capacity, both in terms of the amount of information that can be stored and the duration of the memories (Atkinson & Shiffrin, 1968).

Another highly important memory function in the area of learning, which is related to short-term memory, is working memory (Baddeley & Hitch, 1974). The original working memory model consisted of three components: the central executive, the phonological loop, and the visuospatial sketchpad (Baddeley & Hitch, 1974). The central executive is

described as the attentional controller, and it divides attentional resources to ensure that tasks are executed properly. The phonological loop is responsible for maintaining auditory information and speech, including both internal and external speech, while the visuospatial sketchpad is responsible for maintaining visual and spatial information. These are considered to be subsystems of the central executive (Baddeley, 2000). A few decades later, Baddeley (2000) updated the model to also include a component called the episodic buffer, which allows interaction between working memory and long-term memory (Baddeley, 2000). The episodic buffer holds information in a multidimensional code which allows episodes/chunks from different dimensions to be combined and stored in the buffer. These episodes or chunks can be derived from different sources, for example working memory, perception, and long-term memory (Baddeley, 2000). Working memory capacity can influence our ability to learn. For example, individuals need to be able to keep information active in working memory in order to form new concepts, such as linking two facts together. In addition, the control process of working memory is essential for staying on task, which of course is important in educational settings (Cowan, 2014).

When talking about memories in everyday settings, we usually refer to facts we have learned, or events we have experienced, and stored in our long-term memory. This type of memory is called explicit memory and is available for us to consciously recall. Explicit memory is further divided into episodic memory, which consists of "episodes" or experiences from our lives, and semantic memory, which consists of knowledge and facts (Tulving, 1993). There is also a second type of memory called implicit memory, which mainly consist of unconscious memories. Examples of unconscious memories are procedural memories, for example motor functions such as learning to walk, and previous experiences that (unconsciously) affect our behaviors (Tulving, 1993). In the present thesis, the learning and retention of memories mainly concerns semantic memory.

How are memories formed and stored?

Long-term memory is said to be unlimited, but we are all aware that it is not a given that our memories will be stored forever. Moreover, a small and insignificant detail might be remembered without trying while important information may be completely forgotten despite the intention of remembering it. Learning, or forming memories, depends on three processes: encoding, storage, and retrieval (McDermott & Roediger, 2014). Encoding occurs when we first experience something or attend to

a piece of information. Different factors can affect the initial encoding, and this will in turn affect the consolidation and storage of the memory. After encoding, and over time, memory traces of the experience are created through neural changes in the brain, a process termed consolidation (McDermott & Roediger, 2014). The memory is then stored, in theory, indefinitely. However, memories are forgotten if they are not brought to mind, and both prior and subsequent experiences will interfere with the integrity of our memories (McDermott & Roediger, 2014). This explains why memories are sometimes false or altered. The retrieval process, which brings stored memories to mind, is crucial because this step gives us access to the information we need, as well as strengthening memories and making them available for longer (Karpicke & Roediger III, 2008). Fortunately, this process can be improved by practicing retrieval.

Retrieval practice

Retrieval practice, test-enhanced learning, and test-based learning are all names referring to the same phenomenon. They all refer to the experimental finding that learning which uses strategies requiring the repeated retrieval of information from long-term memory leads to superior long-term retention when compared to other learning strategies, the most common of the latter being re-study (i.e. repeatedly reading the information; Roediger & Karpicke, 2006b). In this thesis, the terms retrieval practice and re-study are used to refer to these contrasting strategies. The term *learning strategy* is used to signify operations and actions used for learning; however, the terms *learning method* and *learning technique* are used in two of the included studies, and *study strategy* is another term found within the research field; these are used interchangeably here.

The enhanced retention seen after practice that incorporates a repeated retrieval component is called the *testing effect*. However, it is important to note that this effect is not exclusive to tests. In fact, it can be elicited by a multitude of different types of learning activities, as long as they incorporate the key component — retrieval from long-term memory.

Overview of the research field

The fact that being tested on a material enhances the memory of that material is not a novel finding. Actually, this idea was studied more than a hundred years ago by Abott (1909), Gates (1917), and Thorndike (1914), and sporadically since then until interest in the phenomenon picked up

again in the early 2000s, resulting in several thousands of research studies to date.

Initially, the research was mainly focused on experimental studies investigating the testing effect across different circumstances, such as different types of information, retrieval practice activities, response formats, and with or without feedback. In fact, studies have demonstrated testing effects on information ranging from word pairs and paired associates (Karpicke & Roediger, 2007), psychological concepts (Wiklund-Hörnqvist et al., 2014), historical facts (Carpenter et al., 2009), prose materials (Butler, 2010; Roediger & Karpicke, 2006a), visuospatial information (Carpenter & Pashler, 2007; Kang, 2010), and bilinear functions (Kang et al., 2011), to complex materials (for a review, see Karpicke & Aue, 2015). Regarding types of retrieval activities, studies have included testing with multiple choice, short answer, and free recall questions (Einstein et al., 2012), and concept mapping (Blunt & Karpicke, 2014). Studies have also compared some of these formats with each other in terms of effectiveness, predominantly showing larger effects from response formats that require more effort (i.e., short answer and free recall; see Carpenter & Delosh, 2006; Kang et al., 2007; McDaniel et al., 2007; Stenlund et al., 2016).

Furthermore, the testing effect does not appear to be a phenomenon that is sensitive to the age of the person as it has been demonstrated in preschool (Fritz et al., 2007), and elementary-school children (Goossens et al., 2016; Karpicke et al., 2016; Lipko-Speed et al., 2014; for a review focusing on younger children see Fazio & Marsh, 2019;), upper-secondary school students (Jonsson et al., 2021; McDermott et al., 2014; Stenlund et al., 2017) and university students (McDaniel et al., 2007; Wiklund-Hörnqvist et al., 2014), as well as in older adults (Coane, 2013; Meyer & Logan, 2013). Another aspect that has received a lot of interest is the use of feedback during retrieval practice. In fact, a classic finding is that, when feedback is not given during retrieval practice, performance may actually be higher for material learned through re-study when tested immediately after learning (Roediger & Karpicke, 2006a; Wheeler et al., 2003), but the re-studied material is then forgotten at a higher rate than the repeatedly tested material. This pattern is generally reversed when feedback is included, and correct-answer feedback can protect from erroneous learning and enhance the effect of retrieval practice (Butler & Roediger, 2008; Pashler et al., 2005). Finally, substantial efforts have also been made to elucidate the neurocognitive processes related to retrieval practice (see van den Broek et al., 2016, for a review).

In order to gain a better understanding of how retrieval practice is best utilized, efforts have been made to determine how many times information must be retrieved from long-term memory, and what the timing of the practice sessions should be, to achieve durable learning. It is common for students to report that, while studying, material is dropped from further study as soon as they feel as though they have learnt it (Kornell & Bjork, 2007). However, research has found that continuing to practice until the material has been retrieved several times during the first learning session improves recall over time (Rawson & Dunlosky, 2011). The problem is knowing when one has practiced enough because too much practice takes time away from practicing other material. Rawson and Dunlosky (2011) suggest practicing until all items have been correctly recalled three times in the first learning session, and then relearning the material in three subsequent sessions. This means that initial learning takes more time and effort, but this is compensated by easier relearning in later sessions.

The suggestion of several separate learning sessions incorporates another powerful strategy — spacing. The spacing effect refers to the finding that learning sessions become more effective when they are spaced apart in time, than when the same amount of learning is conducted closer together in time (Carpenter et al., 2022). Retrieval practice becomes even more effective when combined with spacing, but again, knowing how much time should pass between sessions is not straightforward. The optimal spacing schedule will involve balancing the amount of practice in each learning sessions and time between sessions, as well as taking into account the familiarity of the material, in order to optimize learning and memory retention (see Carpenter et al., 2022, for a review of the combination of spacing and retrieval).

Over time, the emphasis has shifted from experimental studies of retrieval practice to applied research in educational settings. This is an important development to ensure that retrieval practice can be recommended for use in schools. While the testing effect has been well-documented in laboratory settings, it cannot be assumed that it translates to the more complex classroom setting. A number of reviews and meta-analyses focusing on “applied” or “classroom” research have emerged over the last few years (Agarwal et al., 2021; Lamotte et al., 2021; Moreira et al., 2019; Schwieren et al., 2017; Sotola & Credé, 2021; Yang et al., 2021), all of which have come to the conclusion that the benefits of retrieval practice do indeed transfer to the classroom. However, the methodological designs used in classroom studies vary significantly, and it is difficult to discern the ecological validity of studies carried out in classrooms. Examples of

classroom studies range from less applied, where the material was relevant to the course/subject but the intervention did not include exams/course grades (Wiklund-Hörnqvist et al., 2014), to studies that are more applied in the sense that the retrieval practice activity could affect exam scores or course grades but was in fact performed outside of the classroom (Daniel & Broida, 2004; B. C. Johnson & Kiviniemi, 2009), to designs where the intervention was fully integrated as an in-class activity (Agarwal, 2019; Batsell et al., 2017; E. L. Bjork et al., 2014; Freda & Lipp, 2016; McDaniel et al., 2013). As a specific example, Batsell et al. (2017) examined the impact of daily quizzes in an introductory psychology course. Two classes were included in the study, one served as the control condition and were given daily readings, while the experimental group were given a five-question quiz each day. Performance was then measured on three exams that included questions that were identical, similar, or new but related to the quiz questions. The experimental class performed better than the control class on all three types of question (i.e., a testing effect).

Although the research is generally in agreement that retrieval practice is an effective learning strategy, there is no single established theoretical explanation for the effect. However, there appears to be a general consensus that the process of retrieving information from long-term memory alters the memory or memory trace so that subsequent retrieval attempts become easier (Putnam et al., 2016). Exactly what this change consists of, and how it makes subsequent retrieval easier, is the basis of many of the proposed theories. These theories are presented in brief below.

Theoretical explanations for the testing effect

Retrieval effort hypothesis

The retrieval effort hypothesis states that effortful and successful retrieval enhances memory more than successful but less effortful retrieval does (Pyc & Rawson, 2009). This hypothesis is based on the general principles of the desirable difficulties framework proposed by Bjork (1994), which suggests that processing that is difficult but successful is better for memory than processing that is successful but less difficult. Examples of activities that can introduce desirable difficulties are, for example, spacing, testing, and varying the learning conditions. However, these activities can prove to introduce undesirable difficulties depending on the student's capabilities; for example, if the task is too difficult for an individual, the retrieval/processing will not be successful, and the memory will not be strengthened. Therefore, it is important to

differentiate between desirable and undesirable difficulties in this context (Bjork & Bjork, 2011).

The bifurcation model

The bifurcation model proposes an explanation for the classic finding that re-study may result in better performance than retrieval practice in retention tests conducted immediately after the learning session, but that the reverse is seen in later retention tests. The model is based on the idea that retrieval practice without feedback produces a bifurcated, or divided, item distribution. That is, memory traces for items, for example words, that are successfully retrieved are greatly strengthened, while items that are not retrieved are not strengthened at all. In contrast, items in the re-study condition are all strengthened, but to a lesser extent than the retrieved items (Kornell et al., 2011). Assuming that successfully retrieved and restudied items are strengthened to reach above some recall threshold, and that the rate of forgetting is identical for all items, re-study may result in better performance than retrieval practice on an immediate test. Over time, retrieval practice items will remain above the recall threshold for longer, while restudied items will be forgotten (Kornell et al., 2011).

Transfer-appropriate processing

Transfer-appropriate processing is the general idea that memory performance is influenced by how well the processing used during learning is related to the way in which the material will be tested or used. For example, a nursing student preparing for a practical exam in wound suturing should practice actual suturing, not only read about it in a textbook. Both the quality and the strength of the memory traces are contingent on the suitability of the testing situation (Morris et al., 1977). This theory was proposed as an explanation for the testing effect because the processes required when being repeatedly tested on a material match the processes used in a final memory test more than the processes of repeated study.

Elaborative retrieval hypothesis/Mediator effectiveness hypothesis

The elaborative retrieval hypothesis and the mediator effectiveness hypothesis both propose that retrieval practice produces better associations between cue and target than re-study does, and thus retrieval practice leads to better retention. In this case, a cue refers to a hint or prompt that initiates the retrieval of a specific memory – the target. The elaborative retrieval hypothesis claims that retrieval practice is superior to re-study because, when retrieval is attempted, elaboration produces multiple pathways between the cue and target (Carpenter, 2009;

Carpenter & Delosh, 2006). For example, Carpenter (2009) suggested that learning weakly related word pairs, such as *Basket—Bread,* by practicing retrieval may activate additional information (i.e., *Basket—Eggs—Flour—Bread)* and that this elaborative structure benefits future retention. However, such a structure is less likely to occur during re-study because cue and target are presented together. In a later study, Carpenter (2011) further developed this work by specifying that the additional information activated by retrieval may be a word or a concept, and proposed a new hypothesis called the *semantic mediator hypothesis*.

The mediator effectiveness hypothesis states that retrieval practice enhances retention because it results in more effective mediators between cue and target (Pyc & Rawson, 2010). These mediators can be words, phrases, or concepts, and are produced during encoding. Thus, both re-study and retrieval practice generate mediators, but the hypothesis assumes that those generated during retrieval practice lead to better recall (Pyc & Rawson, 2010).

Episodic context account
The episodic context account emphasizes the role of the temporal context of a learning situation for memory retention. During memory encoding, the information becomes associated with the context of the learning situation. The context then acts as a retrieval cue for that information. When the information is later retrieved, the context of the learning situation is reestablished and updated to include the context at the time of retrieval. When multiple contexts are associated with a specific piece of information, each of them can serve as a memory cue (Karpicke et al., 2014; Lehman et al., 2014).

The dual memory theoretical framework
A recently suggested theoretical framework proposes that retrieval practice creates two separate memories of the practiced material—a study memory from the initial study session, and a separate test memory from the initial testing. Thus, material that has only been practiced through study is only supported by the study memory, while material that has been learned through testing can be accessed by either study memory, test memory, or both (Rickard & Pan, 2018).

Empirical evidence for the theoretical explanations
Efforts have been made to find empirical evidence for the different theoretical explanations. It could be argued that the *retrieval effort hypothesis* is the best supported theory because much of the research regarding the mechanisms of the testing effect inadvertently manipulate

retrieval effort; for example, when comparing response formats (Kang et al., 2007; see Rowland, 2014, for a meta-analysis). Examples of studies that have explicitly put this theory to the test include experiments by Pyc and Rawson (2009), where retrieval difficulty was altered by varying the interstimulus interval (i.e. "the number of items between each next practice trial with any given item" (Pyc & Rawson, 2009, p. 438)) and the number of correct retrievals required. The results showed that difficult but successful retrieval increased retention on a final test compared to easy and successful retrieval.

The findings in relation to *transfer-appropriate processing* as an explanation for the testing effect are rather mixed. While some studies find the expected benefit of matching test formats between initial and final tests (Duchastel & Nungester, 1982; Veltre et al., 2015), others do not find such effects (Carpenter & Delosh, 2006; Kang et al., 2007; Stenlund et al., 2016). Interestingly, two reviews have come to different conclusions, with Adesope et al. (2017) reporting that matching practice and final test format result in stronger testing effects, while Rowland (2014) did not find an increase in the testing effect when initial and final test formats matched.

Support for the *mediator effectiveness hypothesis* was found by Pyc and Rawson (2010) when participants were asked to generate mediators while learning Swahili–English word pairs through test—re-study (i.e. a cued recall test with immediate re-study after each trial) or re-study only. The study showed that retrieval practice facilitates both the retrieval and decoding of mediators, which in turn improves recall of the test material. Carpenter (2011) further supported the mediator effectiveness hypothesis in her study investigating the semantic mediator hypothesis, where the participants were not required to generate mediators. Carpenter (2011) proposes that this is a more typical example of the testing effect, and an important contribution which shows that mediators can be activated spontaneously, not only upon request (but see Cho et al., 2017, for contrasting results). However, a number of studies have failed to find support for these hypotheses (Karpicke & Smith, 2012; Lehman et al., 2014; Lehman & Karpicke, 2016), or find fixed results (Coppens et al., 2016).

The *episodic context account* was investigated by comparing the level of recall and intrusions of previous items in three different conditions: control, retrieval practice, and elaboration (Lehman et al., 2014). The results showed that participants engaging under the retrieval practice condition achieved higher correct recall and made fewer prior-list

intrusions. This suggests that retrieval practice alters the representation of the episodic context, which helps people to restrict the search set and enhances correct recall. In contrast, Hong et al. (2019) did not find evidence for context-related details being strengthened by retrieval practice.

Use of retrieval practice

An important aspect to consider when investigating learning strategies is how they work both inside and outside of the classroom. As discussed earlier in this section, the effects of retrieval practice appear to transfer well from lab-based research to classroom research (Agarwal et al., 2021). However, less is known about how students use retrieval practice outside of the classroom—when they self-regulate their learning and choose their own learning strategies. In relation to retrieval practice as an effective learning strategy, self-regulated use of the strategy has been the objective of a few studies, using both behavioral methods and self-report measures.

In a study examining the use of retrieval practice among middle- and high-school students, Tullis and Maddox (2020) found that high-school students reported using retrieval practice to a greater extent than middle-school students did, but both age groups used re-study more than retrieval practice (see Dirkx et al., 2019, for similar findings). In a review by Rivers (2021) investigating students' metacognition about practice testing, the results showed that students tend to choose testing when the material to be studied is easier compared to more difficult material, and when feedback on their performance is given. The results also suggested that students use self-testing in order to monitor their learning to a greater extent than for learning purposes (Rivers, 2021). In behavioral studies, a common finding is that students tend to use optional retrieval practice activities to only a low extent (Corral et al., 2020; B. C. Johnson & Kiviniemi, 2009; G. M. Johnson, 2006), but that that students who do use such activities perform better on subsequent exams. With the effectiveness of retrieval practice in mind, it is crucial that students self-regulate their learning and use the strategy on a voluntary basis. In the following section, the framework of self-regulated learning is briefly introduced and related to the use of retrieval practice as a learning strategy.

Self-regulated learning

Self-regulated learning refers to the ability to assess and manage one's own learning. This is a very important ability to possess because it is related to academic achievement (Richardson et al., 2012; Schneider &

Preckel, 2017), and it is also an important tool throughout life (R. A. Bjork et al., 2013). Zimmerman (2002) has described the self-regulatory process as consisting of three cyclical phases: the forethought phase, the performance phase, and the self-reflection phase. The forethought phase consists of analyzing the task at hand, for example by setting learning goals and planning what strategies to use. The performance phase includes self-control, that is, putting the selected strategies to use and maintaining focus. The self-reflection phase includes self-judgement, which refers to an evaluation of one's performance. This self-reflection can then influence the forethought phase in future study sessions (Zimmerman, 2002).

Unfortunately, it is rare for students to be taught to regulate their own learning (R. A. Bjork et al., 2013; Zimmerman, 2002) and there seems to be a general belief that people do not need to be taught how to manage their learning. However, most people have difficulties regulating their learning, both when it comes to choosing appropriate activities for learning, and in judging how well they have learnt something. Studies show that, overall, students use effective learning strategies to only a low extent and rely more on ineffective strategies, such as re-reading (Dunlosky et al., 2013). This suggests that most students do in fact need to be instructed on how to become an efficient learner. Furthermore, some studies have reported that female students appear to be better at regulating their learning than male students (Pérez et al., 2017; Weis et al., 2013). For example, Marrs and Sigler (2012) found that female students in their study displayed approaches to learning characterized by deeper learning and higher achievement, and engaged in study behaviors related to achievement to a greater extent than male students (see Ruffing et al., 2015 for similar findings). In addition to sex-related differences, individual differences in both cognitive and non-cognitive factors have been linked to differences in self-regulated learning (Pérez-González et al., 2022; Richardson et al., 2012). Hence, it is important to consider how individual differences may be related to the use of retrieval practice.

Individual differences and retrieval practice

Individuals' aptitudes for learning and, in turn, their ability to succeed in educational endeavors, are influenced by several different factors. Cognitive factors, such as working memory capacity (WMC; Ahmed et al., 2019; Alloway & Alloway, 2010) and intelligence (Downey et al., 2014; Furnham et al., 2009; Laidra et al., 2007; Roth et al., 2015), are commonly identified as predictors of academic achievement. Intelligence has been

defined as a "general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience" (Gottfredson, 1997, p. 13). In this thesis, a specific intelligence factor called fluid intelligence is of primary interest because of its relationship to academic achievement (Laidra et al., 2007). Fluid intelligence concerns the aspects of intelligence that involve cognitive processes, such as solving new problems and abstract reasoning, that are not dependent on acquired knowledge. Stored knowledge and learned procedures are instead considered to be part of crystalized intelligence (Nisbett et al., 2012).

However, cognition is not the only important factor. Non-cognitive factors, such as personality traits, self-regulation, and motivational factors also influence academic aptitude. It has been suggested that cognitive factors reflect an individual's maximum performance while personality instead reflects an individual's typical performance (Chamorro-Premuzic et al., 2006; Goff & Ackerman, 1992). Given that academic performance is influenced by both cognitive and non-cognitive aspects, two individuals with the same level of intelligence but with different levels of interest in or propensities towards educational tasks, will perform differently. The question of how individual differences affect retrieval practice has been insufficiently explored. The present thesis examines how cognitive and non-cognitive factors are related to the testing effect, and to the use of retrieval practice as an optional learning strategy. As the first question concerns the enhanced learning observed after retrieval practice compared to re-study, and the latter rather concerns how individuals choose to study, it seems plausible that individual differences in cognitive and/or non-cognitive factors are related to these outcomes in different ways.

Cognitive factors and retrieval practice

Relatively few studies have examined the relationships between cognitive abilities and the testing effect, and the findings are mixed. Factors such as episodic memory, working memory capacity (WMC), processing speed, and different aspects of cognitive ability have been investigated.

In a study including both behavioral and brain-imaging data, measures of several abilities (i.e., fluid intelligence, WMC, short-term memory, episodic memory, and updating) were combined to create a composite score of cognitive ability. The results showed that the testing effect was not dependent on cognitive ability (Jonsson et al., 2021). Brewer and Unsworth (2012) also included a large test battery and found that episodic

memory and general fluid intelligence were negative predictors of the testing effect, which suggests that participants with poorer episodic memory, and those with lower fluid intelligence, benefit more from retrieval practice than participants with high cognitive ability. However, in a replication study focusing on the relationship between the testing effect and episodic memory, Pan et al. (2015) did not find a larger benefit among participants of low ability. Following up on these conflicting results, Robey (2019) combined newly collected data with data from Brewer and Unsworth (2012) and Pan et al. (2015), and ultimately did not find any evidence that the testing effect was related to individual differences in episodic memory or fluid intelligence.

When investigating whether the testing effect differs between age groups, Meyer and Logan (2013) found that older and younger participants benefited comparably from retrieval practice, and that the effect was unrelated to intelligence (as measured using the vocabulary and reasoning subtests of the Wechsler Abbreviated Scale of Intelligence). Also exploring the testing effect for younger students, Karpicke et al. (2016) found that the effect was unrelated to differences in processing speed in elementary school children. Minear et al. (2018) investigated intelligence and working memory in relation to retrieval practice. The results did not show any significant relationships with the overall testing effect; however, when item difficulty (determined by recall accuracy in earlier experiments) was taken into account, fluid intelligence was a negative predictor for easy items, and a positive predictor for difficult items (Minear et al., 2018). This indicates that participants with lower fluid intelligence experienced a larger testing effect for easy items than for difficult items, and participants with higher fluid intelligence experienced a larger testing effect for difficult items than for easy items (see Wenzel & Reinhard, 2019, for similar findings regarding highly intelligent participants).

Several additional studies have examined the relationship between the testing effect and working memory capacity, and the results are mixed. While some studies have been unable to find any indications that working memory capacity is related to the testing effect (Brewer & Unsworth, 2012; Tse et al., 2019; Wiklund-Hörnqvist et al., 2014), other studies report lesser (Tse & Pu, 2012), and greater (Agarwal et al., 2017; Yang et al., 2020) benefit for low-WMC participants. Still other studies report greater benefit for high-WMC participants under conditions where the demands on working memory are high (Zheng et al., 2023). Moreover, a study exploring the relationship between the testing effect and academic achievement found that high-performing students benefited more from retrieval practice than from copying definitions, whereas the opposite was

found for low-performing students, and middle-performing students benefited similarly from both methods (Carpenter et al., 2016). To summarize, some studies indicate that individual differences in cognitive factors are related to the magnitude of the testing effect, whereas other studies report no such relationship. Considering these inconsistent results regarding the impact of cognitive factors on the testing effect, more research is warranted.

Another aspect that it is important to consider, and which has attracted even less interest to date, is the relationship between cognitive abilities and how students use retrieval practice in their own studying. Fellman, Lincke, and Jonsson (2020) investigated how medical students interacted with a massive open online course (MOOC) platform in which quizzing was an optional activity. The results revealed that fluid intelligence (as measured using Ravens Advanced Progressive Matrices) was a positive predictor of quiz use, indicating that high intelligence is related to more frequent use of optional quizzes. In a similar study investigating student engagement with a MOOC platform, Fellman, Lincke, Berge, et al. (2020) found that quiz use was positively related to verbal working memory, again suggesting that high levels of cognitive ability are related to more frequent quiz use. Again, more research on the subject is necessary to be able to draw any conclusions about the relationships between cognitive abilities and optional use of retrieval practice.

Non-cognitive factors and retrieval practice

As mentioned previously in this section, a set of non-cognitive factors that is important for educational achievement and self-regulated learning are personality characteristics. In relation to retrieval practice, it is possible that certain personality characteristics may have an impact on both how students act in specific learning situations and how they organize their learning over time.

In this thesis, the relationship between personality and the use of retrieval practice is investigated, and in this work a trait approach to personality has been adopted. A comprehensive, empirically supported model within the trait approach is the five-factor model of personality (FFM) (McCrae & John, 1992). As the name suggests, the model proposes that personality can be captured by five factors: extroversion, agreeableness, conscientiousness, neuroticism, and openness to experience (henceforth referred to as “openness”; McCrae & John, 1992). People scoring high on measures of these factors can be described as having a tendency to be: extraverted, talkative, and assertive (extroversion); kind, cooperative, and

sympathetic (agreeableness); organized, systematic, and thorough (conscientiousness); anxious, moody, and temperamental (neuroticism); intellectual, creative, and imaginative (openness; Goldberg, 1992).

Two of these factors, conscientiousness and openness, are of particular interest due to their connections to learning and achievement. Conscientiousness has repeatedly been found to be positively associated with academic achievement (Busato et al., 2000; Hakimi et al., 2011; O'Connor & Paunonen, 2007; Poropat, 2009, 2014; Sorić et al., 2017), to be as strong a predictor of educational achievement as intelligence (Poropat, 2009), and to be related to the use of self-regulated learning strategies (Bidjerano & Dai, 2007). Openness has also been found to be positively related to achievement, but to a slightly lesser extent and with mixed findings (Bidjerano & Dai, 2007; O'Connor & Paunonen, 2007; Poropat, 2009, 2014). However, to the best of my knowledge, these factors have not been investigated in relation to retrieval practice. To date, the only personality characteristics that have been included in retrieval practice studies are Need for Cognition (NFC) and grit.

Grit and NFC share similarities with conscientiousness and openness. Grit refers to a person's tendency to persevere and have a passion for long-term goals (Duckworth et al., 2007) and includes two sub-dimensions: *perseverance of effort* and *consistency of interest*. Grit is strongly and positively associated with conscientiousness (see Credé et al., 2017, for a review; Duckworth et al., 2007; Duckworth & Quinn, 2009). This strong association has caused some criticism from those who argue that grit is simply another measure of conscientiousness, and that the dimension of perseverance of effort has better criterion validity and explains more variance in achievement than conscientiousness, and is therefore a more promising measure (Credé et al., 2017). However, others argue that, because the definition of grit encompasses perseverance and passion for long-term goals, both dimensions should be included when the purpose is to measure grit (Duckworth et al., 2021). The second measure, NFC, refers to a person's inclination toward engaging in and enjoying thinking (Cacioppo & Petty, 1982) and has been found to be a positive predictor of academic achievement in both secondary-school students (Colling et al., 2022; Luong et al., 2017) and university students (Grass et al., 2017; Richardson et al., 2012; Strobel et al., 2019). Furthermore, studies have also reported that NFC is positively related to openness (Fleischhauer et al., 2010) and intelligence (Fleischhauer et al., 2010; Hill et al., 2013; von Stumm & Ackerman, 2013).

As previously discussed in this section, the rationale for why non-cognitive factors could be expected to impact upon the effectiveness of retrieval practice is related to the effort put into learning sessions. Based on the abovementioned associations between conscientiousness, openness, NFC, grit, and academic achievement, it does not seem unreasonable to suggest that these constructs may also be related to the learning effects of retrieval practice. Moreover, it is possible that these constructs affect the learning strategies students use, and how they engage in different learning tasks. Thus, relationships between these personality characteristics and the use of retrieval practice in authentic school settings seem plausible.

Research investigating the relationships between non-cognitive aspects, such as personality characteristics, and the testing effect is even more limited than that for cognitive aspects. The relationship between effects of retrieval practice and NFC and grit have been examined in a few studies, and these indicated no such relationships (Stenlund et al., 2017; Wiklund-Hörnqvist et al., 2022). To the best of my knowledge, only one study has been conducted examining the self-regulated use of retrieval practice. Fellman, Lincke, and Jonsson (2020) did not find that use of optional online quizzes was related to individual differences in grit or NFC. This short list of studies indicates that the individual differences research related to personality and retrieval practice deserves more attention.

Aim of the thesis

The overarching aim of this thesis was to investigate the relationships between retrieval practice and individual differences in cognitive and non-cognitive factors related to learning and academic achievement. This was explored in two separate aims. The first was to investigate whether such individual differences are related to the testing effect (i.e., the increased retention observed after repeated retrieval compared to re-study). The second was to implement retrieval practice in an authentic school environment in order to examine students' self-regulated use of retrieval practice as an optional learning strategy, and whether such use is related to individual differences. The purpose of these investigations was to contribute information that could help to determine how retrieval practice should be recommended for use in education.

Study I

In study I, the aim was to examine the testing effect in relation to individual differences in cognitive (WMC) and non-cognitive (Grit, NFC) factors.

Study II

Study II was a replication of the second experiment from Study I. The aim was to provide further support for the results of Study I by utilizing a larger sample.

Study III

In Study III, the overall aim was to implement retrieval practice in an authentic school environment in order to investigate how students use the method when it is provided in the classroom, and as an optional study activity, and whether individual differences in cognitive factors (WMC, fluid intelligence), non-cognitive factors (NFC, Grit, Conscientiousness, and Openness to experience), or sex-related differences are related to optional, self-regulated use of retrieval practice.

Materials and Methods

The measurement of psychological constructs

As mentioned in the introduction, the measurement of constructs such as memory, personality, and intelligence implies operationalizing these constructs into observable behaviors. The use of different instruments, for example, self-report instruments or psychological tests, to measure these constructs provides measurements that vary in reliability and validity. The construct validity of a measure refers to "the degree to which a measure actually measures the latent construct it is intended to measure" (Netemeyer et al., 2003, p. 8), while its reliability concerns "the consistency of measurements: from time to time, from form to form, from item to item, or from one rater to another" (Mueller & Knapp, 2018, p. 367). So, how valid a measure of intelligence is will vary depending on how well the measurement used captures the construct. The reliability of a measurement can be affected both by factors inherent in the test construction and by factors associated with the test environment and the test person (e.g., test anxiety, motivation). The validity and reliability of the measures used in this thesis are presented below and their implications for the conclusions about the results are discussed in the Discussion section.

Measures

Working Memory Capacity (WMC)

WMC was assessed using the Automated Operation Span Task (AOSPAN; Unsworth et al., 2005), a computerized version of the original Operation Span Task, which measures participants' ability to hold letters in their working memory while solving simple math problems. Presentations of math problems and letters are alternated during a series of trials. Firstly, participants are confronted with a math problem and, after solving the problem, they click to go to the next page where they must judge whether the number displayed represents the solution to the problem. After that, a letter is presented. The trials range from three up to seven math problems and letters, and each trial size is repeated three times, giving a total of 75 math problems and 75 letters to memorize. Each trial ends with a recall task in which participants are asked to indicate which letters they saw, in the correct order, from a matrix displaying 12 letters. To ensure that participants put effort into both math problems and letter

memorization, a lower level of 85% correct for the math problems is encouraged. The AOSPAN has been found to have good internal consistency and rest–retest reliability, as well as construct validity (Unsworth et al., 2005).

Fluid intelligence

Raven's Advanced Progressive Matrices (RAPM; Raven, 1990) is a non-verbal measure of abstract reasoning that is used to estimate participants' fluid intelligence. The items consist of geometric patterns with one piece of a 3 x 3 matrix missing. Items are solved by identifying the pattern and selecting the final piece from eight possible alternatives. The level of difficulty increases as the test progresses. The original version consists of two sets of items, with 12 items in set 1 and 36 items in set 2, and the first set is often used as practice items. In this thesis, half of the items were used, resulting in 6 practice items and 18 test items. The total number of correct answers represents the score, with a higher score indicating greater abstract reasoning ability. The RAPM has demonstrated good construct validity (Schweizer et al., 2007).

Need for Cognition

A Swedish adaptation of the Need for Cognition Scale (Cacioppo & Petty, 1982), the Mental Effort Tolerance Questionnaire (METQ; Dornic et al., 1991), was used to measure NFC. The METQ consists of 30 items rated on a 5-point scale from 1 (*Do not agree at all*) to 5 (*Agree completely*). Twelve items reflect positive attitudes (i.e., I like…) and 18 items reflect negative attitudes (i.e., I don't like…), thus, the negative items require reverse scoring, and high scores indicate a high need for cognition. The METQ has been found to have good internal consistency as well as validity both by the original researchers and in more recent studies (Dornic et al., 1991; Stenlund & Jonsson, 2017).

Grit

Grit was measured using GRIT-S, a short adaption of the original GRIT scale (Duckworth et al., 2007; Duckworth & Quinn, 2009). GRIT-S consists of eight items rated on a 5-point scale ranging from 1 (*Not like me at all*) to 5 (*Very much like me*). The scale contains two subscales: four items measure positive attitudes toward *perseverance of effort* (e.g., "I am a hard-working person") and four items measure negative attitudes toward the subscale *consistency of interest* (e.g., "New ideas and projects distract me from what I am doing"). The negatively phrased items are reverse scored so that higher scores on GRIT-S reflect higher levels of grit.

GRIT-S has been found to have acceptable reliability and validity (Duckworth & Quinn, 2009).

Conscientiousness and Openness to Experience

Conscientiousness and openness to experience were assessed using the mini International Personality Item Pool (Mini-IPIP; Donnellan et al., 2006), a 20-item questionnaire measuring the Five-Factor Model of personality. Each of the five factors: neuroticism, extraversion, conscientiousness, openness to experience, and agreeableness, is measured by four items. Of interest in this thesis are the factors conscientiousness (e.g., "Get chores done right away") and openness to experience (e.g., "Have difficulty understanding abstract ideas", negatively phrased). Responses are given on a 5-point scale from 1 (*Very inaccurate)* to 5 (*Very accurate)* and negatively phrased items are reversed during scoring such that high scores reflect higher levels of conscientiousness/openness to experience. The mini-IPIP has demonstrated acceptable internal consistency and construct validity (Donnellan et al., 2006).

Study design

Studies I & II

The experiments in Study I and Study II were designed as classic examples of experiments examining the testing effect. For Experiment I in Study I, a (quasi-experimental) between-subjects design and multiple-choice response format were used. Half of the participants, the control group, learned 30 Swedish–Swahili word pairs by repeatedly studying them, and the experimental group learned the word pairs using retrieval practice. Both groups were introduced to the learning material by reading a list of all word pairs for five minutes before the learning phase began. For the re-study group, the word pairs were then presented on a projector screen using a PowerPoint presentation. Each pair was displayed for four seconds and a total of five times. While the word pairs appeared on the projector screen, the participants confirmed that they had read each pair by pressing a button on individual handheld units belonging to an audience response system (ARS). For the retrieval practice group, the Swahili words were presented on the projector screen for four seconds each and a total of five times. After each item, four options of letters were displayed, and the participants were instructed to select the alternative that represented the second letter in the Swedish translation and press the corresponding button on their ARS unit. Finally, the correct answer was presented for

one second. Retention was then measured after a five-minute break, after one week, and after four weeks.

In Experiment II in Study I, and in Study II, within-subject designs were used. The intervention was conducted in a classroom on the participants' individual computers, and consisted of three phases: a baseline exposure, the learning phase, and three retention tests. During baseline exposure, the participants were introduced to the learning material, 60 Swedish–Swahili word pairs, by presenting each word pair on their computer screens, one at a time, for eight seconds each. In order to examine the effect of retrieval practice within each participant, during the learning phase, half of the learning material was presented as re-study (i.e., both words in the pair was presented), and half was presented as retrieval practice (i.e., the Swahili word was presented). For each participant, the word pairs were randomly assigned to the two conditions, creating unique lists of re-study and retrieval practice word pairs that were interleaved during learning. The participants were instructed to try to remember the words either by reading the presented word pair or by typing in the corresponding Swedish word. This was repeated until each word pair had been practiced six times. To ensure equal exposure for both conditions, re-study word pairs were presented for nine seconds, and for retrieval practice word pairs the Swahili word was shown for eight seconds, followed by one second of correct answer feedback.

After the learning phase, three retention tests were conducted at increasing intervals: five minutes, one week, and four weeks later. The test format was cued recall, identical to the retrieval practice condition during the learning phase except no feedback was given. To be able to measure differences in retention over time, the 60 word pairs were divided into three sets of 20, each including 10 re-study and 10 retrieval practice word pairs. One set of 20 word pairs was then introduced at each retention test, meaning that 20 word pairs were tested in the first retention test, an additional 20 were tested in the second retention test, and the total 60 word pairs were tested in the third retention test. See Figure 1 for a schematic of the procedure. Analyses were then conducted on both the accumulated items at each interval (20, 40, 60 items respectively), and the items that were uniquely tested at each interval (20, 20, 20 items respectively).

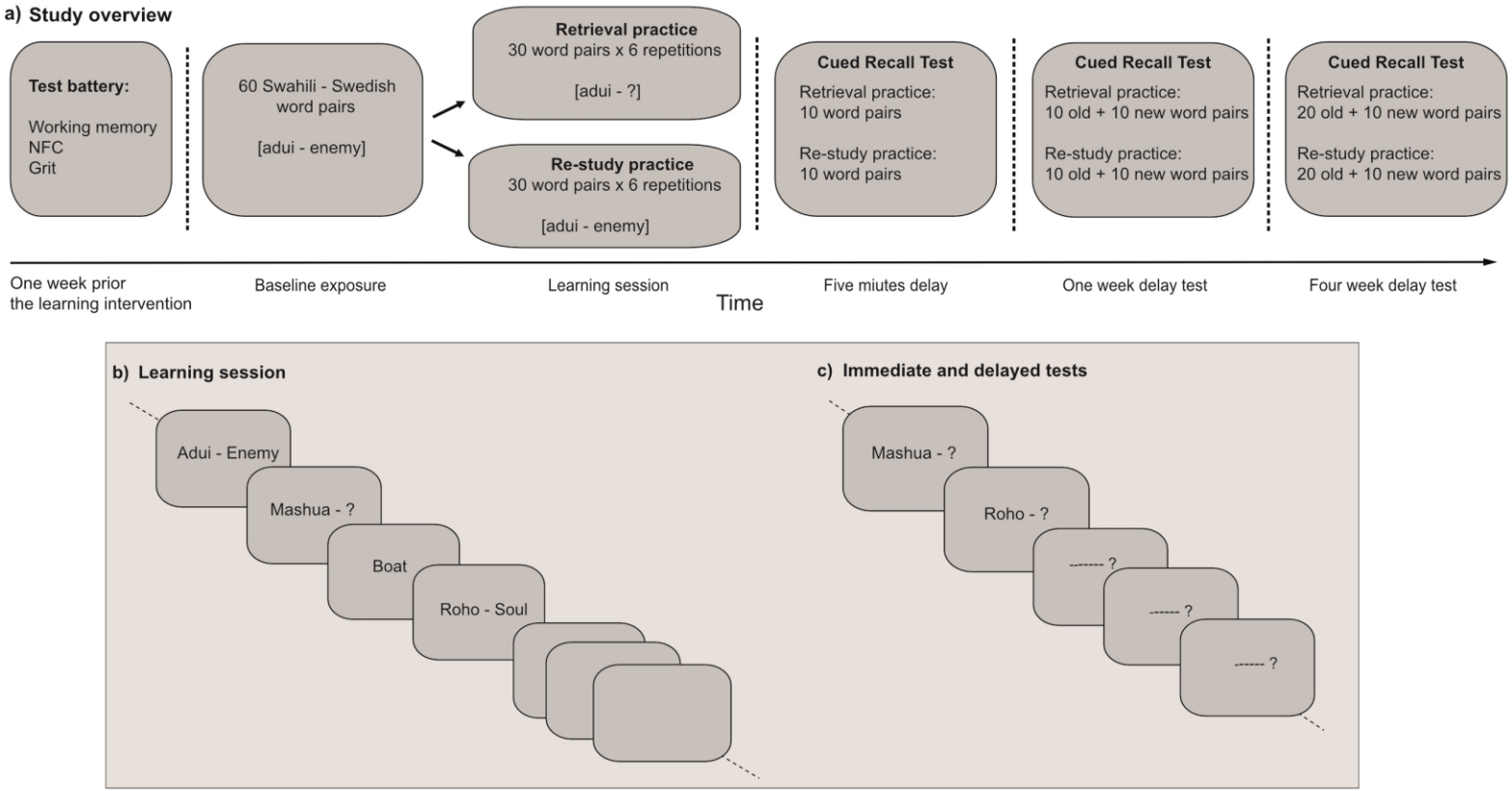

Figure 1. *A schematic overview of the study design for Study I, Experiment II, and Study II (from Bertilsson et al. (2021), slightly altered for clarity; reprinted under CC BY 4.0 license)*

Study III

The data analyzed in the third study was collected during an implementation of retrieval practice in an authentic educational setting. The overarching purpose of this implementation was to explore possible ways of utilizing retrieval practice in schools. The project was a collaboration between our research group and an upper-secondary school, which meant that four teachers were assigned to work part time on the project. To start off the project, the teachers were included in a workshop in which different types of retrieval practice tasks were discussed, and we arranged seminars for them about scientific perspectives and research design. The teachers' experience was of great value during the design process because it was important to create a procedure that was realistic in terms of workload and time in class for the teachers in order to ensure ecological validity. Two of the teachers taught mathematics and two taught Swedish, which meant that those were the subjects our interventions would focus on.

After trying out a couple of different options, two interventions based on online quizzes were created. In total, the intervention in mathematics was conducted in two cohorts, the first cohort spanning four semesters and the second cohort spanning two semesters. The intention was to run the same intervention in two cohorts in Swedish as well, but due to practical difficulties, we were only able to collect data from one cohort spanning two semesters. Before the intervention started, the students were given a lecture about retrieval practice that explained how testing can enhance learning and how one should test oneself in order to achieve enhanced retention. All the students also completed a large test battery measuring cognitive abilities and personality characteristics, a selection of which were included in Study III.

The online quizzes consisted of 20 questions in the form of definitions of mathematical or Swedish concepts. The questions were answered by typing in the corresponding concept. Correct answer feedback was given after each question. Each quiz corresponded to a chapter of the mathematics book, or a section of the Swedish course. The quizzes were available to the students on their learning platform during the weeks that they were working on the corresponding section in the classroom. The students were encouraged to use the quizzes as part of their own studying. Additionally, half of the quizzes were completed in the classroom on one occasion during each week that the quiz was available. This ABAB design was chosen to investigate differences in quiz use between quizzes that were completely optional and quizzes that were also completed in the

classroom. In the present study, data from the first two semesters, a total of eight quizzes from each course, was included. See Figure 2 for a schematic of the procedure.

To be able to measure how much of the material the students learned, each quiz was initiated with a pre-test and ended with a post-test containing the same questions as the quiz. This data was not analyzed in the present study.

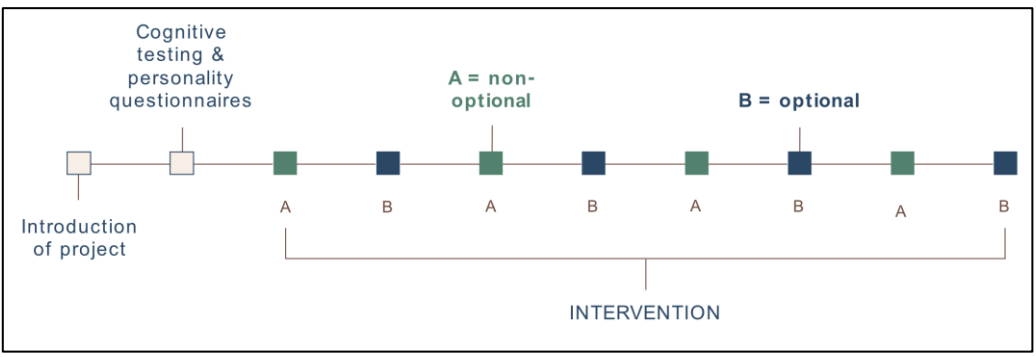

Figure 2. *A schematic overview of the study design for Study III.*

Learning materials

In order to study the testing effect and use of retrieval practice, different learning materials were included in the three studies. These materials are described in the following section.

Studies I & II

In the first two studies, the learning material consisted of Swedish–Swahili word pairs that were originally adapted from studies by Karpicke and Roediger (2008), Nelson and Dunlosky (1994), and Pyc and Rawson (2009), and were later translated into Swedish by Eriksson et al. (2011) and Karlsson Wirebring et al. (2015). The word pairs were learned using retrieval practice and/or re-study. Re-study word pairs were presented together for the participants to read. The retrieval practice conditions consisted of multiple-choice testing, where responses were given using an audience response system (Experiment I, Study I), or the Swahili word was presented, and the participants were asked to type in the corresponding Swedish word (Experiment II, Study I & study II).

Study III

In the third study, the learning materials, presented as online quizzes, consisted of mathematical terms and terms related to the Swedish language course. There were eight quizzes for each subject, each containing 20 questions. Answers were typed in and correct answer feedback was provided after each question. For the mathematics course, the material was based on the corresponding chapters in the course book (algebra & equations, geometry, etc.). The questions were adapted from definitions and the answers were the terms associated with each definition (What do we call a number that can be expressed as the ratio of two integers? = Rational number). For the Swedish course, the material was based on the subtopics included in the course (grammar, literature, etc.). As in mathematics, some questions were adapted from definitions (Which word class can be used to join two clauses? = Conjunction), and other questions asked for a synonym (What is another way of saying "reword"? = Paraphrase). In both subjects, the questions were constructed by senior teachers who were involved in the research project.

Statistical analyses

Analyses investigating the effects of retrieval practice

In both Study I and Study II, ANOVAs were used to investigate mean group differences in performance across the three time points. In Experiment I of Study I, performance differences across time were compared between two independent groups, using mixed between-within ANOVA (also called split-plot ANOVA). In Experiment II in Study I and in Study II, differences in performance across time were compared between two conditions including the same participants, using repeated measures ANOVA. Before the analyses were conducted, the relevant assumptions (i.e., outliers, normality, homogeneity) were controlled and deemed not to pose threats to the integrity of the analyses after identified outliers were removed. The assumption of sphericity was violated in Experiment II of Study I and in Study II; consequently, the degrees of freedom were Greenhouse-Geisser corrected.

In order to control for whether performance also differed between re-study and retrieval practice word pairs at each retention interval, paired *t*-tests were used in Study I and pairwise comparisons for the main effects were examined in Study II.

Analyses investigating relationships

Linear regression was used to examine the extent to which the cognitive and non-cognitive factors explained the variations in the dependent variables (i.e., performance, use of retrieval practice). Before conducting the analyses, the relevant assumptions (i.e., multivariate normality, linearity, autocorrelation, homoscedasticity, multicollinearity) were investigated and deemed not to pose threats to the integrity of the analyses after identified outliers were removed.

In the first experiment, linear regressions were conducted with grit and NFC as independent variables, and performance on the three retention tests for both groups as dependent variables. This resulted in a total of six analyses.

In the second experiment, hierarchical linear regressions were used. This method was selected to enable an exploration of the extent to which the independent variables explain the variation in the dependent variable, while controlling for the effects of other independent variables. Specifically, the purpose was to control for the possible effect of WMC on performance when investigating the effects of grit and NFC.

In Study II, the same hierarchical regression analyses were repeated, but the dependent variable was changed to better reflect the research question: Is the testing effect related to individual differences? Here, the difference in performance between re-study and retrieval practice word pairs at each retention was used, resulting in three regression analyses.

In Study III, hierarchical regression analyses were used to investigate the extent to which differences in optional use of retrieval practice could be explained by individual differences. To control for possible differences between males and females, sex was entered in the first step of the analysis, then the cognitive factors, fluid intelligence and WMC, were entered in the second step, and the non-cognitive factors were added in the third step.

Summary of studies

This chapter summarizes the three studies included in the present thesis.

Study I

Bertilsson, F., Wiklund-Hörnqvist, C., Stenlund, T., & Jonsson, B. (2017). The testing effect and its relation to working memory capacity and personality characteristics. *Journal of Cognitive Education and Psychology*, *16*(3), 241–259.

Aims and methods

The aim of the first study was to investigate the effectiveness of retrieval practice, that is the testing effect, and whether individual differences in cognitive factors and non-cognitive factors were related to any observed benefits of retrieval practice. This was achieved by performing two separate experiments, which had the further purpose of guiding the design of future studies on the same topic.

Experiment I

Experiment I included 39 upper-secondary school students (64.1% females, M_{age}= 17.74 years) who took part in an intervention where they would either learn 30 Swahili–Swedish word pairs using retrieval practice (experimental condition, n = 19) or repeated studying (control group, n = 20). The participants were randomly assigned to either condition.

Procedure
The stimulus material was projected onto a projector screen using Microsoft PowerPoint and participant responses were given and recorded using an audience response system (ARS) consisting of an individual ARS unit with response buttons for each participant.

The learning phase was initiated by a five-minute familiarization period, during which the participants were instructed to read though a list of the word pairs. Next, the word pairs were practiced for five rounds, in the re-study group by reading the two words presented side by side (e.g., Adhama–Heder [Adhama–Honor]) and in the retrieval practice group by means of a multiple-choice testing procedure (Adhama–?). The multiple-choice alternatives consisted of singular letters and the participants were

instructed to choose the alternative that corresponded to the second letter in the Swedish target word (i.e. the letter *e* for the word pair Adhama–Heder). Feedback was given by displaying the correct answer for one second after each trial.

Retention tests
Retention was measured at three time-points: five minutes after the learning phase, after one week, and after four weeks. The procedure for the retention tests was identical to the retrieval practice learning condition, but without the correct answer feedback.

Independent variables
Individual differences related to non-cognitive factors were assessed using the measures grit and NFC. Questionnaires measuring grit and NFC were filled out during the same session as the second retention test.

Statistical analyses
The main analyses consisted of a 3 × 2 mixed between-within ANOVA to investigate whether performance in the three retention tests differed between the two learning conditions, as well as regression analyses to evaluate the association between test performance and grit and NFC.

Results and discussion
There was a significant main effect of retention interval, as well as a significant interaction effect between retention interval and group. The interaction effect illustrates a classic finding in testing effect research; that is, that repeated studying can result in higher retention compared to retrieval practice on a test immediately after studying, but that the performance in the study group then drops at a faster rate over time. The between-subjects effect was nonsignificant, indicating that there was not a statistically significant difference in retention between the two learning conditions.

Moreover, there were no statistically significant associations between the independent variables, NFC and grit, and test performance for either condition.

Limitations
The results indicate that there were some limitations in the study design. Firstly, the lack of a testing effect could potentially be explained by a lack of statistical power, or the tendency toward a ceiling effect in terms of performance. Secondly, since the aim is to investigate how individual differences are related to retrieval practice and re-study, a within-subjects

design, where each individual studies under both conditions, may be a more suitable study design.

Experiment II

In an attempt to overcome some of the limitations of Experiment I, a second experiment was designed. Experiment II used a within-subjects design, the response format was changed from multiple-choice to short-answer, and the number of items was increased. In addition, a measure of WMC was added as an independent variable. Participants were 29 upper-secondary school students (41,4% females, M_{age} = 18.03 years), who learned 60 Swahili–Swedish word pairs, 30 pairs through retrieval practice and 30 pairs through repeated studying.

Procedure

The intervention was performed through a web-based program that the participants accessed on their school-provided computers. Firstly, the participants became familiarized with the material as all the word pairs were presented to them one after the other. The learning phase then consisted of six rounds in which participants learned half of the word pairs through repeated studying (Adhama–Heder) and half through retrieval practice (Adhama–?). Responses to retrieval practice pairs were given by typing in the corresponding Swedish word. Correct answer feedback was then shown for one second. Retrieval practice and re-study word pairs were randomly assigned to each condition, and they were presented in a randomized order during the learning phase.

Retention tests

Retention was measured at three time-points: five minutes after the learning phase, after one week, and after four weeks. In order to investigate the effect of retrieval practice on retention at different time points, the retention test was set up so that one third of the items were added at each retention interval. Hence, the first retention test included one third of the items: 10 retrieval practice items and 10 re-study items. The second retention test included two thirds of the items: 20 retrieval practice items and 20 re-study items. The third retention test included all of the items: 30 retrieval practice items and 30 re-study items. The procedure for the retention tests was identical to that of the retrieval practice items during the learning session but without the correct answer feedback.

Independent variables
Individual differences related to non-cognitive factors were assessed using the measures Grit and NFC. Individual differences related to cognitive factors were represented using a measure of WMC. The assessment of WMC was completed one month before the intervention took place. Questionnaires measuring grit and NFC were filled out during the same session as the second retention test.

Statistical analyses
The main analyses consisted of 2 × 3 repeated-measures ANOVAs, with condition (retrieval practice and re-study) and retention interval (five minutes, one week, and four weeks) as within-subjects variables. Paired sample *t*-tests were used to investigate the difference between retrieval practice and re-study at each retention interval. Finally, the associations between the independent variables, Grit, NFC, and WMC, and retention were explored using hierarchical regression analyses. Analyses for each condition and retention interval, a total of six analyses, were conducted with WMC, Grit, and NFC as predictors. WMC was added in the first step, while Grit and NFC were added in the second step. Separate ANOVAs and regression analyses were conducted with two dependent variables, one containing all the word pairs tested at each retention interval, referred to as accumulated word pairs (i.e., 20 items at the five-minute test, 40 items at the one-week test, and 60 items at the four-week test), and one containing the 20 items that were unique to each respective retention test.

Results

Learning using retrieval practice led to superior performance on all three retention tests compared to re-study. This was in contrast to Experiment I, where the differences in performance between the conditions were not significant. ANOVAs conducted on both accumulated and unique word pairs revealed main effects of retention interval for both retrieval practice and re-study word pairs, suggesting that performance on the retention tests declined across time for both learning techniques. In addition, neither WMC, Grit, nor NFC could predict performance on the retention tests for either condition.

Conclusions

These experiments further support the finding that retrieval practice, compared to re-studying, is a preferable learning technique that does not appear to be sensitive to individual differences in personality or WMC.

Study II

Bertilsson, F., Stenlund, T., Wiklund-Hörnqvist, C., & Jonsson, B. (2021). Retrieval practice: Beneficial for all students or moderated by individual differences? *Psychology Learning and Teaching*, *20*(1), 21–39.

Aims and methods

The second study was designed as a partial replication of Experiment II from Study I with the purpose of adding credibility to the results by replicating the study with a larger sample. As in Study I, the learning material consisted of 60 Swedish–Swahili word pairs; 30 word pairs were learned using retrieval practice and 30 were learned using repeated studying. A total of 151 upper-secondary school students (45.7% female, M_{age} = 17.1 years) participated in this study. The procedure of the learning session and retention tests were identical to Experiment II in Study I.

Independent variables
The measures of WMC, and of Grit and NFC were again used to assess individual differences in relation to cognitive and non-cognitive factors, respectively. The assessments were completed one week before the intervention.

Statistical analyses
As in Study I, changes in retention across time for the two conditions were investigated using a 2 × 3 repeated-measures ANOVA. Differences in retention between retrieval practice word pairs and re-study word pairs at each interval were determined by inspecting pairwise comparisons for the main effects of the ANOVA. Again, hierarchical regression analyses were conducted to investigate the relationship between the independent variables and the testing effect. The differences in performance on retrieval practice and re-study word pairs (i.e., retrieval practice – re-study = testing effect) at each retention interval were used as the dependent variables. WMC was entered in the first step, and Grit and NFC were entered in the second step. All analyses were conducted on both word pairs unique to each retention interval and on accumulated word pairs.

Results

The results of the ANOVA revealed main effects of both retention interval and practice condition, as well as an interaction effect between them, when the analysis was conducted using word pairs unique to each intervention interval as dependent variables. The same analysis conducted using accumulated word pairs as dependent variables also

found main effects of retention interval and practice condition, but no interaction effect. The pairwise comparisons for the main effects of retention interval and practice condition showed that performance differed significantly between all retention tests, and that performance was significantly better for word pairs practiced using retrieval practice compared to repeated studying. Neither WMC, Grit, nor NFC could predict the testing effect at any retention interval. Non-significant results were obtained for both unique and accumulated word pairs.

Conclusions

The results provide further support for retrieval practice as a learning strategy that is beneficial regardless of individual differences in WMC, Grit, and/or NFC.

Study III

Bertilsson, F., Stenlund, T., Sundström, A., & Jonsson, B (Manuscript). Self-regulated use of retrieval practice: Associations with individual differences in non-cognitive and cognitive factors.

Aims and methods

The aim of the third study was to investigate students' self-regulated use of retrieval practice as an optional learning strategy, when integrated into their school activities, and to determine whether individual differences in cognitive (fluid intelligence and WMC), and non-cognitive factors (NFC, grit, conscientiousness, and openness) are related to differences in their use of retrieval practice. In addition, it was deemed relevant to examine whether there are sex-related differences in the use of retrieval practice.

Participants
A total of 146 upper-secondary school students (27% female, M_{age} = 16.2 years) were included in the study, which consisted of two separate interventions. Altogether, 96 students participated in the mathematics intervention only (in two separate cohorts), 26 students took part in the Swedish intervention only, and 24 students took part in both.

Materials
In both interventions, the retrieval practice material consisted of eight quizzes, with a total of 20 items each. The content of the quizzes corresponded to the content of the chapters in the mathematics course book and subtopics of the Swedish course, respectively.

Procedure
The quizzes were made available to the students via their online learning platform. In order to investigate differences in quiz use when quizzes were completely optional (i.e., outside the classroom) and when they were also used in the classroom once each week (referred to as non-optional), an ABAB design was used. That is, optional and non-optional quizzes were alternated between sections/chapters. Quizzes were available during the time when the corresponding chapter/section was being worked on in class, an average of 11 days for quizzes in the Swedish course, and an average of 22 days for the first cohort in mathematics and 24 days for the second cohort.

Independent variables
The measures of fluid intelligence and WMC, and NFC, grit, conscientiousness, and openness were used to assess individual differences in relation to cognitive and non-cognitive factors, respectively. The assessments were completed in two sessions before the intervention started.

Statistical analyses
Descriptive statistics and paired samples *t*-tests were used to investigate quiz use for optional and non-optional quizzes. Independent samples *t*-tests were used to examine differences between males and females. Pearson correlations and hierarchical regression analyses were used to investigate the associations between optional quiz use and sex (step 1), fluid intelligence, WMC (step 2), NFC, grit, conscientiousness, and openness to experience (step 3).

Results

As expected, quiz use was higher for non-optional quizzes than for optional quizzes. The quizzes were used to a greater extent by females than by males, a difference that was statistically significant for both optional and non-optional quizzes in Swedish, and non-optional quizzes in mathematics. Grit and conscientiousness were positively correlated with optional quiz use in mathematics, and NFC, grit, and conscientiousness were positively correlated to optional use of quizzes in Swedish. The only significant positive predictor of quiz use in mathematics was conscientiousness, but for quiz use in Swedish, sex was a significant positive predictor in all three steps of the analysis, and NFC and conscientiousness were positive predictors in the third step, while openness was a negative predictor.

Conclusions

The self-regulated use of retrieval practice, in the form of quizzes, was low. The results also suggest that some groups of students are more likely to choose to use retrieval practice for studying in their free time. In general, females completed more quizzes than males. Also, students with higher scores on non-cognitive factors such as conscientiousness used the optional quizzes to a greater extent in both mathematics and Swedish. In conclusion, when retrieval practice is to be implemented in school settings, it should be included in classroom activities to ensure that all students take part and can reap the benefits.

Discussion

Given the worrying problem of a relatively large proportion of teenagers not finishing school with sufficient grades, effective and evidence-based tools that are easy to use could be valuable resources for teachers and students alike. Retrieval practice has been demonstrated to be an effective and relatively straightforward strategy for enhancing memory retention and learning, but it is important to ensure that the benefits found in lab-based studies apply to all students, independent of variations in both cognitive and non-cognitive factors, and that they translate to applied settings. The overall aim of the present thesis was to investigate the relationships between retrieval practice and individual differences related to both cognitive and non-cognitive factors, with the purpose of contributing information about how retrieval practice as a learning strategy should be recommended for use in education. Specifically, this was explored under two separate aims. The first aim was examined in Studies I and II and focused on the relationship between individual differences in terms of cognitive and non-cognitive factors, and the testing effect. The second aim, examined in Study III, was to explore the use of retrieval practice in an authentic school context, and to investigate the relationship between self-regulated use of the strategy and individual differences, as well as sex-related differences. In the following sections, I discuss the findings of the thesis in relation to previous research in the area. I then address some limitations of the thesis and provide suggestions for future research, as well as practical suggestions for how to utilize retrieval practice.

Is the testing effect related to individual differences in cognitive and/or non-cognitive factors that are important for learning?

Considering the well-documented effectiveness of retrieval practice as a learning strategy, it is important to assess whether it is equally beneficial for all students, or whether the benefit differs depending on individual differences that relate to academic achievement in general.

In Study I, the associations between individual differences and the testing effect were explored in two experiments. The first experiment used a between-subject design and the retrieval practice condition used multiple-choice questions as the test format. The results showed that performance was not significantly different between the retrieval practice and re-study groups (i.e., no testing effect occurred). In the second experiment, a

within-subject design was used, the retrieval practice format was changed to short answer, and a measure of working memory capacity (WMC) was included to investigate whether the more difficult task would reveal an association with WMC. The results showed that a testing effect was achieved; that is, retention was significantly higher for word pairs learned using retrieval practice. However, retention after retrieval practice was not predicted by individual differences in either experiment. Study II was designed as a replication of Experiment II from Study I, using a larger sample and, again, a testing effect was achieved, and it was found that neither WMC, need for cognition (NFC), nor grit were significant predictors of the testing effect.

When synthesizing the results from Studies I and II, the evidence suggests that the testing effect is independent of individual differences in cognitive factors—that is WMC—as well as the non-cognitive factors NFC and grit, because we did not find any significant relationships between retention and these factors. This result suggests that learning using a retrieval practice strategy is effective for students regardless of WMC or non-cognitive aspects. Regarding NFC and grit, these findings are in line with the existing research (Stenlund et al., 2017; Wiklund-Hörnqvist et al., 2022), although very few studies have investigated the relationships between the testing effect and NFC or grit. Regarding WMC, previous studies show inconsistent results, reporting positive (Agarwal et al., 2017; Zheng et al., 2023), negative (Tse & Pu, 2012), or no (Wiklund-Hörnqvist et al., 2014) relationships between WMC and the testing effect. Zheng et al. (2023) argue that these inconsistent findings, specifically the lack of association with WMC, could be explained by study designs that are not demanding enough, in terms of working memory load, for the associations to be revealed. In Experiment II of Study I of this thesis, the task difficulty was increased to avoid ceiling effects on performance, a tendency that was observed in Experiment I. A related concern was that increasing task difficulty might also increase working memory demand. Therefore, a measure of WMC was included in the study in order to assess its relationship with the testing effect. While no associations were found despite it being a more demanding task, a possible explanation for this could be that the working memory demands did not exceed student WMC (Zheng et al., 2023). This hypothesis about demand and capacity is interesting in a broader sense because the same reasoning could be applicable to other cognitive abilities. A couple of studies have reported similar findings regarding a difference in the magnitude of the testing effect related to intelligence and academic achievement, and task difficulty. Minear et al. (2018) reported that high-intelligence individuals achieved a larger testing effect for difficult items, while low-intelligence

individuals achieved a larger testing effect for easier items. Carpenter et al. (2016) did not explicitly manipulate the difficulty of the learning materials, but students were given an exercise in which they either recalled (retrieval practice) or copied (re-study) definitions, representing tasks with different levels of difficulty. The results showed that high-achieving students benefited more from learning with retrieval practice, while lower-achieving students benefited more from re-study (i.e., copying definitions). The findings from Carpenter et al. (2016), Minear et al. (2018), and Zheng et al. (2023) suggest that, when cognitive demands/task difficulty is included as a factor, individual differences may influence the testing effect.

In addition, these findings could be interpreted as providing support for the retrieval effort hypothesis for the testing effect. This hypothesis claims that successful but effortful retrieval enhances memory more than less effortful retrieval. However, one implication that follows from this reasoning is that, if the level of difficulty becomes too high for some individuals, they will not be able to achieve successful retrieval. On the other hand, if the difficulty is too low for high-ability individuals, they will not be required to exert the effort needed for memory to be enhanced. Furthermore, the findings also imply that one explanation for the lack of associations between individual differences and the testing effect found in this thesis and in other research could be due to task demand being relatively low in retrieval practice experiments. While task/item difficulty was not investigated in Studies I and II, including a measure of intelligence in addition to WMC would have been interesting. Neither Study I nor II explicitly investigated any of the theoretical explanations for the testing effect, but the above reasoning regarding low task demand suggests that retrieval effort may not have been responsible for the observed testing effects in these studies. Given that the response formats were identical during the learning sessions and during the retention tests (i.e., multiple choice in Experiment I in Study I, and cued recall during Experiment II in Study I and Study II), it appears more likely that the effect was driven by transfer-appropriate processing. Although the focus of these two studies was to examine individual differences in the testing effect, the results also provide some support for the theoretical underpinnings of the testing effect. More research aimed specifically at this topic is needed, however, in order to empirically test the theoretical explanations.

Overall, the results regarding individual differences in non-cognitive factors appear, so far, to cautiously indicate that retrieval practice is an effective learning strategy irrespective of individual differences.

Regarding individual differences in cognitive factors, the available results are more complex. One similarity among the inconsistent results discussed previously in this section is that all the studies found that retrieval practice is in fact beneficial for all students, but that it might be more, or less, beneficial for individuals with specific characteristics. Hence, further research on the relationship between the testing effect and individual differences should focus on bringing clarity to this confusion. And in this endeavor, the role of task demand/difficulty should be considered. In the meantime, retrieval practice can be considered to be an effective learning strategy.

Is self-regulated use of retrieval practice related to individual differences in cognitive and/or non-cognitive factors?

While a vast amount of research has demonstrated the positive effects of retrieval practice, both in lab-based studies and inside classrooms, research on its optional use has found that students use the strategy to only a low extent. Therefore, the focus for Study III instead shifted to investigating how students use retrieval practice when it is optional, and whether there were differences in use related to sex, or individual differences in cognitive factors (fluid intelligence, WMC), or non-cognitive factors (NFC, grit, conscientiousness, or openness). To determine this, a retrieval practice intervention was implemented in an authentic school environment. In line with previous research (Corral et al., 2020; B. C. Johnson & Kiviniemi, 2009; G. M. Johnson, 2006), we found that self-regulated use of quizzes (i.e., retrieval practice) was low compared to the use of quizzes that were also included as a classroom activity, and this was true both for quizzes implemented in a mathematics course and for those in a Swedish course. Importantly, this difference between completely optional quizzes and quizzes that were also included as a once-a-week in-class activity was observed despite the students being given information about the effectiveness of retrieval practice at the start of the intervention. This finding is further discussed in the future directions section, below.

Additional analyses found that self-regulated use of retrieval practice was positively predicted, in mathematics, by conscientiousness, and in Swedish by sex, NFC, and conscientiousness, and negatively by openness to experience. This indicates that non-cognitive factors are more important for determining whether a student will use optional quizzes than cognitive factors. To the best of my knowledge, only a couple of additional studies have investigated the association between individual differences and self-regulated use of retrieval practice, and their results

differ from the results obtained in this study. Student engagement with retrieval practice activities on an e-learning platform were recorded and related to cognitive and non-cognitive individual differences (Fellman, Lincke, & Jonsson, 2020; Fellman, Lincke, Berge, et al., 2020). The first study found that fluid intelligence (as measured by Raven's Advanced Progressive Matrices), but not NFC or grit, was related to quiz use (Fellman, Lincke, & Jonsson, 2020), and the second found that quiz use was related to verbal working memory (Fellman, Lincke, Berge, et al., 2020). In contrast to Study III, these findings suggest that cognitive factors are stronger predictors of self-regulated use of retrieval practice than non-cognitive factors.

In light of the abovementioned inconsistencies, it is important to consider the different samples used in the studies. Study III included upper-secondary school students with a mean age of 16 years, while Fellman, Lincke, and Jonsson (2020) included medical students with a mean age of 30 years. Demographic data was not disclosed in Fellman, Lincke, Berge, et al. (2020), but the sample was drawn from the same educational platform as in Fellman, Lincke, and Jonsson (2020), implying similar sample characteristics. It does not seem farfetched to suggest that upper-secondary students differ in important aspects from adult students who have chosen to study one of the most competitive higher education programs in Sweden. As mentioned earlier, it has been suggested that personality characteristics reflect typical performance, while intelligence reflects maximum performance (Chamorro-Premuzic et al., 2006; Goff & Ackerman, 1992). Thus, when considering that the participants in Fellman, Lincke, and Jonsson (2020) as well as Fellman, Lincke, Berge, et al. (2020) may represent high-performing and homogenous groups in terms of both cognitive and non-cognitive factors, perhaps it is reasonable to expect that it is differences in cognitive factors that affect how they self-regulate their learning, rather than differences in personality.

Regarding the associations with non-cognitive factors, research specifically related to self-regulated use of retrieval practice is lacking, but research regarding self-regulated learning in general could give some valuable insights into this matter. Support for the importance of conscientiousness and openness in self-regulated learning was found by Bidjerano and Dai (2007), who reported that higher levels of conscientiousness and openness were related to higher levels of five self-regulatory strategies (critical thinking, metacognition, effort regulation, time management, and elaboration). Interestingly, the study also suggested that a combination of lower levels of openness and neuroticism, together with high levels of conscientiousness and agreeableness, was

related to greater effort regulation and better time-management skills, but poorer critical thinking skills (Bidjerano & Dai, 2007). While self-regulated use of retrieval practice was used here, rather than self-regulatory strategies per se, the findings of Bidjerano and Dai (2007) may provide clarification for the observed negative relationship between openness and quiz use found in Study III.

Differences between males and females in terms of the testing effect have not been explored extensively in the existing research. There are examples of studies where a control analysis has been performed to check for sex-related differences, but I have not been able to find any studies reporting significant differences. This is not very surprising as the theoretical explanations for the testing effect do not appear to include mechanisms that one would expect to differ between males and females. However, in terms of self-regulated use of retrieval practice, the conditions are different. Research has repeatedly reported sex-related differences in self-regulated learning (Marrs & Sigler, 2012; Pérez et al., 2017; Ruffing et al., 2015; Weis et al., 2013), suggesting that such differences would apply to self-regulated use of retrieval practice as well. While this has not yet been explored extensively in studies specifically focusing on self-regulated use of retrieval practice, self-report data indicates that females are more likely to use self-testing than males (Gagnon & Cormier, 2019). Furthermore, research on self-regulated learning and use of learning strategies has reported that, in general, females have better study skills than males (Marrs & Sigler, 2012). Importantly, this is in line with the results of Study III, where females in general completed more quizzes than males. Considering these observed differences between males and females, it is important to explore such differences further in relation to self-regulated use of retrieval practice, with the purpose of finding solutions to reduce their impact on differences in achievement.

The non-cognitive variables that significantly predicted quiz use in Study III differed between the two subjects selected. Conscientiousness was a strong positive predictor of optional quiz use in both mathematics and Swedish, but that was the only similarity between the school subjects. For Swedish, sex, NFC (positive), and openness (negative) were also significant predictors. Since this study was not designed with the purpose of investigating differences between the subjects, any ideas about the reasons behind these differences are merely speculative. However, one possible aspect that might influence how quizzes are used in these subjects are their qualitative differences. In mathematics, testing one's abilities, in the form of practicing new mathematical problems, is an inherent part of the subject. In Swedish, a subject where the purpose is to improve

communication and literary abilities, it may require more conscientiousness and motivation to learn how best to use quizzes to study. However, one methodological difference that could also influence the results is that the quizzing period for each quiz was longer in mathematics than in Swedish, which means that more quizzes were completed in general. Further limitations in terms of methodological differences are discussed in the Limitations section, below.

Similarly to the approach taken in Study III in the present thesis, efforts have also been made by other researchers to improve or increase students' self-regulated use of retrieval practice (Ariel & Karpicke, 2017; Einstein et al., 2012). Before taking part in a retrieval practice experiment, the experimental group was given a brief set of instructions that explained the benefits of retrieval practice and how it should be used, while the control group were only asked to learn as many word pairs as possible. The results revealed that informing students of the benefits of retrieval practice improved performance, both in the current experiment and in a follow-up one week later (Ariel & Karpicke, 2017). Interestingly, using a similar inspirational introduction did not appear to be successful in increasing the use of retrieval practice in Study III in the present thesis. However, the temporal scope of the interventions (one week vs. two school semesters) could be an explanation for these differing results. Einstein et al. (2012) achieved better results after having students write a lab report on their own results from an experiment comparing learning after re-study and retrieval practice. An end-of-semester questionnaire revealed that, after this activity, students were 82% more likely to use testing during studying. Further research should investigate how, and to what extent, students can be influenced in terms of self-regulated use of retrieval practice.

Conclusions

To conclude, the results of the studies conducted for this thesis suggest that learning using a retrieval practice learning strategy is beneficial for students regardless of their working memory capacity, their level of grit, or their need for cognition. In terms of how students chose to use a retrieval practice tool when it was made available to them, we found that sex, conscientiousness, openness, NFC, and grit were related to this use. The results imply that females are more likely to engage in using retrieval practice than males. Students who are more conscientious and/or have a higher NFC choose to use retrieval practice to a greater extent than students who are less conscientious and/or have a lower NFC. In addition, openness was negatively related to quiz use, suggesting that students who

are less open to experience use retrieval practice to a greater extent than those who are more open to experience.

Limitations

Studies I and II

One evident limitation of this research is the low number of participants, specifically in Study I. While that study's findings were interesting and promising, they were mainly valuable as a guide for Study II due to the low statistical power, which was confirmed by a power analysis conducted to ensure sufficient power for Study II.

As mentioned previously in the discussion, it could be argued that the choice of independent variables for Studies I and II was not optimal. These studies were highly controlled and used material that is not educationally relevant, which could indicate that the context may not be stimulating enough for individual differences in NFC or grit to come into play. However, this is merely speculation about why no relationships with the testing effect were found. Given the previously identified associations of NFC and grit with both academic achievement and career-related success, a thorough investigation in relation to retrieval practice is justified. Furthermore, only WMC was included as a measure of cognitive factors. While WMC is highly relevant from a learning perspective, so are many other factors, such as intelligence, episodic memory, and executive functioning, which were not included in these studies.

Study III

As in Study I, the low number of participants in the Swedish group is one important limitation. The problem of the small size of this group mainly affected the ability to make reliable comparisons between the two subjects. Ideally, a larger and more heterogenous sample should have been used. Unfortunately, the participating classes were selected by the school and this was out of my control. Another, somewhat related, limitation is that four teachers in two subjects administered the intervention to their individual classes. It is likely that this introduced procedural differences between the classes that may be reflected in the outcome variable. It could be argued that the differences in quiz use between mathematics and Swedish may be connected to teacher-related differences rather than differences between the subjects per se. This lack of control affects the internal validity, but on the other hand the approach provides ecologically valid research in an authentic educational setting.

Validity and reliability of the measures

There are also some limitations related to the quality of the different instruments used to measure cognitive and non-cognitive factors in the studies, and these need to be addressed. As in all psychological research, the measurement of psychological constructs implies issues of validity and reliability. The main constructs used in this thesis are well established in terms of definitions, but nonetheless, there are several different ways to measure them (operational definitions) and the choice of measure that we use to represent a certain construct will influence the results of the research. Using other measures for the same constructs might have led to different results. Therefore, it is important to bear this in mind when interpreting the practical significance of the results obtained here, and in psychological research in general. On the other hand, several of the measures employed here were included in all three studies. This provides support for the observed variations in how individual differences are related to the benefit and use of retrieval practice.

Another potential limitation regarding validity is that only one measure was used to represent each construct, and that some measures included very few items. It would have been valuable to include several measures of each construct to increase the validity, and to use a measure of conscientiousness that was more comprehensive in terms of items, but this was not possible due to logistical reasons. However, the instruments used in the studies included in this thesis have all demonstrated acceptable reliability and validity.

Future directions, and implications

As mentioned earlier in this discussion, further research investigating the relationships between individual differences and the testing effect are warranted. This would assist in both further illuminating the inconsistent findings achieved to date and investigating other factors that have not been explored thus far.

In addition, I think it is time for studies performed in educational settings to take the leap toward performing more ecologically valid studies. While results are more easily interpretable when there is more experimental control, such control also eliminates certain research questions from being investigated. For example, in the review by Agarwal et al. (2021), only studies where the use of retrieval practice was controlled by the instructor were included in order to maintain the internal validity of the

review. This is not wrong per se, but it means that the conclusions can only be generalized to settings where the use of retrieval practice is controlled.

Retrieval practice is a relatively simple strategy to use, but because repeated repetitions of a material are required to produce lasting retention, it can be time-consuming to implement as a classroom activity. Therefore, it would be preferable if students could be relied upon to use this effective strategy when studying outside of the classroom. However, the results of Study III, and previous research on self-regulated use of retrieval practice, suggest the students do not use this strategy to any great extent when it is optional.

The findings presented in this thesis suggest that, while the testing effect is a robust phenomenon that appears to be beneficial regardless of individual differences in either cognitive or non-cognitive factors, its implementation in an educational setting requires consideration due to an overall low level of self-regulated use that is related to non-cognitive factors.

Furthermore, self-regulated use of retrieval practice was found to be related to non-cognitive factors associated with academic achievement, indicating that students who use the strategy may already be high achievers. In addition, it was found that females used retrieval practice to a greater extent than males. These potential issues could be mitigated by making retrieval practice mandatory to some extent, which would ensure that the strategy was used more, and that the beneficial effects would support all students.

Based on the results of this thesis, I want to take the opportunity to share some practical suggestions for teachers, teacher educators, and others who may be interested.

- We know that retrieval practice works, and it appears to work well for most students – use it in the classroom.
- We also know that students of all ages are relatively poor at using effective learning strategies. They do not know what to study or how to do it so they should be supported in using effective learning strategies outside of the classroom as well.
- Teach teachers the science of learning. Giving teachers basic knowledge about the science of learning and an awareness of how to encourage self-regulated learning would provide them with valuable tools that could support students in becoming efficient learners.

- Teach students the science of learning. The fact that students are having problems with self-regulated learning in upper-secondary school and at university suggests that more needs to be done at earlier stages of the educational system. From very young ages, students should be taught why certain tasks are more beneficial, and when to use which tasks in order to learn effectively. If this was an everyday part of education, *how to* learn would not have to be an obstacle for students to overcome.

As mentioned in the introduction, there is not one single cause for the alarming number of teenagers who do not graduate from upper-secondary school. To be frank, I have come to realize that the problem is much larger than something that a single learning strategy can solve singlehandedly. But giving students tools and knowledge from a young age that may allow them to study more effectively, and struggle less, could be one part of the solution. And retrieval practice is a valuable tool to have in one's toolbox.

Acknowledgement

When I first read the quote in the beginning of this book, I thought that it was just another cheesy saying appropriate for a refrigerator magnet. But then I read it again and realized that it captures the essence of my thesis quite well. We are born with a capacity for learning, at least in part consisting of cognitive abilities – the gift. But this capacity does not guarantee success. We also have to be able to acquire lasting knowledge – the skill – and be motivated to do so – the choice.

During my six and a half years as a PhD student, it was certainly not only my cognitive abilities that kept me going towards this distant goal. Perhaps I was helped by a little bit of grit, or maybe a pinch of need for cognition and conscientiousness. Certainly, the skill was not all mine, I have had wonderful people surrounding me, making this possible. While the process from start to finish has had its ups and downs, it has been a great experience, and for the most part a lot of fun!

Of course, I could not have gotten this far without my supervisors, current or previous ones. **Tova,** thank you for all your support during these years, and for always being so positive and reassuring. I think I have left all our meetings feeling more positive about the future than I did going into it. Your feedback is always constructive and on point. You let the small stuff go and focus on what's important, making it easy to trust your instinct. Thank you for putting up with me and for your invaluable guidance. And don't worry, I will keep bothering you for advice in the future! **Bert,** thank you for welcoming me into your research group as a master student and for trusting me to take on this project. You humble, unpretentious character and scientific knowledge was very inspirational to meet as a beginner in the academic world. Thank you for your support and expertise during this project. **Carola,** your enthusiasm for research is simply infectious! Thank you for being an inspiration during my first experiences as a teacher and researcher. **Anna S,** thank you for joining my supervisor team at the most confused stage and giving valuable insight to the final parts of this thesis. Your calm and friendly character, and sharp attention to detail, was just what I needed. I hope that we get the opportunity to work together again in the future!

To the amazing teachers at Balderskolan, **Ingemar, Per, Irene and Linda**, thank you for taking on this slightly chaotic project with me. It would simply not have been possible without your insight. I think we all learned a lot, both about retrieval practice and about doing research in a

school setting. It was not an easy task, but we did it! And a huge thank you to all the students who have been kind enough to participate in the three studies!

The doctoral student group, what a gang! Thank you to everyone, past and present members of the group, for making the (sometimes tough) work environment for doctoral students so warm, inviting, and unpretentious. Thanks for the much-needed coffee in the morning, for all the conversations - work related and other random stuff - and for just having each other's backs. The support and understanding given by those who are down in the trenches alongside you is invaluable. And we have also had a lot of fun! I did contemplate asking for prolongation just so that I could cling on to this group for a little longer. What will I do without our Teams-chat?! But every season comes to an end. At least I can keep crashing your morning coffee, and I am certain that the friendships will remain!

Finally, to my family and friends. **Mamma, pappa, Elin, Ulrika, farmor**, thank you for believing in me and for your constant support. I am very lucky to have you by my side on this walk through life. **Nichlas, Felicia, Clara,** what a joy to spend everyday life with you. Thank you for distracting me from work and showing me what is most important. **Obbolatjejerna, bokklubbstjejerna**, even though there has been less time for hanging out these last few years, I hope you know how much you mean to me. Thank you for all the fun and support! **Carola and Gunnar,** thank you for welcoming me into your family with open arms. And thanks again for taking care of Clara when I tried to edit my second paper while being on maternity leave, I couldn't have finished on time without you!

References

Abott, E. E. (1909). On the Analysis of the Factor of Recall in the Learning Process. *Psychological Monographs: General and Applied*, *11*(1), 159–177. https://doi.org/10.1037/h0093018

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the Use of Tests. *Review of Educational Research*, *87*(3), 0034654316689306. https://doi.org/10.3102/0034654316689306

Agarwal, P. K. (2019). Retrieval practice & bloom's taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology*, *111*(2), 189–209. https://doi.org/10.1037/edu0000282

Agarwal, P. K., Finley, J. R., Rose, N. S., & Roediger III, H. L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory*, *25*(6), 764–771. https://doi.org/10.1080/09658211.2016.1220579

Agarwal, P. K., Nunes, L. D., & Blunt, J. R. (2021). Retrieval Practice Consistently Benefits Student Learning: a Systematic Review of Applied Research in Schools and Classrooms. *Educational Psychology Review*, *33*(4), 1409–1453. https://doi.org/10.1007/s10648-021-09595-9

Ahmed, S. F., Tang, S., Waters, N. E., & Davis-Kean, P. (2019). Executive function and academic achievement: Longitudinal relations from early childhood to adolescence. *Journal of Educational Psychology*, *111*(3), 446–458. https://doi.org/10.1037/edu0000296

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–29. https://doi.org/10.1016/j.jecp.2009.11.003

American Educational Research Association, (AERA), American Psychological Association, (APA), & National Council on Measurement in Education, (NCME). (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association. https://books.google.se/books?id=clI_mAEACAAJ

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

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. *Psychology of Learning and Motivation - Advances in Research and Theory*, *2*(C), 89–195. https://doi.org/10.1016/S0079-7421(08)60422-3

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D., & Hitch, G. (1974). *Working Memory* (pp. 47–89). https://doi.org/10.1016/S0079-7421(08)60452-1

Batsell, W. R., Perry, J. L., Hanley, E., & Hostetter, A. B. (2017). Ecological Validity of the Testing Effect: The Use of Daily Quizzes in Introductory Psychology. *Teaching of Psychology*, *44*(1), 18–23. https://doi.org/10.1177/0098628316677492

Bidjerano, T., & Dai, D. Y. (2007). The relationship between the big-five model of personality and self-regulated learning strategies. *Learning and Individual Differences*, *17*(1), 69–81. https://doi.org/10.1016/j.lindif.2007.02.001

Bjork, E. L., & Bjork, R. A. (2011). Making Things Hard on Yourself, But in a Good Way: Creating Desirable Difficulties to Enhance Learning. In *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society* (pp. 55–64). https://doi.org/10.1017/CBO9781107415324.004

Bjork, E. L., Little, J. L., & Storm, B. C. (2014). Multiple-choice testing as a desirable difficulty in the classroom. *Journal of Applied Research in Memory and Cognition*, *3*(3), 165–170. https://doi.org/10.1016/j.jarmac.2014.03.002

Bjork, R. A. (1994). Memory and Metamemory Considerations in the Training of Human Beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing*. The MIT Press. https://doi.org/10.1016/S0013-4694(97)84006-4

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

Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, *106*(3), 849–858. https://doi.org/10.1037/a0035934

Brewer, G. A., & Unsworth, N. (2012). Individual differences in the effects of retrieval from long-term memory. *Journal of Memory and Language*, *66*(3), 407–415. https://doi.org/10.1016/j.jml.2011.12.009

Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences*, *29*(6), 1057–1068. https://doi.org/10.1016/S0191-8869(99)00253-6

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1118–1133. https://doi.org/10.1037/a0019902

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, *36*(3), 604–616. https://doi.org/10.3758/MC.36.3.604

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1563–1569. https://doi.org/10.1037/a0017021

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1547–1552. https://doi.org/10.1037/a0024140

Carpenter, S. K., & Delosh, E. L. (2006). Impoverished cue support

enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, *34*(2), 268–276. https://doi.org/10.3758/BF03193405

Carpenter, S. K., Lund, T. J. S., Coffman, C. R., Armstrong, P. I., Lamm, M. H., & Reason, R. D. (2016). A Classroom Study on the Relationship Between Student Achievement and Retrieval-Enhanced Learning. *Educational Psychology Review*, *28*(2), 353–375. https://doi.org/10.1007/s10648-015-9311-9

Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). *The science of effective learning with spacing and retrieval practice*. *496*. https://doi.org/10.1038/s44159-022-00089-1

Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin and Review*, *14*(3), 474–478. https://doi.org/10.3758/BF03194092

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology*, *23*(6), 760–771. https://doi.org/10.1002/acp.1507

Chamorro-Premuzic, T., Furnham, A., & Ackerman, P. L. (2006). Incremental Validity of the Typical Intellectual Engagement Scale As Predictor of Different Academic Performance Measures. *Journal of Personality Assessment*, *87*(3), 261–268. https://doi.org/10.1207/s15327752jpa8703_07

Cho, K. W., Neely, J. H., Brennan, M. K., Vitrano, D., & Crocco, S. (2017). Does testing increase spontaneous mediation in learning semantically related paired associates? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1768–1778. https://doi.org/10.1037/xlm0000414

Coane, J. H. (2013). Retrieval practice and elaborative encoding benefit memory in younger and older adults. *Journal of Applied Research in Memory and Cognition*, *2*(2), 95–100. https://doi.org/10.1016/j.jarmac.2013.04.001

Colling, J., Wollschläger, R., Keller, U., Preckel, F., & Fischbach, A. (2022). Need for Cognition and its relation to academic achievement in different learning environments. *Learning and Individual Differences*, *93*, 102110.

https://doi.org/10.1016/j.lindif.2021.102110

Coppens, L. C., Verkoeijen, P. P. J. L., Bouwmeester, S., & Rikers, R. M. J. P. (2016). The testing effect for mediator final test cues and related final test cues in online and laboratory experiments. *BMC Psychology*, *4*(1), 25. https://doi.org/10.1186/s40359-016-0127-2

Corral, D., Carpenter, S. K., Perkins, K., & Gentile, D. A. (2020). Assessing students' use of optional online lecture reviews. *Applied Cognitive Psychology*, *34*(2), 318–329. https://doi.org/10.1002/acp.3618

Cowan, N. (2014). Working Memory Underpins Cognitive Development, Learning, and Education. In *Educ Psychol Rev*. (Vol. 26, Issue 2). https://doi.org/10.1007/s10648-013-9246-y.Working

Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*, *113*(3), 492–511. https://doi.org/10.1037/pspp0000102

Daniel, D. B., & Broida, J. (2004). Using Web-Based Quizzing to Improve Exam Performance: Lessons Learned. *Teaching of Psychology*, *31*(3), 207–208. https://doi.org/10.1207/s15328023top3103_6

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*(5), 952–957. https://doi.org/10.1002/acp.3584

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP Scales : Tiny-Yet-Effective Measures of the Big Five Factors of Personality. *Psychological Assessment*, *18*(2), 192–203. https://doi.org/10.1037/1040-3590.18.2.192

Dornic, S., Ekehammar, B., & Laaksonen, T. (1991). Tolerance for mental effort: Self-ratings related to perception, performance and personality. *Personality and Individual Differences*, *12*(3), 313–319. https://doi.org/10.1016/0191-8869(91)90118-U

Downey, L. A., Lomas, J., Billings, C., Hansen, K., & Stough, C. (2014). Scholastic Success: Fluid Intelligence, Personality, and Emotional Intelligence. *Canadian Journal of School Psychology*, *29*(1), 40–53.

https://doi.org/10.1177/0829573513505411

Duchastel, P. C., & Nungester, R. J. (1982). Testing Effects Measured with Alternate Test Forms. *The Journal of Educational Research*, *75*(5), 309–313. http://www.jstor.org/stable/27539915

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Duckworth, A. L., & Quinn, P. D. (2009). Development and Validation of the Short Grit Scale (Grit–S). *Journal of Personality Assessment*, *91*(2), 166–174. https://doi.org/10.1080/00223890802634290

Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2021). Revisiting the Factor Structure of Grit: A Commentary on Duckworth and Quinn (2009). *Journal of Personality Assessment*, *103*(5), 573–575. https://doi.org/10.1080/00223891.2021.1942022

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

Einstein, G. O., Mullet, H. G., & Harrison, T. L. (2012). The Testing Effect: Illustrating a Fundamental Concept and Changing Study Strategies. *Teaching of Psychology*, *39*(3), 190–193. https://doi.org/10.1177/0098628312450432

Eriksson, J., Kalpouzos, G., & Nyberg, L. (2011). Rewiring the brain with repeated retrieval: A parametric fMRI study of the testing effect. *Neuroscience Letters*, *505*(1), 36–40. https://doi.org/10.1016/j.neulet.2011.08.061

Fazio, L. K., & Marsh, E. J. (2019). Retrieval-Based Learning in Children. *Current Directions in Psychological Science*, *28*(2), 111–116. https://doi.org/10.1177/0963721418806673

Fellman, D., Lincke, A., Berge, E., & Jonsson, B. (2020). Predicting Visuospatial and Verbal Working Memory by Individual Differences in E-Learning Activities. *Frontiers in Education*, *5*(March), 1–11. https://doi.org/10.3389/feduc.2020.00022

Fellman, D., Lincke, A., & Jonsson, B. (2020). Do Individual Differences in Cognition and Personality Predict Retrieval Practice Activities on MOOCs? *Frontiers in Psychology*, *11*(2076), 1–10. https://doi.org/10.3389/fpsyg.2020.02076

Fleischhauer, M., Enge, S., Brocke, B., Ullrich, J., Strobel, A., & Strobel, A. (2010). Same or Different? Clarifying the Relationship of Need for Cognition to Personality and Intelligence. *Personality and Social Psychology Bulletin*, *36*(1), 82–96. https://doi.org/10.1177/0146167209351886

Freda, N. M., & Lipp, M. J. (2016). Test-Enhanced Learning in Competence-Based Predoctoral Orthodontics: A Four-Year Study. *Journal of Dental Education*, *80*(3), 348–354. https://doi.org/10.1002/j.0022-0337.2016.80.3.tb06091.x

Fritz, C. O., Morris, P. E., Nolan, D., & Singleton, J. (2007). Expanding retrieval practice: An effective aid to preschool children's learning. *Quarterly Journal of Experimental Psychology*, *60*(7), 991–1004. https://doi.org/10.1080/17470210600823595

Furnham, A., Monsen, J., & Ahmetoglu, G. (2009). Typical intellectual engagement, big five personality traits, approaches to learning and cognitive ability predictors of academic performance. *British Journal of Educational Psychology*, *79*(4), 769–782. https://doi.org/10.1348/978185409X412147

Gagnon, M., & Cormier, S. (2019). Retrieval Practice and Distributed Practice: The Case of French Canadian Students. *Canadian Journal of School Psychology*, *34*(2), 83–97. https://doi.org/10.1177/0829573518773225

Gates, A. I. (1917). *Recitation as a factor in memorizing* (Issue 40). The Science Press.

Goff, M., & Ackerman, P. L. (1992). Personality-intelligence relations: Assessment of typical intellectual engagement. *Journal of Educational Psychology*, *84*(4), 537–552. https://doi.org/10.1037/0022-0663.84.4.537

Goldberg, L. R. (1992). Development of Markers for The Big-Five Factor Structure. *Psychological Assessment*, *4*(1), 26–42.

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., Tabbers, H. K.,

Bouwmeester, S., & Zwaan, R. A. (2016). Distributed Practice and Retrieval Practice in Primary School Vocabulary Learning: A Multi-classroom Study. *Applied Cognitive Psychology*, *30*(5), 700–712. https://doi.org/10.1002/acp.3245

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*(1), 13–23. https://doi.org/10.1016/S0160-2896(97)90011-8

Grass, J., Strobel, A., & Strobel, A. (2017). Cognitive investments in academic success: The role of need for cognition at university. *Frontiers in Psychology*, *8*(MAY), 1–9. https://doi.org/10.3389/fpsyg.2017.00790

Hakimi, S., Hejazi, E., & Lavasani, M. G. (2011). The Relationships Between Personality Traits and Students' Academic Achievement. *Procedia - Social and Behavioral Sciences*, *29*, 836–845. https://doi.org/10.1016/j.sbspro.2011.11.312

Hill, B. D., Foster, J. D., Elliott, E. M., Shelton, J. T., McCain, J., & Gouvier, W. D. (2013). Need for cognition is related to higher general intelligence, fluid intelligence, and crystallized intelligence, but not working memory. *Journal of Research in Personality*, *47*(1), 22–25. https://doi.org/10.1016/j.jrp.2012.11.001

Hong, M. K., Polyn, S. M., & Fazio, L. K. (2019). Examining the episodic context account: does retrieval practice enhance memory for context? *Cognitive Research: Principles and Implications*, *4*(1), 46. https://doi.org/10.1186/s41235-019-0202-3

Johnson, B. C., & Kiviniemi, M. T. (2009). The Effect of Online Chapter Quizzes on Exam Performance in an Undergraduate Social Psychology Course. *Teaching of Psychology*, *36*(1), 33–37. https://doi.org/10.1080/00986280802528972

Johnson, G. M. (2006). Optional online quizzes: College student use and relationship to achievement. *Canadian Journal of Learning and Technology / La Revue Canadienne de l'apprentissage et de La Technologie*, *32*(1). https://doi.org/10.21432/t2j300

Jonsson, B., Wiklund-Hörnqvist, C., Stenlund, T., Andersson, M., & Nyberg, L. (2021). A Learning Method for All: The Testing Effect Is Independent of Cognitive Ability. *Journal of Educational Psychology*, *113*(5), 972–985. https://doi.org/10.1037/edu0000627

Kang, S. H. K. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory and Cognition*, *38*(8), 1009–1017. https://doi.org/10.3758/MC.38.8.1009

Kang, S. H. K., McDaniel, M. A., & Pashler, H. (2011). Effects of testing on learning of functions. *Psychonomic Bulletin & Review*, *18*(5), 998–1005. https://doi.org/10.3758/s13423-011-0113-x

Kang, S. H. K., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*(4–5), 528–558. https://doi.org/10.1080/09541440601056620

Karlsson Wirebring, L., Wiklund-Hörnqvist, C., Eriksson, J., Andersson, M., Jonsson, B., & Nyberg, L. (2015). Lesser Neural Pattern Similarity across Repeated Tests Is Associated with Better Long-Term Memory Retention. *Journal of Neuroscience*, *35*(26), 9595–9602. https://doi.org/10.1523/JNEUROSCI.3550-14.2015

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., Blunt, J. R., & Smith, M. A. (2016). Retrieval-based learning: Positive effects of retrieval practice in elementary school children. *Frontiers in Psychology*, *7*(MAR), 1–9. https://doi.org/10.3389/fpsyg.2016.00350

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-Based Learning: An Episodic Context Account. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 61, pp. 237–284). Academic Press. https://doi.org/10.1016/B978-0-12-800283-4.00007-1

Karpicke, J. D., & Roediger, H. L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 704–719. https://doi.org/10.1037/0278-7393.33.4.704

Karpicke, J. D., & Roediger III, H. L. (2008). The Critical Importance of Retrieval for Learning. *Science*, *319*(5865), 966–968. https://doi.org/10.1126/science.1152408

Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory and Language*, *67*(1), 17–29. https://doi.org/10.1016/j.jml.2012.02.004

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review*, *14*, 219–224. https://doi.org/10.3758/BF03194055

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, *65*(2), 85–97. https://doi.org/10.1016/j.jml.2011.04.002

Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences*, *42*(3), 441–451. https://doi.org/10.1016/j.paid.2006.08.001

Lamotte, M., Izaute, M., & Darnon, C. (2021). Can tests improve learning in real university classrooms? *Journal of Cognitive Psychology*, *33*(8), 974–992. https://doi.org/10.1080/20445911.2021.1956939

Lehman, M., & Karpicke, J. D. (2016). Elaborative retrieval: Do semantic mediators improve memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1573–1591. https://doi.org/10.1037/xlm0000267

Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1787–1794. https://doi.org/10.1037/xlm0000012

Lipko-Speed, A., Dunlosky, J., & Rawson, K. A. (2014). Does testing with feedback help grade-school children learn key concepts in science? *Journal of Applied Research in Memory and Cognition*, *3*(3), 171–176. https://doi.org/10.1016/j.jarmac.2014.04.002

Luong, C., Strobel, A., Wollschläger, R., Greiff, S., Vainikainen, M. P., & Preckel, F. (2017). Need for cognition in children and adolescents: Behavioral correlates and relations to academic achievement and potential. *Learning and Individual Differences*, *53*, 103–113. https://doi.org/10.1016/j.lindif.2016.10.019

Marrs, H., & Sigler, E. A. (2012). Male academic performance in college: The possible role of study strategies. *Psychology of Men & Masculinity*, *13*(2), 227–241. https://doi.org/10.1037/a0022247

McCrae, R. R., & John, O. P. (1992). An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, *60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*(4–5), 494–513. https://doi.org/10.1080/09541440701326154

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in Middle-School Science: Successful Transfer Performance on Classroom Exams. *Applied Cognitive Psychology*, *27*(3), 360–372. https://doi.org/10.1002/acp.2914

McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger, H. L., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied*, *20*(1), 3–21. https://doi.org/10.1037/xap0000004

McDermott, K. B., & Roediger, H. L. (2014). Memory (Encoding, Storage, Retrieval). *General Psychology FA2018. Noba Project: Milwaukie, OR*, 117–153.

Meyer, A. N. D., & Logan, J. M. (2013). Taking the testing effect beyond the college freshman: Benefits for lifelong learning. *Psychology and Aging*, *28*(1), 142–147. https://doi.org/10.1037/a0030890

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Minear, M., Coane, J. H., Boland, S. C., Cooney, L. H., & Albat, M. (2018). The benefits of retrieval practice depend on item difficulty and intelligence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(9), 1474–1486. https://doi.org/10.1037/xlm0000486

Moreira, B. F. T., Pinto, T. S. S., Starling, D. S. V., & Jaeger, A. (2019). Retrieval Practice in Classroom Settings: A Review of Applied Research. *Frontiers in Education*, *4*(February), 1–16. https://doi.org/10.3389/feduc.2019.00005

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Mueller, R. O., & Knapp, T. R. (2018). Reliability and validity. In *The reviewer's guide to quantitative methods in the social sciences* (pp. 397–401). Routledge.

Nelson, T. O., & Dunlosky, J. (1994). Norms of Paired-Associate Recall During Multitrial Learning of Swahili-English Translation Equivalents. *Memory*, *2*(3), 325–335. https://doi.org/10.1080/09658219408258951

Netemeyer, R., Bearden, W., & Sharma, S. (2003). *Scaling Procedures*. https://doi.org/10.4135/9781412985772

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*(2), 130–159. https://doi.org/10.1037/a0026699

O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, *43*(5), 971–990. https://doi.org/10.1016/j.paid.2007.03.017

Pan, S. C., Pashler, H., Potter, Z. E., & Rickard, T. C. (2015). Testing enhances learning across a range of episodic memory abilities. *Journal of Memory and Language*, *83*, 53–61. https://doi.org/10.1016/j.jml.2015.04.001

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When Does Feedback Facilitate Learning of Words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 3–8. https://doi.org/10.1037/0278-7393.31.1.3

Pérez-González, J.-C., Filella, G., Soldevila, A., Faiad, Y., & Sanchez-Ruiz, M.-J. (2022). Integrating self-regulated learning and

individual differences in the prediction of university academic achievement across a three-year-long degree. *Metacognition and Learning*, *17*(3), 1141–1165. https://doi.org/10.1007/s11409-022-09315-w

Pérez, H. S., Gutiérrez-Braojos, C., & Fernández, S. R. (2017). The relationship of gender, time orientation, and achieving self-regulated learning. *Revista de Investigacion Educativa*, *35*(2), 353–369. https://doi.org/10.6018/rie.35.2.273141

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*(2), 322–338. https://doi.org/10.1037/a0014996

Poropat, A. E. (2014). A meta-analysis of adult-rated child personality and academic performance in primary education. *British Journal of Educational Psychology*, *84*(2), 239–252. https://doi.org/10.1111/bjep.12019

Putnam, A. L., Nestojko, J. F., & Roediger, H. L. (2016). Improving student learning: Two strategies to make it stick. In J. C. Horvath, J. Lodge, & J. A. C. Hattie (Eds.), *From the laboratory to the classroom: Translating the science of learning for teachers* (pp. 94–121). Routledge.

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. https://doi.org/10.1016/j.jml.2009.01.004

Pyc, M. A., & Rawson, K. A. (2010). Why Testing Improves Memory: Mediator Effectiveness Hypothesis. *Science*, *330*(6002), 335–335. https://doi.org/10.1126/science.1191465

Raven, J. C. (1990). *Advances Progressive Matrices: Sets I, II*. Oxford Univ Press.

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, *140*(3), 283–302. https://doi.org/10.1037/a0023956

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological

correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353–387. https://doi.org/10.1037/a0026838

Rickard, T. C., & Pan, S. C. (2018). A dual memory theory of the testing effect. *Psychonomic Bulletin & Review*, *25*(3), 847–869. https://doi.org/10.3758/s13423-017-1298-4

Rivers, M. L. (2021). Metacognition About Practice Testing: a Review of Learners' Beliefs, Monitoring, and Control of Test-Enhanced Learning. In *Educational Psychology Review* (Vol. 33, Issue 3). Educational Psychology Review. https://doi.org/10.1007/s10648-020-09578-2

Robey, A. (2019). The benefits of testing: Individual differences based on student factors. *Journal of Memory and Language*, *108*(May), 104029. https://doi.org/10.1016/j.jml.2019.104029

Roediger, H. L., & Karpicke, J. D. (2006a). 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

Roediger, H. L., & Karpicke, J. D. (2006b). 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

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118–137. https://doi.org/10.1016/j.intell.2015.09.002

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

Ruffing, S., Wach, F. S., Spinath, F. M., Brünken, R., & Karbach, J. (2015). Learning strategies and general cognitive ability as predictors of gender- specific academic achievement. *Frontiers in Psychology*, *6*(August), 1–12. https://doi.org/10.3389/fpsyg.2015.01238

Schneider, M., & Preckel, F. (2017). Variables associated with

achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, *143*(6), 565–600. https://doi.org/10.1037/bul0000098

Schweizer, K., Goldhammer, F., Rauch, W., & Moosbrugger, H. (2007). On the validity of Raven's matrices test: Does spatial ability contribute to performance? *Personality and Individual Differences*, *43*(8), 1998–2010. https://doi.org/10.1016/j.paid.2007.06.008

Schwieren, J., Barenberg, J., & Dutke, S. (2017). The Testing Effect in the Psychology Classroom: A Meta-Analytic Perspective. *Psychology Learning & Teaching*, *16*(2), 179–196. https://doi.org/10.1177/1475725717695149

SFS 2021:1335. (2021). *Förordning om utbildning till lärare och förskollärare*. https://svenskforfattningssamling.se/sites/default/files/sfs/2021-12/SFS2021-1335.pdf

Skolverket. (2022a). *Betyg och studieresultat i gymnasieskolan år 2022*. https://www.skolverket.se/getFile?file=10522

Skolverket. (2022b). *Slutbetyg i grundskolan*. https://www.skolverket.se/getFile?file=10334

Sorić, I., Penezić, Z., & Burić, I. (2017). The Big Five personality traits, goal orientations, and academic achievement. *Learning and Individual Differences*, *54*, 126–134. https://doi.org/10.1016/j.lindif.2017.01.024

Sotola, L. K., & Credé, M. (2021). Regarding Class Quizzes: a Meta-analytic Synthesis of Studies on the Relationship Between Frequent Low-Stakes Testing and Class Performance. *Educational Psychology Review*, *33*(2), 407–426. https://doi.org/10.1007/s10648-020-09563-9

Stenlund, T., & Jonsson, B. (2017). Assessing the Willingness to Elaborate among Young Students: Psychometric Evaluation of a Swedish Need for Cognition Scale. *Frontiers in Education*, *2*(2). https://doi.org/10.3389/feduc.2017.00002

Stenlund, T., Jönsson, F. U., & Jonsson, B. (2017). Group discussions and test-enhanced learning: individual learning outcomes and personality characteristics. *Educational Psychology*, *37*(2), 145–

156. https://doi.org/10.1080/01443410.2016.1143087

Stenlund, T., Sundström, A., & Jonsson, B. (2016). Effects of repeated testing on short- and long-term memory performance across different test formats. *Educational Psychology*, *36*(10), 1710–1727. https://doi.org/10.1080/01443410.2014.953037

Strobel, A., Behnke, A., Gärtner, A., & Strobel, A. (2019). The interplay of intelligence and need for cognition in predicting school grades: A retrospective study. *Personality and Individual Differences*, *144*(July 2018), 147–152. https://doi.org/10.1016/j.paid.2019.02.041

Thorndike, E. L. (1914). Communications and discussion. Repetition versus recall in memorizing vocabularies. *Journal of Educational Psychology*, *5*(10), 596–599. http://psycnet.apa.orgjournals/edu/5/10/596

Tse, C.-S., Chan, M. H.-M., Tse, W.-S., & Wong, S. W.-H. (2019). Can the Testing Effect for General Knowledge Facts Be Influenced by Distraction due to Divided Attention or Experimentally Induced Anxious Mood? *Frontiers in Psychology*, *10*, 969. https://doi.org/10.3389/fpsyg.2019.00969

Tse, C.-S., & Pu, X. (2012). The effectiveness of test-enhanced learning depends on trait test anxiety and working-memory capacity. *Journal of Experimental Psychology: Applied*, *18*(3), 253–264. https://doi.org/10.1037/a0029190

Tullis, J. G., & Maddox, G. B. (2020). Self-reported use of retrieval practice varies across age and domain. *Metacognition and Learning*, *15*(2), 129–154. https://doi.org/10.1007/s11409-020-09223-x

Tulving, E. (1993). What Is Episodic Memory? *Current Directions in Psychological Science*, *2*(3), 67–70. https://doi.org/10.1111/1467-8721.ep10770899

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/BF03192720

van den Broek, G., Takashima, A., Wiklund-Hörnqvist, C., Karlsson Wirebring, L., Segers, E., Verhoeven, L., & Nyberg, L. (2016).

Neurocognitive mechanisms of the “testing effect”: A review. *Trends in Neuroscience and Education*, *5*(2), 52–66. https://doi.org/10.1016/j.tine.2016.05.001

Veltre, M. T., Cho, K. W., & Neely, J. H. (2015). Transfer-appropriate processing in the testing effect. *Memory*, *23*(8), 1229–1237. https://doi.org/10.1080/09658211.2014.970196

von Stumm, S., & Ackerman, P. L. (2013). Investment and intellect: A review and meta-analysis. *Psychological Bulletin*, *139*(4), 841–869. https://doi.org/10.1037/a0030746

Weis, M., Heikamp, T., & Trommsdorff, G. (2013). Gender differences in school achievement: The role of self-regulation. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00442

Wenzel, K., & Reinhard, M. A. (2019). Relatively unintelligent individuals do not benefit from intentionally hindered learning: The role of desirable difficulties. *Intelligence*, *77*(October), 101405. https://doi.org/10.1016/j.intell.2019.101405

Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, *11*(6), 571–580. https://doi.org/10.1080/09658210244000414

Wiklund-Hörnqvist, C., Jonsson, B., & Nyberg, L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology*, *55*(1), 10–16. https://doi.org/10.1111/sjop.12093

Wiklund-Hörnqvist, C., Stillesjö, S., Andersson, M., Jonsson, B., & Nyberg, L. (2022). Retrieval Practice Is Effective Regardless of Self-Reported Need for Cognition - Behavioral and Brain Imaging Evidence. *Frontiers in Psychology*, *12*(February), 1–10. https://doi.org/10.3389/fpsyg.2021.797395

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, *147*(4), 399–435. https://doi.org/10.1037/bul0000309

Yang, C., Sun, B., Potts, R., Yu, R., Luo, L., & Shanks, D. R. (2020). Do working memory capacity and test anxiety modulate the beneficial effects of testing on new learning? *Journal of Experimental Psychology: Applied*, *26*(4), 724–738.

https://doi.org/10.1037/xap0000278

Zheng, Y., Sun, P., & Liu, X. L. (2023). Retrieval practice is costly and is beneficial only when working memory capacity is abundant. *Npj Science of Learning*, *8*(1), 1–7. https://doi.org/10.1038/s41539-023-00159-w

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2