SWPS University

Ewa Butowska

The harder the better? Learning by errors

PhD Thesis

Supervisor: dr hab. Katarzyna Zawadzka, prof. USWPS

Warszawa, 2023

Table of contents

List of Figures

List of Tables

Abstract

Learning by guessing which involves a guessing attempt before the correct answer is revealed has recently been proposed as an effective, readily implemented learning strategy compared to reading – where learners are presented with the full study material outright and are not asked to guess at the answer first. So far, guessing has been shown to improve learners' ability to associate the study items with each other, for instance learning that the word "pond" was presented together with "frog" during study. However, this benefit was observed only when studying semantically related materials such as related word pairs or trivia questions and their answers. When studying unrelated materials such as foreign words and their translations, the guessing benefit seemed to be limited to item memory, which reflects the memory for studied words without associating them with their foreign counterparts (remembering studying "ongelma" and "problem", but not the fact that they were presented together at study, and thus mean the same thing). In Study 1, we investigated whether guessing can also benefit memory for associations when studying foreign translations, which is key in language acquisition as learners need to associate a new foreign word with its specific translation. More specifically, we tested if adding a restudy phase would benefit translations previously learnt in the guess condition more than those studied in the read condition. Study 2 introduced contextual sentences with embedded foreign words which varied the degree of prediction error: some sentences were highly suggestive but also misleading of the meaning of the foreign word (i.e. resulted in high prediction error) and others were ambiguous and not suggesting any specific meaning (i.e. resulted in low prediction error). None of these two manipulations revealed the hypothesized guessing benefit over reading evident in enhanced associations' between foreign words and their meanings. Instead, we showed dissociations between the effects of guessing and restudying and also between guessing and learning via prediction error. Study 3 changed study materials to trivia questions and investigated the role of timing in driving the differences between guessing and reading. The results underlined the role of the time spent with the question and showed no guessing performance advantage over reading once the learning duration was equated. What is more, guessing decreased performance when studying unfamiliar novel responses and was comparable to reading when studying more familiar, previously known responses to the questions. Additionally, guessing at a related question did not help to learn the following test questions, directly undermining the role of response generation in guessing benefits over reading. In short, the current results discourage from using guessing as an effective strategy as its benefits seem to be restricted to few types of learning materials with little educational relevance.

Keywords: Guessing · Errorful learning · Desirable Difficulties · Restudy · Item memory · Associative memory · Prediction error · Trivia · Education

Streszczenie

Uczenie się przez zgadywanie, które polega na podjęciu próby odgadnięcia poprawnej odpowiedzi przed jej ujawnieniem, zostało niedawno zaproponowane jako skuteczna, łatwa do wdrożenia strategia uczenia się w porównaniu z czytaniem - gdzie uczący się otrzymują pełny materiał do nauki od razu i nie są proszeni o odgadnięcie odpowiedzi. Jak dotąd wykazano, że zgadywanie poprawia zdolność badanych do asocjowania ze sobą części nabywanego materiału, na przykład uczenia się, że słowo "staw" było prezentowane razem ze słowem "żaba" podczas nauki. Jednak korzyści te zaobserwowano tylko podczas studiowania semantycznie powiązanych materiałów, takich jak powiązane pary słów lub pytania z wiedzy ogólnej i odpowiedzi na nie. Podczas studiowania niepowiązanych semantycznie materiałów, takich jak obce słowa i ich tłumaczenia, korzyść ze zgadywania wydaje się być ograniczona do pamięci uczonych słów bez asocjowania ich z ich obcymi odpowiednikami (pamiętanie słów "ongelma" i "problem", ale nie tego, że były one zaprezentowane razem, a zatem oznaczają to samo). W badaniu 1 przestudiowaliśmy, czy zgadywanie może również korzystnie wpływać na pamięć asocjacji podczas studiowania tłumaczeń obcych słów, co jest kluczowe w przyswajaniu języka, ponieważ uczący się muszą powiązać nowe obce słowo z jego konkretnym tłumaczeniem. Dokładniej rzecz ujmując, sprawdziliśmy, czy wprowadzenie możliwości ponownego przestudiowania materiału przyniosło większe korzyści parom tłumaczeń wcześniej uczonym poprzez zgadywanie niż tłumaczeniom uczonym poprzez czytanie. W badaniu 2 obce słowa umieściliśmy na końcu zdań w języku ojczystym, które dostarczały kontekst uczeniu się obcych słów. Zdania te różniły się stopniem błędu predykcji: niektóre zdania były wysoce sugestywne, ale także wprowadzały w błąd co do znaczenia obcego słowa (tj.

skutkowały wysokim błędem predykcji), a inne były niejednoznaczne i nie sugerowały żadnego konkretnego znaczenia obcego słowa (tj. skutkowały niskim błędem predykcji). Żadna z tych manipulacji nie wykazała przewidywanej przewagi zgadywania nad czytaniem, widocznej we wzmocnionej pamięci powiązań między obcymi słowami a ich znaczeniem. Zamiast tego wykazaliśmy, że efekty uczenia się przez zgadywanie i ponownego uczenia się, a także efekty zgadywania i uczenia się poprzez błąd predykcji są od siebie niezależne. W badaniu 3 zmieniliśmy materiały do nauki na pytania z wiedzy ogólnej i zbadaliśmy rolę czasu w kreowaniu różnic pamięciowych między zgadywaniem a czytaniem. Wyniki podkreśliły rolę czasu przeznaczonego na naukę pytań i nie wykazały przewagi zgadywania nad czytaniem po zrównaniu czasu trwania nauki. Co więcej, zgadywanie obniżyło efektywność uczenia się nieznajomych odpowiedzi i było porównywalne z czytaniem podczas studiowania bardziej znajomych, wcześniej kojarzonych odpowiedzi na pytania. Dodatkowo, próba odpowiedzi na powiązane pytania nie pomagała w nauce pytań testowych, co bezpośrednio podważa rolę generowania odpowiedzi w efektywności zgadywania w stosunku do czytania. Podsumowując, obecne wyniki odradzają stosowanie uczenia się przez zgadywanie, ponieważ jego korzyści wydają się być ograniczone do materiałów o niewielkim znaczeniu edukacyjnym.

Słowa kluczowe: Zgadywanie · Uczenie się przez popełnianie błędów · Pożądane trudności · Ponowne uczenie się · Pamięć bodźca · Pamięć asocjacji · Błąd predykcji · Pytania z wiedzy ogólnej · Edukacja

Introduction

Learning Strategies: An Overview

The learning process is one of most important and essential in human lives. Therefore, it is no surprise that a question of how to optimize one's learning has been of considerable interest to researchers for decades (e.g. Atkinson, 1972). In today's world especially, the issue becomes extremely valid. Many people face requirements of life-long learning, stemming from professional and personal circumstances. Also, rapidly evolving technologies and interfaces used in everyday lives increase the toll on our learning ability. Another circumstance requiring learning may be travelling or working in an international environment, which is inherently combined with acquiring new languages in order to communicate effectively. Therefore, the potential benefits of establishing the most effective ways of acquiring new information are quite evident – they would translate into a high quality and effective educational system, would be desirable in professional training, travelling, and – what is of particular relevance nowadays – they could improve the quality of life of an aging population, which is characterized by increased rates of forgetting (Huppert & Kopelman, 1989).

Learning can be approached in a variety of ways as there is no one fixed way of acquiring information. On the contrary, learners choose how they want to study from a spectrum of different options, methods, techniques, and approaches. Crucially, not all learning approaches are created equal – some result in better memory than the others. For example, one of the most effective learning approaches is testing (also referred to as retrieval practice, for review see Roediger & Karpicke, 2006) which, after an initial study phase, requires learners to retrieve studied material from memory rather than re-presents it. It has been shown that such retrieval practice is more effective than being presented with the study material for additional restudy sessions (Karpicke & Roediger, 2007). This effect is usually observed on a delayed test, i.e., not immediately after the study session – when restudying can result in better memory performance, but after a time lag – ranging from several minutes to several days – in which case the results are reversed and retrieval practice outperforms the shorter-lasting beneficial effects of restudying (e.g. Roediger, et al., 2011). This delayed benefit of testing is of particular educational relevance as the final test is usually preceded by a substantial period of time from the last study session, which makes testing a very practical learning tool.

Another well-researched learning strategy is spacing - which refers to distributing several study sessions over time instead of cramming those into a single learning episode. It turns out that spacing, which requires more effort to recall previously studied material, benefits memory compared to having one long learning session (Benjamin & Tullis, 2010). Also, increasing the lags between study sessions has an impact on this strategy's effectiveness and takes form of a nonmonotonic function – initially increasing the gaps between study sessions sharply increases final performance; however, at some point, further increases are associated with a gradual decline in performance (Cepeda et al., 2009).

Another learning strategy deemed effective is elaboration, which can be described as trying to make connections between various topics, figuring out how they are similar or different, which can also be defined as aiming to understand the material rather than just memorize it 'by heart' (McDaniel & Donnelly, 1996). What is also important in this strategy is describing how the studied material is related to the already existing knowledge base (Wong, 1985). Elaboration may also require learners to answer the question of *why* studied phenomena occur, or *why* somebody did something, which increases memory for the studied information compared to simply reading it without answering such elaborative interrogations (Pressley et al., 1987).

In schools students are often required to learn the so-called declarative concepts. They are usually abstract and denoted by some key terms, which relate to a variety of real-life scenarios such as e.g. positive reinforcement in psychology or diminishing returns in economy. When studying these abstract definitions, connecting them with various concrete examples (e.g., giving somebody a raise for good work performance as an example of positive reinforcement or spending much more time than necessary on a given task as an example of diminishing returns) results in better ability to later correctly match novel examples with appropriate declarative concept than spending additional time studying the definitions alone (Rawson et al., 2014). Thus, using concrete examples when studying abstract phenomena is also considered a good, effective learning approach.

Another effective learning strategy is dual coding which employs more than just one modality and instead relies on combining words with pictures. Mayer and Anderson (1992) showed that presenting simultaneously oral and visual narratives of how to mend a bike tire resulted in better performance on a subsequent problem solving test compared to a successive presentation of an animation and narration, or to presenting just the animation or just narration. The authors concluded that "pictures and words are most effective when they occur contiguously in time or space" (Mayer & Anderson, 1992) which suggests that for learning to be successful, the materials should be represented in more than just one modality.

Desirable Difficulties

In more traditional approaches to learning, the most common one was to avoid making errors while studying new information or learning new skills (Fillingham et al.,

2003) – the so called ‘errorless learning’. This approach is still quite prevalent, especially among clinical populations undergoing cognitive rehabilitation (Middleton & Schwartz, 2012), or young children (Warmington et al., 2013). The main premise of this approach lies in an observation that errors committed during the learning process may later interfere with the to-be-remembered, correct information. Thus, when a specific cue is presented at test, it may potentially be associated both with its correct response (the target), and also with the erroneous response produced during the learning phase, which increases the retrieval competition among these items and further translates into a decrease in memory performance (Anderson & Neely, 1996). By simply avoiding committing errors as part of learning experience in the first place, those negative phenomena should be less likely to occur.

However, more recent research has shown that there is more to learning than avoiding errors and trying to make the learning process as smooth as possible. Quite the contrary – recent findings suggest that making the study session more challenging may be desired in terms of the positive outcome it provides on the final test (Bjork & Bjork, 2011). These added challenges or difficulties are deemed ‘desirable’ as they engage encoding and retrieval processes which support learning and comprehension of the study material, especially when long-term retention is tested (Bjork et al., 2013). Thus, the concept of ‘desirable difficulties’ implies that learning is more effective when it requires effort and engages participants in more active study mode compared with passive and disengaged learning, which may accompany more traditional presentation of the learning materials. Learning strategies which are most often referred to as desirable difficulties in the literature will be described shortly below.

One desirable difficulty that can be applied in educational settings is variable practice, which can be described as approaching the study topic from different perspectives and trying to learn about it in different contexts as opposed to focusing on one single aspect of the studied topic. Probably the most famous and striking illustration of the benefits of varied practice was obtained by Kerr and Booth (1978), who asked children to practice throwing beanbags at a target with their eyesight occluded. Half of them practiced throwing repeatedly from the same fixed distance, and another half varied the distance be several feet closer or farther to the target. Later, all children were tested using the same distance as practiced by the first group which should intuitively favor it over the second one as they repeatedly practiced throwing at the exact same distance. The results were, however, quite the opposite – the latter, varied-practice group outperformed the constant practice group and was better able to throw at the target. Other instances of varying the practice conditions include e.g., studying in two different rooms rather than twice in the same (Smith et al., 1978) or studying words cued by different rather than the same orienting tasks (Zawadzka et al., 2021). The main mechanism by which varied learning results in better test performance is that the learned skill or material becomes less context dependent compared to skills or materials practiced in a repeated, redundant context, which makes the material studied in various contexts easier to recall on a final test (Di Vesta & Peverly, 1984).

Another desirable difficulty is the already mentioned spacing of the study sessions in time as opposed to massing those into one long episode of learning. The advantage of spread-out learning episodes has been demonstrated with a number of learning materials (see Carpenter et al., 2012, for a review) and generally showed better performance on the final test compared to massing. What is more, the effects of spacing are much more long-lasting compared to very short-lived memory resulting from massing (Bjork & Bjork, 2011). In other words, when students cram all the materials into one study session the night before the exam, they may obtain a passing grade, but they are unlikely to remember much a week or two afterwards. In contrast, those students who would spread their study throughout the whole academic year are much more likely to retain the information for long after the test.

One more desired difficulty which can be easily implemented in classroom environment is interleaving which can be described as studying several different topics or skills within one study session. Interleaving as opposed to blocking practice assumes mixing different to-be-learnt materials rather than grouping those together. It was shown to be effective when learning motor skills (Shea & Morgan, 1979) and also mathematical formulas (Rohrer & Taylor, 2007) or artistic styles of various famous painters (Kornell & Bjork, 2008). In the latter study, participants were presented with six pictures by 12 famous artists. Interleaving a given painter's pictures among other pictures resulted in a better ability to correctly identify the artist's new paintings compared to presenting this artist's works in one go before switching to the next painter. The positive effect of interleaving is often explained by the necessity to 'reload' memories each time the learning material is changed as such 'reloads' enhance learning, especially at a later time (Bjork & Bjork 2011). Conversely, when some skill or information is already highly accessible (as in blocked practice), additional repeated practice will result in less effective learning.

The last desired difficulty as described by Bjork and Bjork (2011) – and, arguably, the one that has received the most attention in the last two decades – is the already-mentioned testing (also known as retrieval practice). This concept refers to active retrieval of an answer to a question from memory as opposed to being passively presented with one. It has been argued that testing is an effective learning strategy because when people attempt to retrieve the correct answer, a prior learning context is reinstated and when the retrieval attempt is successful – the context representation is updated to include the features of both current as well as retrieved learning context. Such updated context representation can later help to limit the search set when attempting to recall the response and direct the memory search towards the target (Karpicke et al., 2014).

Still, the learners should have some cognitive potential and basic knowledge structure to be able to make use of all of these challenging learning conditions effectively, or else they may make the study session too challenging up to a point at which the added difficulties such as testing, spacing or interleaving, may become undesired and hinder future memory outcome. Thus, when researching new challenging learning strategies it is important to make sure that the level of difficulty they present is actually beneficial rather than harmful to memory performance, especially at longer time intervals.

Metacognitive Illusions

One of the main reasons why people do not spontaneously use desired difficulties while learning stems from the fact that these strategies *appear* ineffective to learners compared with those stripped of the added challenge like, e.g., reading the study materials in a straightforward fashion. The discrepancy between the actual and subjective state of our memory is described in the literature on metamemory, defined as the individual's knowledge of and awareness of own memory (Flavell & Wellman, 1975). According to Koriat’s (1997) cue-utilization framework, people base their judgments of learning, i.e., predictions of memory performance on the upcoming or completed test, on different types of cues referring to study conditions, and also the difficulty and novelty of the learning material. One of the most powerful cues we use to determine how well we have learnt some given material is the processing fluency during encoding (e.g. Finn & Tauber, 2015). In general, less challenging study conditions will lead to more fluent processing. As people confuse this feeling of fluency with the feeling of mastery of the given study materials – they tend to choose less challenging (and often less effective) study conditions instead of making use of desirable difficulties.

What is more, such metacognitive illusions are quite prevalent and not easy to mend – people quite consistently predict higher performance in the less effective learning conditions. For instance, learners prefer reading over testing (Karpicke & Roediger, 2008), massing over spacing (McCabe, 2011), or blocking over interleaving (Kornell & Bjork, 2008). Also, these illusions are hard to overcome, as participants continue to undervalue more effective learning conditions even after having experience with the final test which evidenced the positive effect of desirable difficulties on performance, unless given substantial guidance to appreciate those challenging yet effective strategies (Tullis et al., 2013). In other words, less effective strategies such as blocking, massing, or reading seem to people to be optimal for fluent learning, but in fact they yield inferior long-term performance and people tend to confuse the fluency of learning experience afforded by those less effective strategies with learning effectiveness (Bjork & Bjork, 2011). As a result, people are more likely to choose and stick with less effective learning approaches, which has a negative result on their future performance. This only further motivates the research into the mechanisms and effectiveness of desirable difficulties because students are not likely to spontaneously choose and develop them when engaging in self-study.

Errorful Learning

As already mentioned throughout the thesis, one of the most researched and effective learning strategies is testing. It involves an initial study phase which is later replaced with retrieval practice of studied material rather than with additional time to reread it. Interestingly, this retrieval does not need be successful to improve memory performance (Kornell et al., 2015). This means that when the retrieval attempt fails and participants produce no response, learning of new information is still enhanced (Slamecka & Fevreiski, 1983). What is more, even when participants produce an error when engaging in the retrieval process, the learning of corrective feedback – i.e., the correct response – will be more effective compared to when no retrieval was attempted (Kornell et al., 2009; Yan et al., 2014). This phenomenon – often referred to as errorful learning – is crucial for educators who face the problem of students' reluctance to err in their responses. It also directly contradicts the tradition of an errorless approach to learning in suggesting that to err is not only human but also highly beneficial in terms of supporting learning and thus should be actively encouraged by instructors (Wong & Lim, 2022). The main focus of the current thesis will thus be on errorful learning.

The current project aimed at further investigating a relatively novel approach based on increasing learning difficulty in such a way that it would lead to a long-term performance gain. It was grounded in the errorful approach to learning and followed a recent surge in the studies on the effectiveness of learning by guessing. One way of introducing errors is having people guess an answer to a question before presenting them with correct answer, as in such a case the majority of guesses tend to be incorrect. This strategy implies committing errors as opposed to the more traditional approach termed errorless learning. The two approaches – errorful and errorless learning – differ substantially in their predictions. The premise of the former is that producing an error – if it is somehow related to the desirable answer – has a potential to aid memory because the generation of the answer, even if it is not correct, can enhance the encoding of the following feedback. The premise behind the latter is that if an error is not produced, it will not be remembered or it will not interfere with the correct answer. The present project takes as its starting point the errorful learning approach, given the plethora of recent research – described below – suggesting its utility for effective learning, and aims to provide a test of how learning by errors affects memory.

The theoretical base for the current project follows from a set of findings by Kornell et al. (2009). In their seminal study, the authors asked participants to memorize weakly associated pairs of words (e.g. *pond-frog*). In one learning condition, only the first word appeared and participants were instructed to guess and type in what the second one might be for 8 seconds (thus, when presented with the word *pond*, they could type in e.g., *water, lake, fish*, etc.). After the guessing stage, corrective feedback was delivered in the form of the full pair (i.e., *pond-frog*) for 5 seconds. A control read condition simply presented complete pairs (*pond-frog*) straightaway for 13 seconds, hence the learning trials in the two learning conditions were equally long. Importantly, guessing the second word resulted in better memory compared to studying the full pairs, as measured by the final memory test. When presented only with the first of the associated words (*pond*) at test – people were better able to recall the second associated word (*frog*) when they previously guessed at its identity compared to when they were presented with it outright.

When studying guessing, one key issue is the format of the test used to assess strategy effectiveness. In general, in studies on guessing effectiveness there are two types of memory representations that can be investigated and that are measured by different test formats: *item memory* and *associative memory*. The former one refers to the memory of what was studied: for example, remembering that the words *pond* and *frog* were presented in the study phase. The latter one refers to the memory for the associations between studied materials: remembering that the words *pond* and *frog* were studied *together*. The test format most commonly used to reflect item memory is simple recognition which presents studied items intermixed with lures – which are here novel, previously unseen items – and requires learners to recognize the studied words as such and reject the lures. In order to perform well on a recognition test, learners simply need to correctly identify which items were shown in the study phase – hence the test reflects item memory.

The test formats most commonly used to reflect associative memory are cued recall and associative recognition. In cued recall tests, learners need to produce a specific word in response to a presented cue. For instance, when studying weakly-related word pairs – one word from a previously studied pair is presented as cue on test and learners need to recall and type in its specific word associate (the target). Here, in order to perform well, learners need to firstly remember what words were studied and also associate them with their specific cues to produce the correct response for a given cue. In the associative recognition test – learners are also presented with the cue and are instructed to pick its associate from a set of lures, which are here other studied words, but which were paired with different cues. Therefore, in order to perform well here – learners also need to form an association between specific cues and their associates – or else they will not be able to choose the correct response because all the lures were present in the study phase and thus are equally familiar. Thus, cued recall and associative recognition tests are suited for investigating associative memory, as their performance is determined by whether learners can form associations between studied items.

As already mentioned, guessing has been shown to improve memory for associations when studying weakly-related word pairs (Kornell et al., 2009). What is more, similar results have been obtained with other types of materials – such as general-knowledge or trivia questions with their answers (Kornell, 2014). Thus, trying to incorrectly guess an answer to a trivia question, before being presented with a corrective feedback, resulted in a better ability to correctly answer the question at test compared to being presented with the correct response outright, as was the case in the read condition. Guessing at questions concerning a passage from a book before reading it has also been shown to increase participants' performance on the final test (consisting of the same questions) compared to being allowed more time for reading the passage (Richland et al., 2009). Importantly, this means that error-prone guessing provides clear benefits compared to merely reading the to-be-learned materials. The fact that this effect has been demonstrated with such a variety of learning materials and by numerous researchers, gives us a fair premise to further pursue research on this learning strategy. .

Notably, the guessing strategy has a great educational potential. It does not require costly investments and a lot of time and resources to implement. If properly understood, it could become a very useful and practical learning strategy, easily implemented into classroom environments. However, in this context it is important to note that guessing also has its limitations. Without understanding when and why the effectiveness of guessing is limited, implementing this strategy in classrooms might be premature. Thus, the present series of experiments will focus on the known limitations of the effectiveness of guessing, trying to understand why they occur and to find ways to overcome them.

One of the caveats to the effectiveness of guessing as a learning strategy refers to the semantic relationship between the cue (first word or question) and the target (second word or answer). The already described paradigm used by Kornell et al. (2009) employed related pairs of words. Crucially, when unrelated words were used (e.g.*, table-sky*), no benefit of guessing – or even a cost of guessing – occurred on the subsequent cued recall test (Huelser & Metcalfe, 2012; Knight et al., 2012; Seabrooke et al., 2021). This is an important limitation to guessing effectiveness and suggests that this learning strategy does not enhance associations between studied items when they are semantically unrelated.

Another crucial limitation refers to temporal aspects of the learning process. More specifically, the timing of feedback presentation determines whether guessing is an effective learning strategy. Vaughn and Rawson (2012) compared two learning conditions using weakly-related word pairs as stimuli – when the correct answer was presented right after the guessing attempt, and when the guessing attempts and feedback presentations stages were blocked, so that the participants were first asked to guess at a series of words and were not presented with the feedback until completing the whole study list; this blocking resulted in a delay of several minutes being introduced between the guessing attempt and feedback presentation. Only in the former condition, i.e., when the feedback was immediate, did guessing lead to better memory performance than reading, which suggests that for guessing to be effective, the corrective feedback needs to be presented directly after the guessing attempt.

What is more, the very process of guessing also has to meet some criteria to be effective. For instance, Bridger and Mecklinger (2014) asked participants to guess at words which were denoted by a three-letters long stem, such as e.g. bro___? (thus, participants could type in *brother, broom,* etc.). Such guessing process, constrained by the first three letters of the stem, resulted in worse memory for the correct words (as dictated by the stems) compared with simply reading the words without making incorrect guessing attempts during learning.

These mixed findings concerning the effectiveness of guessing led to the proposition of specific mechanisms responsible for the associative benefit of guessing, which operate only when certain boundary learning conditions are met. According to Grimaldi and Karpicke (2012), a guessing attempt activates a set of semantically related concepts. For instance, guessing at the word *pond* will activate a set of semantically related words such as *lake, water, fish* etc. Similarly, guessing at trivia questions such as e.g.: *What color will red and blue make?* will activate a set of potential semantically related answers such as *purple, pink, orange* etc. Such activation of related semantic concepts has thus been offered as the main mechanism of guessing effectiveness. The search set account of guessing proposes that the semantic activation of relevant concepts spreads also to the correct response which allows for its enhanced encoding upon feedback presentation (Grimaldi & Karpicke, 2012). In other words, even when as a result of the guessing attempt an incorrect response is given, the correct one also becomes activated as it belongs to the same semantic category, and this pre-activation will enhance its processing upon corrective feedback presentation. Therefore, the associative benefits of guessing are limited to studying semantically related materials. When studying unrelated word pairs like e.g. *table-sky*, guessing at the first word will not activate the relevant semantic network, including the target word. Quite the contrary - only when the cue is somehow related to the target will the guessing attempt at the cue benefit target encoding. This is why the guessing benefits via the mechanism of semantic activation can only be obtained when studying materials which have some intrinsic semantic connection between the cues and target.

The search set account of guessing effectiveness by Grimaldi and Karpicke (2012) readily explains the above-mentioned limitations of guessing. Foremost, it aligns with the multiply replicated results of poor guessing performance when studying unrelated word pairs (e.g. Huelser & Metcalfe, 2012; Knight et al., 2012). The worse performance after guessing (compared to reading) on the final cued recall test results from the fact that guessing at word associates which are not related cannot activate the correct response which in turn does not allow for its enhanced encoding upon feedback presentation.

What is more, the account explains a lack of guessing benefit with delayed feedback as reported by Vaughn and Rawson (2012). That is because the semantic activation of potential responses when guessing at simple materials such as related word pairs is based on priming processes which are rather short-lived and can benefit target encoding only when it is presented within a narrow time-window (Kornell, 2014). Interestingly, Zawadzka et al. (2023) showed that guessing with delayed feedback can be effective, provided that the semantic activation which accompanied the guessing attempt can be somehow reconstructed at the feedback presentation stage. The authors achieved such semantic reactivation of cue-related concepts by presenting feedback in the same context as the guessing attempted by placing the relevant word pairs on identical photo backgrounds. Thanks to such manipulation, guessing with delayed feedback outperformed read condition on the final cued recall test.

Finally, the account accords with the null effect of guessing obtained by Bridger and Mecklinger (2014) who constrained participants' guessing by giving them word stems. Such constrained guessing should not activate a broader semantic network, which can later help to encode corrective feedback and increase performance. That is because participants focus solely on words starting with specific three letters, and when they find one – they will produce it as their guess, which is programmed to be incorrect. In other words, when the guessing attempt is too constrained and incorrect, activation from it cannot spread to include the right response and so will fail to increase the final performance.

Errorful Learning of Foreign Vocabulary

The search set account underlines the role of pre-existing semantic relationship between the cue and its target in producing the benefits of guessing. However, somewhat contrary to the already described limitations of guessing – it was also observed that guessing can benefit performance when studying unrelated materials (Potts & Shanks, 2014). More specifically, in a series of experiments Potts and Shanks asked participants to study either definitions of novel, unfamiliar English words (e.g. *hispid-bristly, valinch-tube, frampold-quarrelsome*) or unfamiliar Euskara translations (*igel-frog, urmael-pond, untxi-rabbit*) which both were instances of novel learning where participants did not know the answers and also were very unlikely to correctly generate one during the guessing attempt. What is more, the incorrect responses were unlikely to be related to the targets, which, according to the search set account (Grimaldi & Karpicke, 2012) is key to observe the guessing benefits. The learners studied the unfamiliar words by either guessing at the response followed by corrective feedback, by reading the full translation outright, or by selecting from a choice of translations followed by feedback. The final recognition test revealed that generating incorrect responses followed by feedback led to higher memory for the correct answer than both reading and making incorrect choices. These results seemed truly groundbreaking inasmuch as they demonstrated that guessing can be effective when learning not only semantically related materials, in which case incorrect guessing can at least tap onto the identity of the target, but also when guessing is truly uninformed by semantics and produced responses do not relate to the correct response. What is more, the authors showed that people were unaware of this beneficial effect of guessing and judged this strategy as least effective (compared to reading and choosing correct translation from given candidates). This suggest that people do not spontaneously chose to learn by guessing and that this potentially effective yet underrated strategy may need promoting as such so that more people's learning can start benefiting from it.

The already-mentioned search set account struggles to account for such results as the studied materials possessed no intrinsic binding relationship between the cues and targets which can inform the guessing process to activate relevant responses. For instance, if somebody is presented with a an Euskara word such as *unxti* and is asked to guess at its translation, the semantic activation accompanying such a guessing attempt will in all likelihood not include the correct translation (*rabbit*), as for the learners with no previous knowledge of Basque language the relationship between the words *unxti* and *rabbit* is opaque. The question remains, how such benefits of guessing over reading emerged when studying unfamiliar translations?

As underlined earlier, the test format is of substantial relevance when studying learning by guessing. Namely, it is of note that the benefit of guessing over reading as reported by Potts and Shanks (2014) was obtained with a simple recognition measure. This means that guessing here enhanced merely the memory for the studied words. Such item memory benefit of guessing is explained by different mechanisms than its associative benefit when studying related materials. The mechanism does not rely on semantic activation processes and therefore can be postulated to operate regardless of whether the study materials are semantically related or not. Namely, guessing was proposed to increase curiosity regarding what the answer may be compared to reading, where the correct response is provided right away (Potts & Shanks, 2019). This curiosity was postulated to increase attention to the corrective feedback, which in turn would make its encoding more effective (Overman, et al., 2021). Thanks to the fact that guessing is associated with more attention devoted to feedback, regardless of whether it is related to the cue or not – the item benefits of this learning strategy seem universal and not determined by the semantic relationship binding the study items. Therefore, guessing at a foreign translation or unfamiliar words can improve memory for the studied words.

After the initial enthusiasm brought to research on learning by Potts and Shanks (2014), who were the first to show that guessing can benefit learning of novel, previously unknown words, a more thorough investigation of the nature of such benefit was bound to take place. As already mentioned – Potts and Shanks evidenced a guessing benefit for item memory, which is not informative of what is key in foreign vocabulary learning – the associations between the foreign words and their meanings. Thus, Seabrooke et al. (2019) compared guessing effectiveness when learning rare English words and Euskara vocabulary for two different memory types:

item and associations memory within one experimental design. Item memory was measured by a recognition test which used novel, previously unstudied words as lures and showed the guessing benefit over reading, replicating the results obtained by Potts and Shanks (2014). Associations memory was measured by cued recall and also by an associative recognition test, in which the lures consisted of previously studied translations. Thus, in order to score well on this test, learners had to form specific association between given translation and their respective foreign words, as the lures were all presented during the learning phase, only with different foreign counterparts. These associative tests, however, revealed no significant benefit, and even costs of guessing over reading the full pairs outright. The authors concluded that incorrect guessing at words which do not have preexisting semantic associations with their counterparts strengthens the studied words but, crucially, does not strengthen the associations between them and their translations.

This dissociation within the benefits of guessing when studying novel, unfamiliar materials – i.e., memory improvement when item memory is measured and no such improvement in memory for associations – can be readily explained by the two proposed accounts of guessing effectiveness. As the search set account (Grimaldi & Karpicke, 2012) suggests, for the associative benefits of guessing to emerge, the cues need to be capable of activating the target at the guessing attempt, or otherwise the association between those will not be enhanced. Guessing at cues which have no intrinsic relationship with their targets such as foreign words and their translations will therefore be futile in terms of enhancing this association. This is likely why Seabrooke et al. (2019) did not observe the associative benefits of guessing when studying unfamiliar foreign translations and obscure English words. However, for the item memory benefit to emerge, no semantic, pre-experimental relationship needs to exist between the studied items, as the mechanisms here rely not on semantic elaboration but rather on attentional up-regulation which accompanies the guessing attempt (Overman et al., 2021). Therefore, guessing at a foreign translation, even when these guesses are far from the correct answers, can still enhance the memory for the studied words, as evidenced by Potts and Shanks (2014) and also by Seabrooke et al. (2019) when the recognition test (i.e., sensitive to item memory) was used.

A similar demonstration of the two mechanisms of guessing – one operating where there is a semantic relationship between the cues and targets, and another operating regardless of such relationship has been presented in a study by Zawadzka and Hanczakowski (2019), which used homographs as study materials. The participants were asked to guess at the pair associates of word such as e.g. *arms* – which could be interpreted as a body part or as referring to weapons. Although associations memory was enhanced (compared to reading the full pairs) only when participants' guess tapped the same meaning of the cue as the later presented target (e.g., a participant guessed *hands* when the correct target was *legs*), item memory increased even when the guess did not semantically overlap with the target, i.e. when participants incorrectly interpreted the ambiguous cue as belonging to a different semantic category (e.g., a participant guessed *military*). Thus, whereas for the associative benefits of guessing to occur – the studied material needs to be semantically related, the item memory benefit will occur regardless of whether the studied materials are semantically related or not.

The Present Project

In the current project we decided to investigate the mechanism of guessing, which involves an unsuccessful guessing attempt before the correct feedback is presented and which was shown to benefit memory performance compared to simply reading the full material outright (Kornell et al., 2009). In doing so we follow up on the recent studies which promote guessing as an effective learning strategy and encourage educators to implement it in their teaching practice. For instance, Kornell et al. (2009) concluded that that “educators and learners should introduce tests as learning events, even if doing so increases initial error rates” (p. 997), Richland et al. (2009) stated “tests can be valuable learning events, even if learners cannot answer test questions correctly” (p. 254), Potts and Shanks (2014) concluded that “In a more educationally relevant learning scenario, we found that generating errors could be helpful to memory even during the learning of novel material” (p. 666), Grimaldi and Karpicke (2012) wrote that “it is clear that attempting retrieval has the potential to improve learning” (p. 512), and Knight et al. (2012) proposed that “unsuccessful tests in which errors are retrieved can enhance retention” (p. 745). These findings are exciting inasmuch they contradict a long tradition of errorless approach to learning and instead suggest that making errors does not only harm, but can even benefit memory (Wong & Lim, 2022). These numerous recommendations for learning which involves unsuccessful retrieval attempts, including incorrectly guessing at the answer, encouraged us to further investigate this potentially useful study technique.

At the same time there are some inconsistencies in the existing literature, which suggest that guessing does not always enhance learning. Its effectiveness is determined by the nature of studied materials, and more specifically, by whether they are semantically related or not (e.g. Knight et al., 2012). Guessing also differentially benefits item and associative memory (Seabrooke et al., 2019) and is prone to temporal factors such as feedback delivery timing (Vaughn & Rawson, 2012). Minding these limitations of guessing, the current project aims to better understand its underlying mechanisms as they seem to depend on the learning paradigm and do not operate universally.

The main aim of the current project is to investigate whether guessing can be an effective way to acquire foreign vocabulary. We chose these study materials due to their clear educational relevance, especially minding the requirements of today's world in which travelling and working in multinational teams is a standard. What is more, there are already studies suggesting that guessing was shown to improve memory when studying novel foreign words and their translations (Potts & Shanks, 2014). At the same time, the benefit of guessing at such unrelated word pairs seems to be restricted to item memory and does not include the associations memory (Seabrooke et al., 2019). Item memory refers only to recalling which words were studied without associating them with each other, which is key in foreign vocabulary acquisition. In other words, learners will not benefit from remembering that they studied how to say given word in foreign language if they cannot associate this native word with foreign translation. Thus, for guessing to be really useful when acquiring new vocabulary, aside from enhancing item memory, it would also need to enhance the associations between foreign words and their native counterparts. The first main aim of the current project, explored in Studies 1 and 2, was to investigate if that can indeed be the case.

Study 1[1]

The current project concentrates on foreign vocabulary acquisition. Focusing on learning translations of foreign language words has several obvious advantages. First, the materials are ecologically valid, and improving foreign language acquisition is important from a societal point of view. Second, as mentioned, some researchers (Potts & Shanks, 2014) succeeded in demonstrating the beneficial effects of foreign vocabulary learning when participants tried to guess the meaning first. As mentioned, those positive results are so far constrained to recognition benefits representing of item memory, and the question we asked is whether guessing still may prove effective when association memory is of interest.

Considering foreign language acquisition – the association between a given foreign word and its translation is key. One may know a bunch of foreign words, but without understanding their meaning, this knowledge becomes far from useful. This reasoning was confirmed by a series of experiments performed by Seabrooke et al. (2019). They observed a guessing benefit using foreign words as stimuli on recognition, but on cued recall, which is a measure sensitive to memory for associations between cues and targets – guessing remained inferior to reading. Therefore, if unrelated pairs of words are used, guessing does strengthens targets; however, the association between them and the cues is not enhanced.

As mentioned, guessing is one strategy that has been shown to involve desirable difficulty: formulating a guess as to what the correct answer may be requires more effort than reading the answer outright and generally results in better

[1] The four experiments described as Study 1 have been published in an article: Butowska, E., Hanczakowski, M., & Zawadzka, K. (2022). You won't guess that: On the limited benefits of guessing when learning a foreign language. *Memory & Cognition*, *50*(5), 1033-1047.

learning. However, to date insights regarding the effectiveness of guessing have been mostly gleaned from single-trial learning procedures, in which participants are presented with the study material just once and do not have a chance to go back to it before the final memory test. Importantly, it is not known how these effects are modulated by introducing additional study opportunities, i.e., restudy. This omission is important, firstly, because very rarely learning sessions are limited to a singular encounter with material. Much more often learning occurs in multiply sessions simply because one such session is not enough to acquire new information. Also, as mentioned, temporal factors, such as e.g. whether feedback was immediate or delayed (Vaughn & Rawson, 2012) have already been shown to be important in modulating the effects of guessing, which further warrants research into guessing from a perspective of more than one learning phase.

Minding the above, we wanted to investigate whether giving our participants an opportunity to restudy material will affect guessing effectiveness when associating foreign words with their translations. As mentioned, guessing when used on unrelated words is beneficial only when item memory is measured, i.e. guessing helps to remember which words were studied but does not help to recall them as a specific translation of presented foreign word. However, if we can make this advantage reflected in the memory for the associations between foreign words and their meanings (as evidenced by a cued recall test) – that would greatly advance educational applications of guessing as a learning strategy, as foreign language acquisition relies on associations between words, tapped by cued recall. In other words, what we aim to teach our participants is using foreign vocabulary, rather than correctly resolving simple tests.

We hypothesized that guessing would benefit memory when people are presented with the to-be-remembered words twice. There are several lines of research supporting this hypothesis. The first one is related to numerous observations of beneficial effects of retrieval for new learning, a pattern generally referred to as test-potentiated learning (e.g., Izawa, 1971). Retrieval has been shown to improve learning across a variety of topics, including foreign vocabulary pairs (Arnold & McDermott, 2013a) of the sort used in guessing studies. The findings of Arnold and McDermott (2013b), who manipulated the presence of tests following the initial study phase as well as the presence of a subsequent restudy phase, are of special relevance here. Not only do their findings show that retrieval improves memory, but also that restudy results in greater memory benefits when preceded rather than not preceded by a test. Although guessing prior to the presentation of the study materials is different from tests given after initial study, an argument can be made that both situations involve retrieval attempts and hence may produce similar patterns of potentiated learning at restudy. According to Kornell et al. (2015), retrieval effort rather than retrieval success is what matters for test-potentiated learning, in which case unsuccessful guessing may indeed be similar to a test in maximizing the effectiveness of restudy.

What adds to the potential important role of restudying in guessing research is the work by Reder et al. (2016). The authors showed better memory for complex materials (Chinese characters) made up of previously known elements, compared to unknown parts of such characters. The work on source memory similarly shows that associating items with their sources is more effective when items themselves are made more familiar by being primed (Gagnepain et al., 2008; but see Kim et al., 2012). Such findings map onto a proposal by Popov and Reder (2020) who postulated the existence of encoding resources that can be depleted and then restored by the passage of time. According to this concept, when items are more familiar, fewer resources are required for their encoding and the remaining resources can be spent on encoding associations, supporting contextual memory. Adapting the encoding-resources hypothesis to the issue of learning novel materials through guessing starts with the observation that guessing improves memory for individual components of the to-be-learned pairs (Seabrooke et al., 2019). This improvement in memory for individual cues and targets may result in facilitated encoding of associations at restudy. In other words, translations of foreign language words, once strengthened by guessing, could be then easier to bind with their foreign counterparts, which would show as an advantage on the final cued-recall test.

Furthermore, research on successive relearning and learning to criterion (Vaughn et al., 2016) may provide some relevant insight too. In short, when people have an opportunity to study material several times, higher initial learning criterion becomes attenuated, which leads to relearning override effect. This means that, even when some material is initially more practiced (i.e. learnt to a higher criterion), such memory advantage will disappear after allowing learners to restudy the whole material so that what was previously better remembered will be remembered similarly well to the rest of the material, which is described as the relearning override effect. Referring this to the current design – even if people initially learn better in read condition, this advantage over guessing may diminish when people are given a chance to relearn.

Last but not least, the recursive reminding framework (e.g., Tullis et al., 2014) also underlines the role of restudying learning stimuli. Recursive reminding refers to recalling of an earlier instance of being presented with a study material which is somehow related to what is studied now. For instance, when studying a word list a word *Santa* can remind learners that they previously saw the word *Christmas* at study, hence the memory for the word *Christmas* will be embedded in the memory for the more recently seen word *Santa*. Thus, a higher order memory representation is created by incorporating memory for one event in that of another, more recent event (Jacoby & Wahlheim, 2013). Translating this to our study using foreign translations – seeing a given pair of words for the second time (during restudy) can be beneficial, as it reminds participants of the first encounter in which they attempted to guess the correct translation, which can help to create a richer memory representation comprising of the cue (foreign word), initial guess (incorrect translation), and the target (correct translation).

Therefore, in order to investigate if guessing can enhance associations between foreign words and their translations once participants are allowed a restudy opportunity, we adapted the standard procedure used in guessing studies involving learning foreign vocabulary which additionally incorporated a restudy phase. In Study 1 we asked participants to complete a learning session for Finnish-Polish word pairs in order to investigate how restudy affects memory performance in the guessing paradigm. Half of the pairs was presented in the guess condition requiring guessing the translation before being presented with corrective feedback in the form of the right answer, and the other half was presented intact in the read condition in which the foreign word and its translation were presented simultaneously outright. The fact that Finnish is largely unknown in Poland and that translations are difficult to deduce was why we decided to use this language. A restudy phase and a cued-recall test followed this initial study phase. Once more, after a single presentation, we predicted better test performance in the read condition, replicating previous findings

(Seabrooke et al., 2019, Experiment 5). However, we expected restudy to be more beneficial for items from the guess condition due to stronger memory representations of targets that should be easier to bind with their Finnish counterparts. Basically, we anticipated that restudy would either limit the benefits of reading, or even reverse it. Thus, ultimately, our objective was to determine whether initial guessing would be an equivalent or even a superior learning strategy to reading given the chance to restudy novel materials.

To sum, minding all the above-mentioned literature, we predicted an interaction as follows: we expected to observe superior performance (in cued recall) in the read condition over guessing when participants are granted only one encounter with stimuli, however, after adding a restudy phase we expected the pattern to reverse, ideally revealing the benefits of guessing or, at the very least, a lack of guessing costs.

Experiment 1.1

In the present experiment, we aimed to determine whether restudy changes the pattern of differences across guess and read conditions when learning foreign language vocabulary. We asked our participants to learn Finnish-Polish pairs via either reading or guessing with immediate feedback, and then we manipulated whether these pairs were presented for restudy via reading. We assessed memory both by the means of simple recognition, to confirm that guessing strengthens individual targets, and cued recall, to first confirm that guessing yields costs compared to reading for associative memory, and then establish whether this cost is ameliorated by restudy.

Method

Participants. Sixty students and graduates from Warszawa and Łódź (11 male; age range 18-57, mean: 26.4) with no previous knowledge of Finnish took part in the experiment in exchange for course credit or gift cards. We excluded two participants as they failed to provide any guesses during the learning phase. This resulted in a final sample of 58 participants. The study was approved by the Department of Psychology Ethics Committee at the SWPS University in Warszawa.

Materials and Design. Sixty-four pairs of Finnish-Polish nouns were used as study materials (see Appendix 1.1), and two additional pairs were used for practice. Finnish words were used as cues and their Polish translations were the targets. As shown in Figure 1.1, the study list was divided in two, with each half being assigned to one study-test block. Within each block, 16 words from the study list were assigned to the guess condition, and the remaining 16 to the read condition, with the assignment of words to conditions being counterbalanced across participants. In the read condition full word pairs were presented in the middle of the screen. In the guess condition the Finnish cue appeared first and participants were instructed to guess and type in its Polish meaning. After that, corrective feedback – the Polish translation – was presented.

After the presentation of the 32 pairs, half of them (eight from the read condition and eight from the guess condition) was presented for restudy. The pairs were always presented for restudy in full, and the assignment of words to the restudy versus no-restudy conditions was counterbalanced across participants. The restudy phase was followed by a cued-recall test for 16 of the already presented pairs, four from each condition. Participants were presented with Finnish words as cues and asked to type in their Polish translations. After the first cued-recall test, the second study-test block followed, which was identical to the first block bar the replacement of all materials. After the second cued-recall test, a final four-alternative forced-choice recognition test was administered for the remaining half of the pairs which was not tested yet (16 from each block, eight from each condition). The cue was again the Finnish word, which appeared with the correct translation (the target) and three novel lures presented in a random order. The lures were Polish words with the same number of letters as the target, and none of them was presented in any of the earlier phases of the experiment.

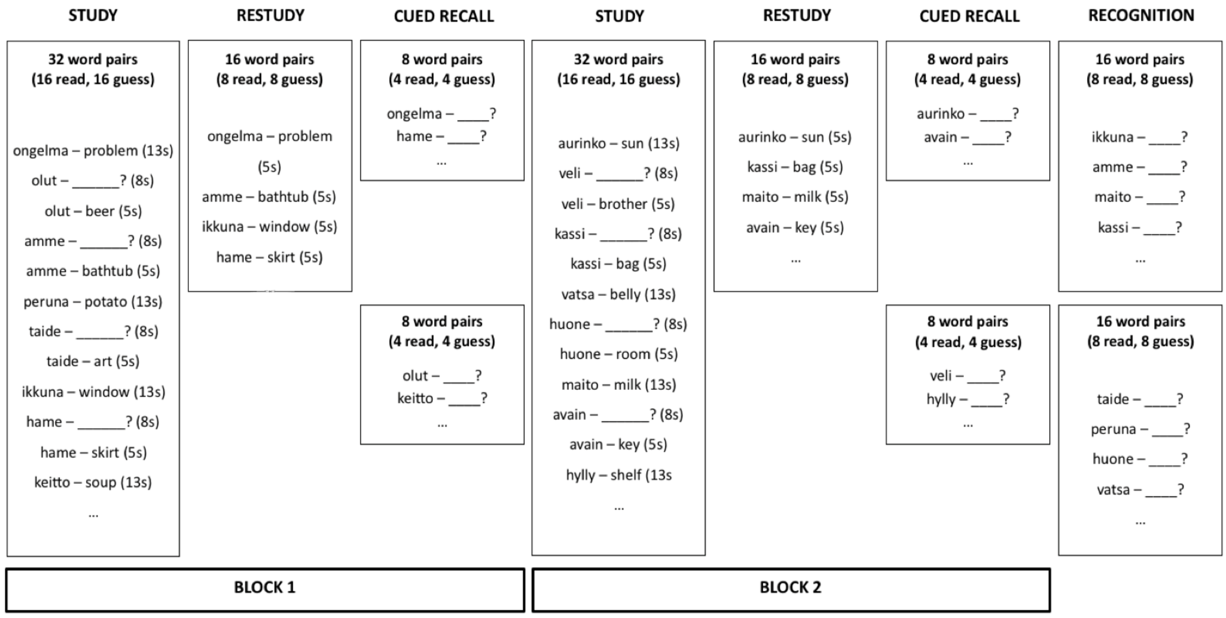

Figure 1.1 A schematic design of Experiment 1.1. For convenience, Finnish-English (rather than Finnish-Polish) translations are presented.

In sum, the study had a 2 (learning condition: guess, read) x 2 (restudy condition: restudy, no restudy) x 2 (test format: cued recall, recognition) within-subject design. The assignment of words to conditions, study-test blocks, and test types, as well as the order of presentation of all items at study and test were counterbalanced across participants. Of primary interest were the cued-recall results, while the simple-recognition test was included as a manipulation check to ensure that guessing in our paradigm indeed strengthens the targets.

Procedure. Participants were tested individually or in small groups. They were instructed that their task would be to learn Polish translations of Finnish words for a future test. They were then informed that they would encounter two types of study trials in the procedure. On some trials, both words would be presented for study for 13 seconds, while some other trials would require guessing the meaning of the presented Finnish word within eight seconds before being presented with the correct translation for five seconds. Participants were also informed that afterwards they would see some of the pairs presented for a second time for study. After these initial instructions, participants were given a short training phase consisting of studying and being tested on two pairs (one from each learning condition) to ensure a good understanding of the task. After the practice phase, participants completed the first study phase and then a restudy phase. In the restudy phase all pairs were presented for five seconds for participants to read. The restudy phase was followed by instructions for the first cued-recall test which underscored that participants had to retrieve the correct translations of the Finnish words and not their own guesses. The time for responding in this test was not limited. The cued-recall test was followed by the second study-test block. After the second cued-recall test, participants were given a final recognition test for pairs from both blocks that were not tested before.

Results and Discussion

During the learning phase, nine participants correctly guessed the meaning of a single Finnish word (sisko-sister). This resulted in an average of 0.5% correctly guessed items, which were removed from subsequent analyses. This is common practice in studies on guessing, because they focus on errorful learning and so are not interested in the learning process during which no errors are committed, and which involves different mechanisms. Thus, when participants successfully guess the correct response in the learning phase, the corresponding test items need to be excluded from the analysis. Overall, participants failed to type in their guesses on .06 of trials (SD = .10). We performed an additional control analysis excluding the trials on which our participants failed to provide their guess for Experiment 1.1, 1.2, and 1.3. However, no difference was observed compared to standard analyses including the trials with no typed guesses. For simplicity, here and in the subsequent experiments we report the analyses which include both typed guesses and questions left blank throughout the experiment, unless stated otherwise.

Recognition. Table 1.1 presents the mean level of accuracy on the multiple-choice simple-recognition test depending on the learning conditions. We performed a repeated-measures analysis of variance (ANOVA) with learning condition (guess, read) and restudy opportunity (restudy, no restudy) as factors. This revealed a significant main effect of learning condition, $F(1,57) = 7.87$, $p = .007$, $\eta_p^2 = .12$. Overall, guessing ($M = .94$, $SD = .09$) benefitted final recognition performance to a greater extent than reading ($M = .90$, $SD = .14$). Also, there was a significant main effect of restudy, $F(1,57) = 31.61$, $p < .001$, $\eta_p^2 = .36$. Restudying items benefitted final accuracy, $M = .95$, $SD = .09$, compared to $M = .89$, $SD = .14$, for non-restudied items. Finally, there was a significant interaction between the two factors, $F(1,57) = 6.01$, $p = .017$, $\eta_p^2 = .09$. This interaction arose because there was a significant difference between the two learning conditions when no restudy opportunity was available, $t(57) = 3.09$, $p = .003$, $d = 0.41$; however, this difference was no longer significant for word pairs that were restudied, $t(57) = 0.82$, $p = .417$, $d = 0.11$. Still, as the results after restudy were near ceiling, we refrain from interpreting this interaction as it might simply constitute a statistical artefact.

Table 1.1

Recognition Performance as a Function of Learning Condition and Restudy Opportunity in Experiment 1.1.

	Learning Condition	
Restudy Opportunity	Guess	Read
No Restudy	.92 (.01)	.86 (.02)
Restudy	.96 (.01)	.95 (.01)

Note. Standard errors of the mean are given in parentheses.

Together, the significant main effect of learning condition and, more specifically, the difference in recognition performance between the read and guess conditions with no restudy, demonstrate that item memory was improved after attempting to guess the target compared to merely reading the cue-target pair. This result is consistent with previous findings (Potts & Shanks, 2014; Seabrooke et al., 2019), and supports our assumption that guessing leads to strengthening of targets. This allowed us to investigate whether greater target strength translates into better encoding of associations at restudy, as measured by the cued-recall test.

Cued Recall. Table 1.2 presents the mean level of cued-recall performance depending on the learning conditions in Experiments 1.1-1.3 (All cued-recall analyses reported in Experiments 1.1-1.3 were performed on the combined data from block-1 and block-2 tests). A 2 (learning condition: read, guess) x 2 (restudy opportunity: restudy, no restudy) ANOVA revealed a significant main effect of learning condition, $F(1,57) = 20.54$, $p < .001$, $\eta_p^2 = .27$. Overall, reading benefitted performance to a greater extent than guessing ($M = .48$, $SD = .32$ in the read condition, compared to $M = .40$, $SD = .30$ in the guess condition). Also, there was a significant main effect of restudy, $F(1,57) = 264.23$, $p < .001$, $\eta_p^2 = .82$. Restudying the translations resulted in better overall final memory performance compared to not restudying (*M* = .63, *SD* = .25, and *M* = .25, *SD* = .25, respectively). Most importantly, and contrary to our predictions, the interaction of the two factors was not significant, $F(1,57) = 0.25$, $p = .621$, $\eta_p^2 = .004$.

Table 1.2

Cued-Recall Performance as a Function of Learning Condition and Restudy Opportunity in Experiments 1.1-1.3.

	Learning Condition	
Experiment and Restudy	Guess	Read
Experiment 1.1		
No restudy	.20 (.03)	.29 (.03)
Restudy	.59 (.03)	.67 (.03)
Experiment 1.2		
No restudy	.16 (.02)	.20 (.03)
Restudy	.31 (.03)	.39 (.03)
Experiment 1.3		
No restudy	.17 (.02)	.20 (.03)
Restudy	.33 (.03)	.37 (.03)

Note. Standard errors of the mean are given in parentheses.

Forward Testing Effect. The forward testing effect refers to a finding that inserting tests into study phase facilitates learning of new information, presented after such test (Chan et al., 2018; Yang et al., 2018). To verify whether the forward testing effect was present in our data, we conducted another ANOVA on cued-recall performance results with an additional block factor (apart from learning condition and restudy).

We performed this analysis for items tested with a cued-recall test only, as there was only one recognition test at the end of the experiment. We observed a significant main effect of block, $F(1,57) = 7.73$, $p = .007$, $\eta_p^2 = .12$. Overall, the accuracy in block 1 ($M = .41$, $SD = .20$) was lower than in block 2 ($M = .46$, $SD = .23$) and this difference was significant $t(57) = 2.78$, $p = .007$, $d = 0.37$. This is likely due to the forward testing effect for learning new materials.

Our first experiment did not demonstrate any positive effect of guessing on cued-recall performance, even after restudying the unfamiliar foreign vocabulary translations. As predicted, reading presented better recall performance without the opportunity for restudy. However, this benefit of reading over guessing remained even after restudying the foreign pairs. That was the case despite the fact that the recognition results confirmed previous findings of guessing advantage over reading. In short, our results replicate the finding that guessing did indeed improve target memory, which was crucial for our predictions. This suggests that our materials are well suited to evidence previously observed benefits of guessing, which refer to better memory for which words were studied. Contrary to our predictions, though, better target memory did not, in this instance, aid in the encoding of cue-to-target associations.

Experiment 1.2

The results of Experiment 1.1 failed to produce any benefits of guessing when foreign language translations were used as study materials, despite the addition of a restudy phase and the fact that targets in the guess condition were better remembered compared to those that were merely read. It is thus reasonable to evaluate the procedure used in this experiment in order to see why restudy may not have produced any memory improvements in the guess condition compared to the read condition.

One of the primary concerns regarding guessing as a learning technique is that it leads to mistakes, which can potentially interfere with memory for the targets. In fact, a whole tradition of errorless approach to learning strategies stands on the above premise, especially in clinical context (e.g. Jones & Eayrs, 1992). However, some recent research indicates that if errors are recalled during test, they might serve as a scaffold for accurate retrieval. Metcalfe and Huelser (2020), using the guessing paradigm, showed that performance in the guess condition was higher on trials on which participants could retrieve their original error compared to those on which the error could not be accessed. This finding is consistent with the work on interference by Wahlheim and Jacoby (2013), who – using the classic AB-AC interference paradigm, in which the second word associate is replaced with another related word during the second presentation – demonstrated that the usual effect of proactive interference of studying an AB pair for memory of an AC pair is reversed when participants detect a change in target words upon presentation of an AC pair, and – crucially – recollect the original AB pair in the final test. For instance, when learners notice that the word pair *knee-bone* was changed in the second presentation to the pair *knee-bend*, their memory for the target pair *knee-bend* will be 'enriched' by recollection of the similar pair presented in the first list and therefore will increase.

What is more, Pyc and Rawson (2010) showed that recalling one's own erroneous responses on the final test increased performance, which suggests that such incorrect responses produced during the learning phase can later serve to elicit targets provided that they are retrieved at test (but see Leggett & Burt, 2021). This suggests that a potentially interfering response (such as an incorrect guess) can also serve as a mediator, i.e., information linking cues with targets. What is more, in Pyc and Rawson’s study the learning phase consisting of tests and restudy sessions produced mediators which were both more likely to be retrieved at final test and also more likely to elicit targets compared to only restudying the material, without testing. This can be compared to our paradigm in which the guess condition consists of both a guessing attempt (which is an instance of an unsuccessful test) and is followed by restudy, and our read condition solely presents word pairs for study and then for restudy, without any testing. Following results by Pyc and Rawson (2010), our guess condition should produce better mediators than read condition. Thus, a potentially interfering response (in the form of an incorrect guess at a foreign translation) can serve as a powerful mediator, which – when produced at test – should help to retrieve the correct translation in response to its foreign counterpart.

It is vital to note that in Experiment 1.1 we had no way of knowing whether our participants were able to recollect their guesses at retrieval. If these were not retrieved, then they could not have served as self-generated episodic mediators for accessing the correct translations of the Finnish words. Therefore, in Experiment 1.2 we introduced a direct measure of *guess recollection* during test to assess the affordance of errors as potential episodic mediators. In addition to that, to make our participants think back to the initial learning phase, we also introduced a measure of *guess detection* during restudy by asking participants whether a given pair was studied in the read or guess condition. This incentive to recall guesses during restudy may enable their integration with targets that still need to be learnt, facilitating their usage as scaffolding for recall during a final test, and improving performance in the guess versus read condition.

Method

Participants. Sixty university students and graduates (14 male; age range 19-46, mean: 28.5) took part in the experiment in exchange for course credit or gift cards. After testing the first 39 participants, we were forced to cease face-to-face data collection due to the Covid-19 pandemic. The remaining 21 participants were thus tested online via a video link, with constant supervision from the experimenter. We continued with on-line testing for the rest of the experiments reported here.

Two participants had to be excluded due to procedure errors, another three because their final accuracy was close to zero, and another four because they failed to type in any guesses during the learning phase. This gave us a final sample of 51 participants. A relatively high percentage of exclusion was most likely resulting from a switch to online testing as well as a relatively difficult procedure.

Materials, Design and Procedure*.* The design of Experiment 1.2 is presented in Figure 1.2. All the materials and design were the same as in Experiment 1.1 except for the following changes. In the restudy phase, after the presentation of each pair, a question asking whether this pair was previously presented in the read or guess condition was presented in a forced-choice alternative format, and participants had to choose one of the options in order to advance to the next pair. This measure will be referred to as *guess detection*. Following Metcalfe and Huelser (2020; see also Yan et al., 2014), in the cued-recall tests we also asked the same question about the learning condition after each item. If the ‘guess’ option was chosen, a follow-up question appeared, asking to type in the initial guess. This was the direct measure of *guess recollection*. We also eliminated the recognition test from the procedure. This allowed us to have twice as many trials for analyses – that is, we had

32 pairs per each cued-recall test compared to 16 in Experiment 1.1. Finally, we replaced the often correctly guessed word pair ‘sisko-sister’ with a new one.

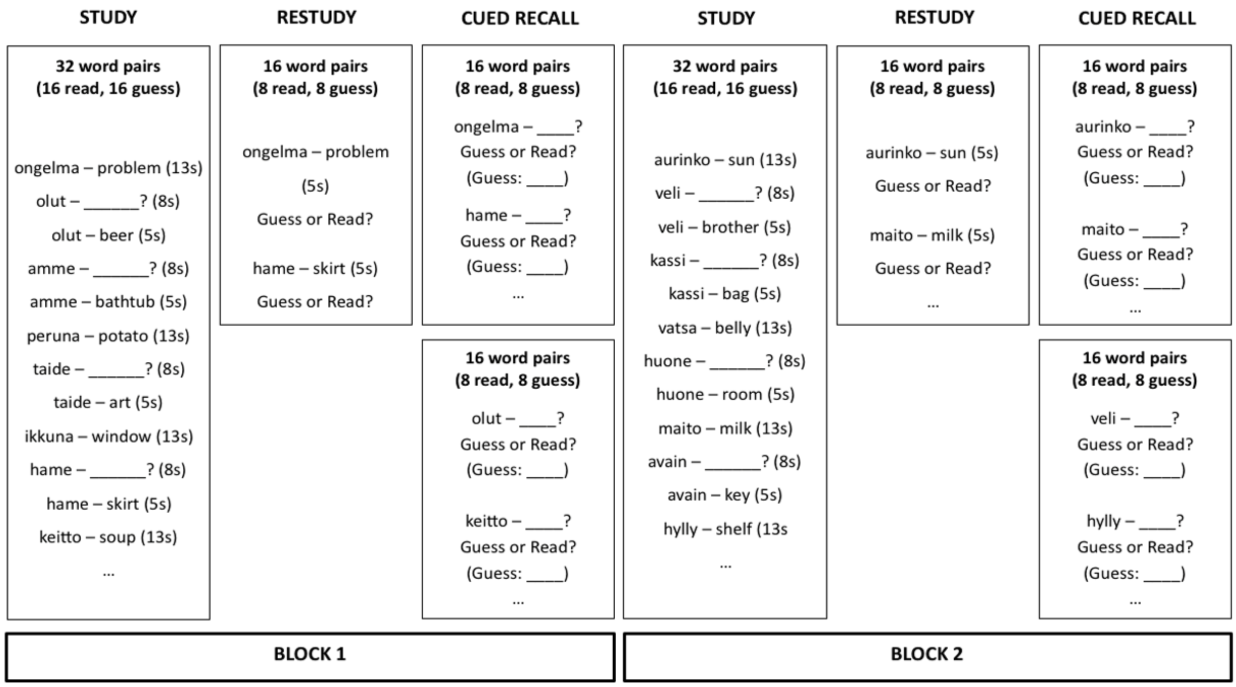

Figure 1.2. A schematic design of Experiments 1.2 and 1.3. In Experiment 1.2, guesses were unconstrained, while in Experiment 1.3 each guess had to begin with the same two letters as the cue. On cued-recall tests participants were also asked to type in their guesses for each pair they believed was studied in the guess condition.

Results and Discussion

During the learning phase, three participants correctly guessed the meaning of a word. This resulted in an average of 0.18 % correctly guessed items, which were removed from subsequent analyses. Overall, participants failed to type in their guesses on .13 (SD = .21) of trials during learning phase.

Cued Recall*.* Table 1.2 presents mean cued-recall performance across conditions. A repeated-measures 2 (learning condition: guess, read) x 2 (restudy opportunity: restudy, no restudy) ANOVA revealed a significant main effect of learning condition, $F(1,50) = 11.80$, $p = .001$, $\eta_p^2 = .19$. Overall, reading benefitted final recall accuracy compared to guessing ($M = .29$, $SD = .23$ and $M =.23$, $SD = .20$, respectively). Also, there was a significant main effect of restudy, $F(1,50) = 113.36$, $p < .001$, $\eta_p^2 = .69$. Restudying the translations resulted in an overall higher accuracy of .35 (*SD* = .22) compared to .18 (*SD* = .17) for not restudied items. The interaction between these factors was not significant, $F(1,50) = 1.77$, $p = .190$, $\eta_p^2 = .03$. These results replicate those of Experiment 1.1.

Notably, the benefit of restudy was attenuated in the present experiment. In Experiment 1.1, restudying increased final recall performance on average by .39. In Experiment 1.2, this average dropped to .17. This was most likely the result of an additional task we asked our participants to perform guess detection during restudy phase, which might have increased task load and resulted in a reduced benefit of restudy.

Guess Detection at Restudy. We divided restudied items in the guess condition depending on whether participants correctly classified them as guesses in the restudy phase, which we refer to as guess detection. Note that this analysis excludes all items that were assigned to the no-restudy condition. The majority of items ($M = .72$, $SD = .19$) were correctly classified at restudy as belonging to the guess condition. Cued-recall performance for items with and without guess detection at restudy can be seen in Figure 1.3. There was a significant difference in cued-recall performance between items correctly identified as being studied in the guess condition compared to those which were incorrectly labelled at restudy as read, $t(45) = 2.89$, $p = .006$, $d = 0.43$, with an advantage for items with guess detection.

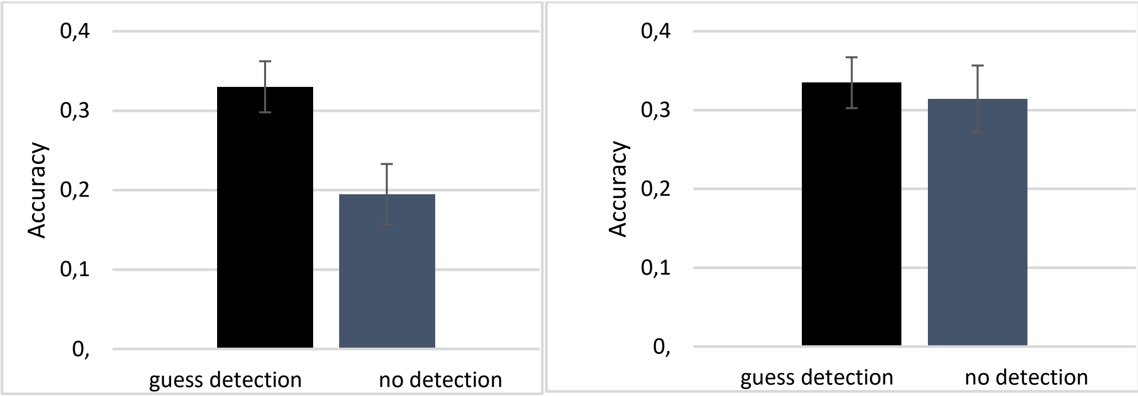

Figure 1.3. Cued-recall performance in Experiments 1.2 (left panel) and 1.3 (right panel) conditionalized on correct guess detection at restudy.

Guess Recollection at Test. The design of Experiment 1.2 allowed us to compare final cued-recall performance for items with and without successful guess recollection at test. In the following analyses we only included items with guesses typed in during the initial study phase, as for those items we could determine whether the guess retrieved at test matched that made at study. This is reflected in varying degrees of freedom across the analyses. Our participants remembered their guesses on .25 (*SD* = .22) of trials without a restudy opportunity and on .32 (*SD* = .22) of trials with restudy. The difference between these two rates was significant, $t(50) = 4.05$, $p < .001$, $d = 0.57$, suggesting that at least some guesses were retrieved and strengthened in the restudy phase, providing an opportunity for integration with the to-be-remembered items represented at restudy.

Table 1.3 presents cued-recall performance depending on restudy opportunity and guess recollection. A repeated-measures 2 (restudy opportunity: restudy, no restudy) x 2 (guess recollection: present, absent) ANOVA revealed a significant main effect of restudy, $F(1,35) = 19.43$, $p < .001$, $\eta_p^2 = 0.36$. Overall, restudying resulted in higher final accuracy (*M* = .37, *SD* = 0.32) compared to performance for translations that were not restudied (*M* = .19, *SD* = .22). Also, there was a significant main effect of guess recollection, $F(1,35) = 8.62$, $p = .006$, $\eta_p^2 = .20$. Pairs with guesses recollected at test were remembered better, *M* = .37, *SD* = .34, than pairs without correct guess recollection, *M* = .21, *SD* = .21. There was no significant interaction between these factors, *F*(1,35) = 0.28, *p* = .598, η_p^2 = .01. We also compared items with correct guess recollection at test with items from the read condition. There was no difference in cued-recall performance between the read condition and those items in the guess condition for which guess recollection was successful. This applies both to restudied, *t*(44) = 1.08, *p* = .286, *d* = 0.16, and non-restudied items, *t*(36) = 1.08, *p* = .287, *d* = 0.18. Additionally, we found a significant correlation between mean guess recollection rates and mean accuracy in guess condition across participants, *r*(49) = .425, *p* = .002.

Table 1.3

Cued-Recall Performance as a Function of Guess Recollection and Restudy Opportunity in Experiments 1.2 and 1.3.

	Guess Recollection	
Experiment and Restudy	Yes	No
Experiment 1.2		
No restudy	.26 (.04)	.15 (.02)
Restudy	.47 (.05)	.28 (.03)
Experiment 1.3		
No restudy	.29 (.04)	.16 (.02)
Restudy	.41 (.04)	.30 (.03)

Note. Standard errors of the mean are given in parentheses.

Guess Detection at Test. We compared guess detection rates for restudied and not restudied items. The rates were 51 % and 57% for not restudied and restudied items, respectively. There was a significant difference between the conditions, $t(49) = 2.24$, $p = .029$, $d = 0.32$.

Table 1.4 presents cued-recall performance depending on restudy opportunity and guess detection. A repeated-measures ANOVA revealed a main effect of restudy, $F(1,50) = 30.97$, $p < .001$, $\eta_p^2 = .38$. Restudying pairs resulted in the overall accuracy of .29 (*SD* = .26) compared to the accuracy rate of .15 (*SD* = 0.19) for not restudied pairs. Also, there was a significant main effect of guess detection, $F(1,50) = 24.74$, $p < .001$, $\eta_p^2 = .33$. Pairs correctly identified at test as initially guessed were recalled at the rate of .29 (*SD* = .25) compared to .15 (*SD* = 0.20) for incorrectly identified items. There was no significant interaction between these factors, $F(1,50) = 3.23$, $p = .078$, $\eta_p^2 = .06$.

Table 1.4

Cued-Recall Performance as a Function of Guess Detection and Restudy Opportunity in Experiments 1.2 and 1.3.

	Guess Detection	
Experiment and Restudy	Yes	No
Experiment 1.2		
No restudy	.20 (.02)	.10 (.02)
Restudy	.37 (.04)	.20 (.03)
Experiment 1.3		
No restudy	.22 (.03)	.13 (.02)
Restudy	.35 (.03)	.29 (.04)

Note. Standard errors of the mean are given in parentheses.

Forward Testing Effect. As earlier, we conducted another ANOVA with additional block factor (apart from learning condition and restudy). We observed a significant main effect of block, $F(1,50) = 8.41$, $p = .006$, $\eta_p^2 = .14$. Overall, the accuracy in block 1 ($M = .23$, $SD = .17$) was lower than in block 2 ($M = .30$, $SD = .21$) and this difference was significant $t(50) = 2.90$, $p = .006$, $d = 0.41$.

Experiment 1.2 replicated the findings of Experiment 1.1, again showing costs of guessing – relative to reading – in cued recall after studying foreign word translations. More importantly, this cost was not mitigated by restudy: after restudying the word pairs, reading still remained superior to guessing. These results, which emerged despite the inclusion of the guess detection question at the time of restudy, again undermine the usefulness of guessing as a learning strategy for novel materials such as foreign language vocabulary.

We found a similar pattern of outcomes to Metcalfe and Huelser (2020) when focusing exclusively on the guess condition and contrasting items with and without guess recollection: target recollection was greater for items for which guesses were remembered rather than forgotten. Also, restudy by itself seemed to increase the chances that one's guesses would be recollected at test, suggesting that participants at least sometimes were reminded of their guesses during restudy. This did not, however, result in an overall improvement in performance, since even pairs for which guesses were recollected were not remembered better than pairs from the read condition. This pattern presents a potential scenario in which differences across pairs for which guesses were versus were not recollected at test reflect a type of an item-selection artifact, by which items that are more likely to be correctly remembered are the same items for which guesses are more likely to be remembered

Before we can conclude that guessing cannot outperform reading foreign words and their translations – whether one can gain access at test to one's initial guess or not – it is worth noting that the overall rate of guess recollection in

Experiment 1.2 was only around 28%. This offers a different way to interpret the findings in which guess recollection does indeed enhance performance above that seen in the read condition, but with the materials that we used it simply occurs too rarely for this difference to be reliably observed. According to this account, the problem with using the guessing strategy for learning novel materials is that participants too rarely remember their guesses, even if they engage in restudy when they could be reminded of their initial guesses. Our main aim in designing Experiment 1.3 was thus to make it easier for participants to recall their initial guesses and thus to increase the overall guess recollection rate. To this end, participants in Experiment 1.3 had their guesses constrained: they had to begin with the same two letters as the Finnish cue.

Experiment 1.3

In this experiment, we sought to increase the prevalence of guess recollection at test. Without any prior knowledge of the to-be-learned Finnish words, we reasoned that participants' guesses were less likely to be sensibly related to the cues and more likely to rely on some random contextual features (e.g., the preceding study item). Such guesses would be very difficult to remember when contextual features change between study and restudy and then between restudy and test. Without access to semantic features of the Finnish words, the easiest strategy for generating guesses that could be retrieved later would be to rely on phonetic information embedded in a cue. Thus, in Experiment 1.3 we explicitly instructed participants to generate their guesses based on the first two letters of the cue word. In this case, the generated words would not so much be guesses at the meaning of the cue word – participants could quickly discern that Polish translations do not start with the same two letters as their Finnish counterparts – but still could serve as potential mediators, similar to those used in the keyword technique for foreign vocabulary acquisition (Lawson & Hogben, 1998). Once more, we were curious to see if recalling such self-generated episodic mediators at test would enhance performance compared to performance in the read condition.

Method

Participants*.* Sixty SWPS University students (11 male; age range 20-52, mean: 29.5) took part in the experiment in exchange for course credit. They were all tested online, in the same way as in Experiment 1.2. We excluded one participant due to technical difficulties which terminated the experiment, another two participants as they were found to be making notes during the learning phase, another three who failed to provide guesses and maintained floor-level accuracy, and one person who did not understand the instructions. This gave us a final sample of 53 participants.

Materials and Design*.* The design was the same as in Experiment 1.2, and can be seen in Figure 1.2. Because in this experiment all guesses had to start with the same two letters as the Finnish cue, we had to replace some our study materials so that all words would satisfy the following conditions: 1) each Finnish word had to start with two letters that could also serve as the initial two letters of a Polish word; 2) each combination of two initial letters had to be unique; and 3) no Polish translation could start with the same two letters as any of the Finnish associates (see Appendix 1.2).

Procedure. The procedure was modelled on that from Experiment 1.2, with the following exception. During the learning phase, we instructed our participants to provide guesses starting with the same two letters as the Finnish word. For instance, for the cue PISTE (Finnish for *dot*) its first two letters (PI____) were presented next to it, and the guess had to start with them. The sole aim of this manipulation was to increase the guess recollection rate – we assumed that the two initial letters should constitute a good cue to remind oneself what the initial guess was.

Results and Discussion

During the learning phase, no correct guesses occurred, simply because they were constrained by the first two letters of the cue which were never the same as for the target. Therefore, no items were excluded from the analysis on the basis of having been guessed correctly. Overall, participants failed to type in their guesses on .24 (SD = .23) of trials during learning phase.

Cued Recall*.* Table 1.2 presents mean cued-recall accuracy levels across conditions. A 2 (learning condition: read, guess) x 2 (restudy opportunity: restudy, no restudy) repeated-measures ANOVA revealed a significant main effect of restudy, $F(1,52) = 84.23$, $p < .001$, $\eta_p^2 = .62$. Overall, restudying items resulted in final cued-recall performance of .35 ($SD = .22$) compared to items deprived of the restudy opportunity which were retrieved at a rate of .18 ($SD = .17$). The main effect of learning condition was not significant, $F(1,52) = 3.47$, $p = .068$, $\eta_p^2 = .06$, even though there was again a trend toward better performance in the read than in the guess condition, $M = .28$, $SD = .23$, and $M = .25$, $SD = .19$, respectively. The interaction of the two factors was not significant, $F(1,52) = 0.01$, $p = .917$, $\eta_p^2 < .001$.

Guess Detection at Restudy*.* Out of all items from the guess condition, 67% were correctly classified as such at restudy. Cued-recall performance for items with and without guess detection can be seen in Figure 1.3. There was no significant difference in the final cued-recall performance between items correctly identified at restudy as being from the guess condition compared to those which were incorrectly labelled as read: $t(49) = 0.22$, $p = .827$, $d = 0.03$. This stands in contrast to our findings from Experiment 1.2.

Guess Recollection at Test. As in Experiment 1.2, we compared final cued-recall performance for items with and without successful guess recollection. In this analysis we only included items with guesses typed in during the learning phase. The guess recollection rates were .29 (*SD* = .18) for non-restudied items and .40 (*SD* = .25) for restudied items. This difference was significant, $t(50) = 4.19$, $p < .001$, $d = 0.59$, again suggesting that guesses were spontaneously retrieved during restudy.

Table 1.3 presents cued-recall performance depending on restudy opportunity and guess recollection. A 2 (restudy opportunity: restudy, no restudy) x 2 (guess recollection: yes, no) repeated-measures ANOVA revealed a significant main effect of restudy, $F(1,43) = 18.47$, $p < .001$, $\eta_p^2 = .30$. Overall, targets from restudied pairs were retrieved more often, *M* =.35, *SD* = .27, than those from pairs that were not restudied, *M* = .22, *SD* = .25. Also, there was a significant main effect of guess recollection, $F(1,43) = 8.41$, $p = .006$, $\eta_p^2 = .16$. Targets from pairs for which guesses were recollected were retrieved more often at test, *M* = .35, *SD* = .31, compared to *M* =.23, *SD* = .21 for items with unrecalled guesses. The interaction between these factors was not significant, $F(1,43) = 0.28$, $p = .598$, $\eta_p^2 = .01$. Finally, we compared items with correct guess recollection at test with items from the read condition. There was no significant difference in the cued-recall measure between read items and guessed items with correctly reported guesses. This applied both to restudied items, $t(46) = 0.58$, $p = .567$, $d = .08$, as well as those presented once, $t(44) = 1.87$, $p = .069$, $d = .28$. This shows that guessing does not outperform reading as a learning strategy for foreign word pairs even under conditions in which it was most effective – with successful guess recollection at test.

Contrary to Experiment 1.2, we found no significant correlation between mean guess recollection rates and mean accuracy in guess condition across participants, $r(51) = .025$, $p = .863$.

Guess Detection at Test. We compared guess detection rates for restudied and not restudied items. The rates were 43 % and 56% for not restudied and restudied items, respectively. There was a significant difference between the conditions, $t(51) = 4.97$, $p < .001$, $d = 0.69$.

Table 1.4 presents cued-recall performance depending on restudy opportunity and guess detection. As in Experiment 2, a repeated-measures ANOVA revealed a main effect of restudy, $F(1,51) = 35.51$, $p < .001$, $\eta_p^2 = .41$. Overall, restudy benefitted final cued-recall performance compared to a single study opportunity ($M = .32$, $SD = .26$, and $M = .17$, $SD = .18$, respectively). Also, there was a main effect of guess detection, $F(1,50) = 9.38$, $p = .004$, $\eta_p^2 = .16$. Pairs which were correctly identified at the final tests as being guessed during the learning phase were recalled at the rate of .29 ($SD = .24$) compared to .21 ($SD = .23$) for pairs which were incorrectly identified. There was no significant interaction, $F(1,50) = 0.47$, $p = .496$, $\eta_p^2 = .01$.

Forward Testing Effect. As earlier, we conducted another ANOVA with additional block factor (apart from learning condition and restudy). We observed a significant main effect of block, $F(1,52) = 10.13$, $p = .002$, $\eta_p^2 = .16$. Overall, the accuracy in block 1 ($M = .23$, $SD = .17$) was lower than in block 2 ($M = .29$, $SD = .19$) and this difference was significant $t(52) = 3.18$, $p = .002$, $d = 0.44$.

Experiment 1.3 replicated the previous two experiments in failing to show a benefit of guessing on cued recall after restudying foreign language translations. In this experiment, the main effect of learning condition was not significant but the numerical trend was consistent with the overall benefit of reading found in previous experiments. The bottom line is that guessing was once again not superior to reading, even after restudying the material and – consistent with Experiment 1.2 – even after correctly recollecting the guesses at test.

One point to note about the present results is that we did not manage to substantially increase the overall guess recall rate by providing first two letters of a word. In Experiment 1.2, in which participants tried to guess the meanings of the Finnish words, they remembered .32 of their guesses after restudy, while the same proportion in the present study was .40, which, albeit somewhat larger, was still not even half of the guesses initially generated during study. Therefore, there is still a chance that guessing might be a useful learning approach if it were implemented in a way that participants would recall the majority of their guesses at the test. The problem with this argument is, however, that remembering associates of unfamiliar foreign words is the very same difficulty that guessing is supposed to ameliorate. If participants find it difficult to associate translations with their respective foreign words, it is perhaps unsurprising that also their guesses are difficult to associate and thus not likely to be later recollected. Thus, it appears that the guessing strategy is suboptimal for acquiring foreign words – a problem which restudy does not serve to remedy – because guesses are not likely to serve as good episodic mediators, possibly because they are rarely recollected at test in the first place.

Experiment 1.4

All experiments reported so far show a clear pattern of results. Adding a restudy phase clearly improved cued-recall performance, but this did not interact with the manipulation of learning condition: performance was improved to a similar extent regardless of whether the pairs were previously read or required guessing. The question remains, however, of why exactly learning conditions and restudy did not interact in our study despite previous studies showing that better item memory for words constituting a pair should facilitate creation of associations between those items (Reder et al., 2013, 2016; Vaughn et al., 2016). It might be because of the materials that we used, with unfamiliar and thus difficult cues being too challenging to associate with the strengthened targets. Thus, in Experiment 1.4 we changed the materials to weakly related pairs of words – stimuli which consistently show the guessing benefit in cued recall (e.g., Bridger & Mecklinger, 2014; Hays et al., 2013; Kornell et al., 2009). If we again fail to see any interaction, we should be able to conclude that adding a restudy phase to the guessing paradigm does not alter the results, regardless of the materials used.

Method

Participants*.* Sixty-one students of the SWPS University (13 male**;** age range 19-47, mean: 27.6) took part in the experiment in exchange for course credit. We planned to test 60 participants, as in previous experiments, but we tested all people who signed up for the study.

Materials and Design*.* Experiment 1.4 was based on Experiment 1.1, with the following differences. We changed the materials from Finnish-Polish translations to weakly related pairs of words (see Appendix 1.3). As there are no association norms in Polish, we chose pairs of words with an average forward association strength of .05 from association norms in English (Nelson et al., 2004) and translated them into Polish. As in the previous experiments, we used 64 word pairs for the experiment proper (with two additional ones that were used in a training phase). In contrast to Experiment 1.1, but in line with Experiments 1.2 and 1.3, a cued-recall test was administered for all pairs. Also, there was a single learning and testing phase instead of two study-test blocks, simply because it was easier to remember related word pairs in one's native language compared to unknown Finnish-Polish translations. After pilot testing, we observed that performance after restudy was at ceiling. For this reason, we implemented a 20-minute delay before the cued-recall test to lower the average performance. During this time, participants completed an unrelated experiment.

Procedure. Participants were presented with a single list of 64 weakly related word pairs for study. Each pair was presented for 13 seconds in the read condition. In the guess condition each cue was first presented alone for eight seconds, during which time participants had to guess what the target might be; after that time, they were presented with the full pair for five seconds. For half of the pairs from each condition, a restudy phase followed during which full pairs were presented for five seconds. After the restudy phase, participants completed an unrelated experiment which took approximately 20 minutes. Finally, all cues were presented one by one and participants had to type in their corresponding targets or skip to the next pair if no target was retrieved.

Results and Discussion

In the learning phase, 36 paired associates were guessed correctly, which constituted 1.8% of all trials. These trials were removed from subsequent analyses. Overall, participants failed to type in their guesses on .09 (SD = .18) of trials during learning phase.

Cued Recall*.* Table 1.5 presents mean cued-recall performance across conditions. A 2 (learning condition: read, guess) x 2 (restudy opportunity: restudy, no restudy) repeated-measures ANOVA revealed a significant main effect of learning condition, $F(1,60) = 35.09$, $p < .001$, $\eta_p^2 = .37$. Overall, guessing benefitted final recall accuracy compared to reading, $M = .69$, $SD = .23$, and $M = .58$, $SD = .28$, respectively, replicating previous findings using related word pairs. Also, there was a significant main effect of restudy, $F(1,60) = 70.04$, $p < .001$, $\eta_p^2 = .54$. Restudying pairs resulted in higher accuracy – $M = .72$, $SD = .25$ – compared to $M = .55$, $SD = .25$, for not restudied items. The interaction between the two conditions was not significant, $F(1,60) = 1.83$, $p = .182$, $\eta_p^2 = .03$.

Table 1.5

Cued-Recall Performance as a Function of Learning Condition and Restudy Opportunity in Experiment 1.4.

	Learning Condition	
Restudy Opportunity	Guess	Read
No Restudy	.62 (.03)	.48 (.03)
Restudy	.76 (.03)	.68 (.03)

Note. Standard errors of the mean are given in parentheses.

In accordance with previous research, guessing outperformed reading when weakly related pairs of words were used as stimuli. Still, the overall pattern of results from our first three experiments was replicated in Experiment 1.4 as well, with significant main effects of both restudy and learning condition but no interaction between the two. This shows that no matter whether reading or guessing leads to better memory performance, the benefit persists even after an additional learning session.

General Discussion of Study 1

Study 1 aimed to investigate if guessing can join the already-established umbrella of challenging yet effective learning strategies referred to as desirable difficulties. More specifically, we tested if it can be used to effectively study foreign vocabulary. While being aware that so far guessing has been shown to be ineffective in learning associations between foreign words and their translations (Seabrooke et al., 2019), we introduced several theory-based modifications to the learning paradigm and hypothesized that they would allow for demonstrating a guessing advantage over reading. Nevertheless, we consistently failed to observe the guessing benefit and instead showed a reading advantage in learning associations between foreign translations. That held regardless of whether the translations were studied once or twice, directly contradicting our prediction that restudy should facilitate encoding of foreign word-to-translation associations. Experiment 1.1 demonstrated that despite the fact that guessing did strengthen the targets, as evidenced by better recognition performance, reading was a more effective strategy for learning associations as reflected in cued-recall results. This chimes with the findings of Seabrooke et al. (2019), who were the first to demonstrate that contrary to what Potts and Shanks (2014) postulated, guessing might not be the best strategy for foreign vocabulary acquisition. Importantly, this pattern held regardless of whether translations were studied once or restudied. Experiments 1.2 and 1.3 confirmed this finding and extended it by showing that even when participants remembered their guesses – and so could in principle use them as episodic mediators for target retrieval – guessing still was unable to outperform reading as a learning strategy for foreign language vocabulary.

To support this conclusion even more, we performed a Bayesian repeated-measures ANOVA in JASP (JASP Team, 2023) on the combined data from Experiments 1.1-1.3. This analysis (total N = 162) showed that the evidence for the main effects on their own was extreme: $BF_{(inclusion)}$ = 157,153.26 for the main effect of learning condition, and $BF_{(inclusion)}$ = 8.601e+64 for the main effect of restudy. In contrast, there was moderate evidence against the interaction of learning condition and restudy, $BF_{(inclusion)} = 0.126$. Therefore we can safely conclude that while it matters what learning strategy people use and whether they restudy the materials, the two factors are independent of one another. This stands in stark contrast to our prediction that strengthened targets should be easier to associate with their foreign equivalents when a restudy opportunity is available.

To evaluate the generalizability of the results from Experiments 1.1-1.3 regarding the ineffectiveness of restudy in ameliorating performance differences across the read and guess conditions, we conducted Experiment 1.4 with weakly related word pairs as study materials. With these materials, an advantage of guessing over reading in measures of both associative (e.g., Grimaldi & Karpicke, 2012; Huelser & Metcalfe, 2012; Kornell et al., 2009) and target memory (Zawadzka & Hanczakowski, 2019) is robustly obtained. In line with the many previous observations, Experiment 1.4 replicated the guessing benefit in cued-recall performance. However, we again failed to observe an interaction between the learning condition and restudy. Therefore, in a series of four experiments we confirmed that both costs (Experiments 1.1-1.3) and benefits (Experiment 1.4) of guessing are not modified by adding a restudy opportunity.

At the same time we acknowledge some limitations to the paradigm introduced in Study 1 of the current thesis. We realize that guessing at the identity of an unfamiliar novel word whose relationship with its translation is opaque to participants is bound to be random. It clearly differs from the guessing processes in different paradigms like, e.g., guessing at an answer to a trivia question or a related pair associate – which both have a clear semantic relationship with their target and enable a more informed guessing attempt. What is more, in more applied settings, when studying a foreign language, people usually use some contextual cues regarding meaning of the yet unknown words. Thus, in Study 2 we investigated whether the addition of contextual cues would improve the chances of finding the benefits of guessing for learning foreign vocabulary.

Study 2[2]

Having in mind that guessing at bare novel translations both diverges from more real-life instances of foreign language learning and also can become quite random, here we decided to investigate whether embedding foreign words within a contextual sentence affording participants to formulate more informed guess at its meaning will present guessing as a more effective strategy than reading. For instance, it is much easier to guess at the meaning of the Finnish word *lattia* when it is presented within a sentence such as e.g. 'A dog was lying on the *lattia*' compared to guessing at the word alone: *lattia* - ?. As already said, this approach may also be more suited to study foreign language acquisition compared to the one used in Study 1 of the present thesis, as foreign words are rarely studied in isolation. Apart from this applied prerequisite for the change in study materials, there is also a solid theoretical base about learning from the so-called prediction error – the degree of discrepancy between the correct answer and the predicted one – which merits the idea that more informed and confident, yet still incorrect guessing at foreign translations, should benefit learning their associations with native meanings.

The distinction between informed and random guessing outlined above can be reformulated as guessing with a varied magnitude of the resulting prediction error when the correct answer is revealed. The magnitude of the prediction error depends on the specificity of the prediction, with specific predictions – when negated – leading to a large prediction error and largely unspecified predictions leading to little prediction error (Henson & Gagnepain, 2010). For example, a recent football result of

[2] The four experiments described as Study 2 are currently under review: Butowska, E., Hanczakowski, M., & Zawadzka, K. (2023). *What role for prediction error in errorful learning?*

a match between Moldova vs Poland (3:2) led many Polish fans to a substantial prediction error, because they were certain that Poland would win with much lower ranked Moldova. However, trying to guess the result of El Clásico, played between FC Barcelona and Real Madrid does not usually involve such specific prediction regarding the final outcome as both teams play very well, which makes it hard to predict which one will win and so does not lead to a substantial prediction error.

Thus, when guessing in the errorful learning paradigm is informed by the nature of semantically related materials, participants can formulate specific guesses as to what the target could be, which means that when the correct target is revealed – which always differs from their guess – a large prediction error ensues. When guessing is not constrained, as in the case of materials lacking a transparent semantic structure, there is less room for prediction error when the correct target is revealed. This is because when studying such ambivalent materials, learners cannot form specific, confident predictions regarding the correct response. Therefore, even when learners predict some incorrect response, the corrective feedback will not strongly contradict such (barely existing) predictions, simply because they were rather uninformed and formulated without much faith and so cannot result in a substantial prediction error. The question addressed in Study 2 of the current thesis is whether this difference in terms of prediction error is what underlies the benefits of guessing for associative memory – regarding the link between foreign words and their meanings.

That this could be the case is suggested by a rapidly growing literature demonstrating the importance of prediction error in driving learning in general. Studies on this topic go back to the foundational theory of learning developed by Rescorla and Wagner (1972), where the effectiveness of classical conditioning was assumed to be a function of the magnitude of the prediction error. More recently, various studies on declarative memory have demonstrated that prediction error augments memory (Fazio & Marsh, 2009; Greve et al., 2017; Haeuser & Kray, 2021; Rommers & Federmeier, 2018; see Quent et al., 2021, for a review). For example, Greve et al. (2017) manipulated prediction error by training an association between a scene and either a single face or multiple faces, before presenting the same scene with a new face for learning. Multiple pairings during training should lead to less specific predictions as to the associated face presented for subsequent study than single pairings. Indeed, Greve et al. demonstrated that less specific predictions lead to poorer learning than more specific predictions that were negated when a new face was presented for study. Interestingly, in this case learning due to prediction error has been linked to improved associative memory – the pairing of scenes with new faces – and also other research suggests that prediction error primarily affects memory for associations (Kafkas & Montaldi, 2018). This seems consistent with the observation that the locus of the benefits of guessing for semantically related materials lies precisely in improved memory for cue-to-target associations (Knight et al., 2012; Kornell et al., 2009), which contrasts with the guessing benefits restricted to the item memory observed for semantically unrelated materials (Potts & Shanks, 2014, Seabrooke et al., 2019), where the role of prediction error should be minimal. This ability of prediction error to enhance associative memory is precisely why we assumed that such mechanism can allow guessing to be more effective than reading when learning foreign vocabulary, in which forming associations between translations is key.

In order to investigate the role of prediction error in learning by guessing, we needed to manipulate its magnitude in both the guess and read conditions and see if it interacts with these learning conditions, potentially favoring performance when using the guessing strategy. This was informed by research suggesting that prediction error specifically drives learning when participants are explicitly asked to guess at the target. In a recent study, Brod et al. (2022) showed enhanced memory for unexpected answers in a numerical general knowledge task, but only when participants were asked to predict what the answer might be before the correct answers were revealed, and not when they were asked to state what their predictions would have been only after seeing the correct answers. Similarly Potts et al. (2019) reported that producing possible translations of a foreign word after already seeing the correct answer produced similar memory performance to reading the full translations without any guessing attempts, whereas explicitly guessing at the target identity resulted in superior performance. These results suggest that guessing can potentially facilitate learning via prediction error, as the performance benefitted only when the erroneous responses were given before and not after feedback delivery.

How do the existing studies on learning by guessing fit into this proposed new perspective? As already argued, the logic of prediction error seems well-suited to explaining why some materials are capable of revealing associative benefits of guessing – those that allow for formulating relatively specific guesses as to the identity of the target, resulting in informed guessing – while other materials are not, as they result in random guessing. Regarding other relevant findings, the prediction error account can be addressed in two different ways. The first one is to keep the predictiveness of the cue constant and vary the extent to which the revealed target contradicts the initial assumptions. There are a couple of studies that implemented this approach, although they were not designed specifically to address the discussed potential account of guessing. At first blush, their results do not seem to align with the prediction error perspective. First, Seabrooke et al. (2022) looked at learning by guessing of foreign language translations in a design where participants needed to guess not only the meaning of the foreign word but also the category (animals or clothing) to which this word belonged. They found better recognition of targets when the correct category was identified – a situation in which guesses were closer in meaning to targets. While this is opposite to what the prediction error account would suggest, it is important to note that these results concerned target memory only. For associative memory – measured via cued recall – no benefits of guessing over reading were observed. This null result is thus not inconsistent with the prediction error account simply because with foreign vocabulary translations one would not expect specific guesses to be formulated in the first place and thus, with less room for prediction error, no associative benefits of committing errors would be expected. In other words, the study by Seabrooke et al. speaks to a different mechanism of guessing, one that is responsible for enhanced target memory, observed for all types of materials, and not the one that is responsible specifically for associative benefits when specific guesses could be formulated.

Second, a number of studies used homographs as study materials (Burt et al., 2021; Cyr & Anderson, 2018; Metcalfe & Huelser, 2020; Zawadzka & Hanczakowski, 2019). Here, when guessing is required, participants can either commit an error that is congruent with the meaning of the correct answer (e.g., cue: *party* – guess: *politics* – correct answer: *member*) or incongruent with it (e.g., guess: *balloons*). Cued-recall performance can then be assessed separately for congruent and incongruent guesses. It could be argued that prediction error is larger when guesses are incongruent with the meaning of the correct answer, in which case the prediction error account would expect better cued-recall performance as compared to congruent guesses. However, studies with homographs have either found an opposite pattern, with cued-recall performance benefitting from congruent guesses more than from incongruent guesses (Burt et al., 2021, Cyr & Anderson, 2018; Zawadzka & Hanczakowski, 2019), or found no difference in benefits conferred by guessing in these two conditions (Metcalfe & Huelser, 2020). Although, again, at first blush these results are inconsistent with the prediction error account, one could argue that this design is not perfectly suited for revealing the role of prediction error. The vital point is that while prediction error may well be larger for incongruent guesses, it is highly unlikely that this is the only process that differentiates the congruent and incongruent conditions. For incongruent guesses, the presentation of the target relating to a different meaning of a cue necessitates a host of processes related to reformulating the meaning of the cue – the search for an alternative meaning of the word ‘party’ in the aforementioned example, which one first interpreted as a social event but now one needs to understand as a political organization. These demanding processes of reinterpreting cues and linking the novel meaning of the cue to a matching target are likely to consume attentional resources to such an extent as to undermine subsequent associative memory. It thus seems likely that studies in which greater prediction error stems from an alternative interpretation of a homographic cue confound the magnitude of prediction error with processes that operate against the memory enhancement that the prediction error account would predict.

In this situation, a different route for assessing the prediction error account potentially holds more promise in directly revealing the workings of this mechanism in the guessing paradigm. This route is to manipulate the specificity of guesses afforded by the study materials, in which case more specific guesses, when disconfirmed, should result in greater prediction error and consequently better associative memory.

In essence, the challenge is to replicate the distinction between semantically related and unrelated materials in a way that would keep the to-be-learned materials across conditions as close to each other as possible, while still affording formulating guesses of varying specificity. A recent study by Gambi et al. (2021) provides the methods for accomplishing this goal, together with initial findings that seem at least consistent with the prediction error account of learning by guessing.

In their study, Gambi et al. (2021) presented novel words (e.g., *cheem*) in the context of either more or less constraining sentences (*Now, Peppa will eat the cheem* or *Now, Peppa will get the cheem*, respectively). Even before the sentences were presented, participants saw pictures of either a novel or a familiar object (e.g., apple) that could fit into the constraining sentences. Participants were expected to formulate a prediction – more specific for the constraining sentences – that the sentence would end with the name of a familiar object, only to be contradicted when a novel word, assumed to describe the novel object, was instead presented. According to the prediction error account, more specific predictions formulated for constraining sentences should lead to better learning of novel words – novel objects associations. This was indeed the pattern observed for adult participants in the study by Gambi et al. This result is consistent with the assumption that prediction error drives associative memory and shows that such effects can be revealed even for materials that lack any pre-experimental semantic association, as long as specific predictions can be derived from the accompanying contextual information such as constraining sentences. The study of Gambi et al., however, was not concerned with learning by guessing *per se* and thus it did not control for the process of formulating predictions – it created conditions under which participants were assumed to implicitly formulate predictions for all sentences and did not compare conditions with explicit predictions resulting from guessing and implicit predictions resulting from reading.

To reiterate, in Study 2 we aimed at directly assessing the role of prediction error in mediating the benefits of guessing. We used foreign language vocabulary as our study materials and hoped to demonstrate benefits of guessing for those materials in terms of supporting memory for word-translation associations. We reasoned that we failed in doing that in Study 1 because our participants had hard time formulating specific and informed guesses regarding the meaning of bare foreign words, and that such associative benefits can potentially be found when the guessing process becomes more informed and based upon some contextual information. For this purpose, we followed Gambi et al. (2021) and used the method of contextual sentences that varied the constrains they put on target words that were presented at the end of these sentences. Thus, the constraint could be either high – e.g., *A footballer kicked a* <u>*stick*</u>, or low – e.g., *In the park, he found a* <u>*stick*</u>. Because the focus here is on errorful learning, correct answers for all sentences were chosen in a way that would be inconsistent with most participants' responses. Memory performance for the guess condition, in which participants were explicitly required to provide their guesses concerning the target identity (which were subsequently negated by the actual translation of a foreign word) was compared against the baseline of the read condition, where participants were simply expected to read sentences ending in foreign words and their translations. To the extent to which prediction error drives the benefits of learning by guessing, we expected these benefits to emerge only with high-constraint sentences. Experiments 2.1a and 2.1b verified this prediction in relation to associative memory tested with foreign words serving as cues for retrieving correct translations. To foreshadow, both experiments resulted in unexpected findings suggesting that high-constraint sentences undermine associative learning. To shed light on these findings, Experiment 2.2 focused again on associative memory by using contextual sentences as cues in the final test, while Experiment 2.3 assessed the role of prediction error and guessing in target memory. The experiments together make a strong claim that the mechanisms of learning by guessing and learning via prediction error are distinct from each other, sometimes affecting different aspects of memory representations and sometimes affecting the same aspect independently of each other.

Experiments 2.1a and 2.1b

Experiments 2.1a and 2.1b assessed the potential role of prediction error in driving the associative benefits of guessing. Participants were presented with high- and low-constraint sentences in Polish (participants' native language) that ended with a Finnish word. For the guess condition, they were asked to provide their best guess as to what the Finnish word could mean. For the read condition, the translation of the Finnish word was presented outright. Because the focus of the current thesis is on errorful learning, the final words were chosen in such a way that they were unexpected in the context of a high-constraint sentence. Participants were unlikely to predict the meaning of the Finnish words correctly on the basis of either low- or high-constraint sentences, with the former ones being too general to allow specific guesses, and the latter ones leading participants to dominant incorrect responses instead of sensible but not dominant correct answers. The procedure ended with a forced-choice associative recognition test, in which participants were asked to choose the correct Polish translations of the previously studied Finnish words, with translations of other studied words serving as lures. As the correct translation and the lures were all studied in the learning phase, correct responding on this test required associating given translation with its Finnish counterpart.

Following Gambi et al. (2021), we assumed that high-constraint sentences should result in greater prediction error than low-constraint sentences. If prediction error drives augmented associative learning in the guessing paradigm, we expected the benefits of guessing to be revealed when learning Finnish-Polish translations in the context of high-constraint sentences. Experiments 2.1a and 2.1b were identical, with Experiment 2.1b attempting to replicate the patterns of Experiment 2.1a with a larger group of participants, recruited from an online platform rather than the student population.

Method

Participants*.* Thirty-nine undergraduate students from the SWPS university (age range 19-45 years, mean: 32) participated in Experiment 2.1a in exchange for course credit and 60 participants recruited via the Prolific platform (age range 18-45 years, mean: 31.5) participated in Experiment 2.1b in exchange for monetary compensation. All participants were fluent Polish speakers with no previous knowledge of Finnish. The study was approved by the Department of Psychology Ethics Committee at the SWPS University.

Materials and Design*.* Sixty-four pairs of Finnish-Polish words were used as study materials (see Appendix 2.1), and four additional pairs were used for practice. Each Finnish word was presented for study at the end of a sentence written in Polish. For each Finnish word, two sentences were created. One was a high-constraint sentence, for which participants were expected to have some confidence when trying to guess – albeit incorrectly – the meaning of the novel Finnish word, while the other was a low-constraint sentence, ambiguous in terms of the possible meaning of the embedded Finnish word.

An additional study with 20 participants recruited via Prolific was conducted to verify whether our materials indeed varied in terms of the constraint imposed by the sentences. In this study, participants were presented with Polish sentences ending with Finnish words and were asked to rate on a scale from 1 to 6 their confidence in being able to provide the correct translation. The time for providing these judgments was not limited. Participants were not asked to provide their actual guesses, and the correct translations were only presented at the very end of the procedure. Each participants saw half of the words embedded within sentences from the low-constraint condition and half within the sentences from the high-constraint condition, with the assignment of words to conditions counterbalanced between participants. This study showed that participants were more confident in being able to predict the translations in the high (M = 3.75, SD = 0.88) compared to the low-constraint condition (M = 2.42, SD = 0.91), $t(19) = 6.14$, $p < .001$, d = 1.37, consistent with the aims of our study.

The study list was divided in two, with each half being assigned to one study-test block. Within each block, 16 words from the study list were assigned to the guess condition, and the remaining 16 to the read condition. In the read condition, full translation was presented simultaneously below the sentence ending with the Finnish word. In the guess condition, the Finnish word alone appeared first below the sentence and participants had to guess and type in the word that they thought could be its Polish equivalent. After that, corrective feedback – the Finnish word with its Polish meaning – was presented.

The study phase was followed by a four-alternative forced-choice associative recognition test for the translations of the studied Finnish words. Participants were presented with Finnish words as cues which appeared with the correct translation (the target) and three familiar lures presented in a random order. All lures were chosen from the other translations which were studied in the earlier phase of the experiment, and so each studied word was presented four times throughout the test phase, once as a target and three times as a lure. Thus, the study had a 2 (learning condition: read, guess) x 2 (sentence type: high-constraint, low-constraint) within-participants design. The assignment of word pairs to conditions was counterbalanced across participants. A schematic representation of the experimental design for these and all subsequent experiments is presented in Figure 2.1.

Procedure. Participants were tested individually online. They were instructed that their task would be to learn Polish translations of Finnish words for a future test. They were then told that they would encounter those Finnish words in various sentences written in Polish. Also, they were informed about two types of study trials in the procedure. On some trials, besides the sentence ending with the Finnish word, the correct translation of the word would be presented outright for study for 13 s. Some other trials would require guessing and typing in a word that could be the translation of the presented Finnish word within 8 s, before being presented with the correct translation for 5 s. Participants were encouraged to use the sentence context for formulating their guesses. After these initial instructions, participants underwent a training phase consisting of studying and being tested on four word pairs (one from each learning condition). Then, participants completed the first study phase followed by a forced-choice associative recognition test. The test was followed by an identical second study-test block, albeit with different study and test materials.

<u>STUDY PHASE</u>

	GUESS	**READ**
LOW CONSTRAINT	On her trip to the zoo, she saw a very beautiful *kilpikonna*. kilpikonna - ??? (8 s) On her trip to the zoo, she saw a very beautiful *kilpikonna*. kilpikonna - tortoise (5 s)	On her trip to the zoo, she saw a very beautiful *kilpikonna*. kilpikonna - tortoise (13 s)
HIGH CONSTRAINT	She was scratching her arm because she was bitten by a *kilpikonna*. kilpikonna - ??? (8 s) She was scratching her arm because she was bitten by a *kilpikonna*. kilpikonna - tortoise (5 s)	She was scratching her arm because she was bitten by a *kilpikonna*. kilpikonna - tortoise (13 s)

<u>TEST PHASE</u>

Experiments 2.1a & 2.1b	ASSOCIATIVE RECOGNITION	kilpikonna - …… sister **tortoise** office chair *
Experiment 2.2	CUED RECALL OF TRANSLATIONS	On her trip to the zoo, she saw a very beautiful……. (**tortoise**)
Experiment 2.3	SIMPLE RECOGNITION	kilpikonna - …… accident football portrait **tortoise** cupboard **

* All lures were also translations of other studied Finnish words
**All lures were novel, unstudied words

Figure 2.1. A schematic design of the study conditions in Experiments 2.1a, 2.1b, 2.2 and 2.3. For convenience, the sentences and the lures are translated into English.

Results and Discussion

During the learning phase, 15 targets (8 from low-constraint and 7 from high-constraint sentences) were correctly guessed in Experiment 2.1a and 40 (14 from low-constraint and 26 from high-constraint sentences) in Experiment 2.1b. This resulted in an average of 1.2% (1.3% for low-constraint and 1.1% for high-constraint sentences) and 2.1% (1.5% for low-constraint and 2.7% for high-constraint sentences) trials with a correct answer, respectively, which were removed from subsequent analyses, as is common practice in studies on errorful learning (Kornell et al., 2009).

Table 2.1 presents performance on the forced-choice associative recognition test across conditions in Experiments 2.1a and 2.1b. We performed a repeated-measures analysis of variance (ANOVA) with sentence type (high-constraint, low-constraint) and learning condition (guess, read) as factors. In Experiment 2.1a this revealed a significant main effect of sentence type, $F(1,38) = 12.69$, $p = .001$, $\eta_p^2 = .25$. Overall, learning with low-constraint sentences ($M = .68$, $SD = .24$) resulted in better final associative recognition performance compared to learning with high-constraint sentences ($M = .61$, $SD = .26$). There was no main effect of the learning condition, $F(1,38) = 0.02$, $p = .88$, $\eta_p^2 = .001$, and the predicted interaction was also not significant, $F(1,38) = 0.29$, $p = .59$, $\eta_p^2 = .008$. Given our specific predictions concerning the interaction, we also computed a Bayes Factor for this effect, $BF_{(inclusion)} = 0.31$, which provides moderate evidence against the interaction.

Table 2.1

Forced-choice Associative Recognition Performance as a Function of Learning Condition and Sentence Type in Experiments 2.1a and 2.1b.

	Learning Condition	
Experiment and Sentence Type	Guess	Read
Experiment 2.1a		
High constraint	.60 (.04)	.62 (.04)
Low constraint	.68 (.04)	.67 (.04)
Experiment 2.1b		
High constraint	.65 (.03)	.72 (.02)
Low constraint	.69 (.02)	.69 (.03)

Note. Standard errors of the mean are given in parentheses.

For Experiment 2.1b, the main effect of sentence type was not significant, $F(1,59) = 0.13$, $p = .724$, $\eta_p^2 = .002$, but the main effect of learning condition was, $F(1,59) = 4.66$, $p = .035$, $\eta_p^2 = .07$. Overall, reading ($M = .71$, $SD = .20$) conferred a small benefit for final recognition performance compared to guessing ($M = .67$, $SD = .20$). This was, however, qualified by a significant interaction of the two factors, $F(1,59) = 5.51$, $p = .022$, $\eta_p^2 = .085$. The Bayes Factor for this interaction was $BF_{(inclusion)} = 2.31$, which provides only anecdotal evidence. This interaction was examined further in two ways. First, *t*-tests revealed that in the guess condition being presented with high-constraint sentences during learning resulted in worse subsequent associative recognition performance compared to low-constraint ones $t(59) = 2.15$, $p = .036$, $d = 0.28$. In the read condition, there was no significant difference between the sentence types, $t(59) = 1.39$, $p = .17$, $d = 0.18$. Second, *t*-tests revealed that for high-constraint sentences guessing resulted in worse subsequent associative recognition performance compared to reading, $t(59) = 3.47$, $p < .001$, $d = 0.45$. For low-constraint sentences, there was no significant difference between the learning conditions, $t(59) = 0.01$, $p = .995$, $d < 0.01$.

Even though results of both experiments are not fully consistent, they nevertheless point in the same direction. Learning foreign vocabulary was generally more effective when foreign words were embedded in low-constraint sentences rather than high-constraint sentences. This effect emerged independently of the requirement to produce overt guesses in Experiment 2.1a, while in Experiment 2.1b it emerged only when participants were required to provide explicit guesses, and not when they were merely asked to read sentences and translations of the final Finnish word. Generally, thus, it seems that high-constraint sentences had the power of undermining rather than enhancing the learning of the meaning of foreign words embedded in them.

For the current study, we predicted the guessing benefits to emerge only when specific guesses as to the meaning of the target could be formulated based on high-constraint sentences. The results were starkly inconsistent with this notion. While the mode of learning had no effect whatsoever on associative recognition performance in Experiment 2.1a, it had an effect that was exactly opposite to the predicted one in Experiment 2.1b, whereby trying to predict the meaning of the Finnish word undermined subsequent recognition performance when applied to high-constraint sentences. Our prediction was formulated to test the hypothesis that prediction error drives associative learning in guessing. We reasoned that if participants are required to guess the meaning of a foreign word embedded in a high-constraint context sentence, this should lead to relatively specific guesses as to the identity of the target, which then, when negated by the correct translation, should result in a large prediction error and the accompanying strengthening of the association between the foreign word and its correct translation, benefitting subsequent memory. The fact that no such effect occurred undermines the viability of the prediction error account of associative benefits of guessing.

While prediction error may not be responsible for the associative effects of guessing, the current results are also surprising from the perspective of extant theories arguing that prediction error drives learning more generally (Greve et al., 2017; Rumelhart & McClelland, 1986; Schomaker & Meeter, 2015). One could assume that the guess and read conditions are largely equated in terms of the contribution of prediction error to learning – with explicitly formulated guesses in the former condition and implicitly formulated guesses in the latter having comparable effects on learning – in which case learning via prediction error would be a mechanism independent from guessing. But this does not explain why the observed effects of prediction error were actually negative in Experiments 2.1a and 2.1b. Both the theory of learning via prediction error and the previous work using the method of high- and low-constraint sentences (Gambi et al., 2021) would lead us to believe that, if anything, high-constraint sentences should result in better learning than low-constraint sentences, which is the opposite of what was observed generally in Experiment 2.1a and more specifically for the guess condition in Experiment 2.1b.

Why should greater prediction error result in poorer learning? One possibility is that while greater prediction error leads to better learning of some aspect(s) of the stimulus, it is not the aspect that was assessed in Experiments 2.1a and 2.1b. We focused here on associative learning, which has been postulated to benefit from larger predictions error (Greve et al., 2017; Kafkas & Montaldi, 2018). This was defined for the current purpose as the strength of associations between foreign words and their translations. However, the way we implemented the difference between informed and random guesses had little to do with the stimuli for which associations needed to be created in order to perform successfully in the final test. This predictiveness at encoding was manipulated by varying constraints of contextual sentences that were not present as cues in the final test. It is thus possible that greater prediction error *did* support better associative learning, but only between elements of memory representations that served to formulate guesses – that is contextual sentences – and the answers that defied these predictions. Such an effect would not be detectable in our associative recognition test which assessed the strength of associations between foreign language words and their translations. Moreover, it is possible that such enhanced learning of associations between sources and targets of guesses, that we have yet to demonstrate within the present paradigm, comes at a cost for other aspects of memory representations. If greater prediction error leads to stronger associations between sentences that were used to generate guesses and answers that defied those guesses, then this may come at a cost to other types of associations that could be generated at the same time, such as associations between foreign words and their translations. This cost would be revealed as worse performance for the high-constraint sentence condition that we observed in Experiment 2.1a – and partially also in Experiment 2.1b. The next experiment tested the underlying assumption of this hypothesis – that prediction error does indeed enhance learning but only between the sources and targets of guesses.

Experiment 2.2

The present experiment again looked at the potential benefits of greater prediction error for associative learning in the guessing paradigm. Here we tested whether greater prediction error leads to stronger associations between the sources of guesses, which in this case were contextual sentences that pointed to potential translations of foreign words embedded in them, and the targets of those guesses, which were the translations themselves. For this purpose, we repeated the design of Experiments 2.1a and 2.1b but changed the final test to sentence-cued recall. We used cued recall here rather than associative recognition because although both tests tap associative learning, cued recall is more demanding and thus more appropriate for assessing memory for relatively easy to learn stimuli such as sentences and their endings.

Note that the more specific version of the prediction error account tested here – the benefits of prediction error to associations between sources and targets of guesses – is still a viable account of the benefits of guessing, as these benefits are most commonly observed when the same cues are used to generate guesses at encoding and prompt retrieval at test (Kornell et al., 2009; Zawadzka & Hanczakowski, 2019; but see Pan et al., 2019). If prediction error does strengthen associations between the sources and targets of guesses, and at the same time it remains a mechanism behind guessing, then we would expect the same interaction that was predicted (although not obtained) for Experiments 2.1a and 2.1b: the benefits of guessing over reading should emerge only in the condition in which informed rather than random guesses are formulated. In other words, here we assessed again whether the benefits of guessing would emerge for high-constraint sentences but not for low-constraint sentences.

Method

Participants. Forty-four students from the SWPS University (age range 19-50 years, mean: 34.5) with fluency in Polish but no previous knowledge of Finnish participated in the experiment in exchange for course credit. Five of those were excluded due to a failure to understand test instructions (i.e., trying to recall Finnish, not Polish words) or due to zero accuracy on the final test. This resulted in a total of 39 participants.

Materials, Design, and Procedure*.* The materials, design, and procedure closely resembled those from Experiment 2.1a and 2.1b, except for two differences. First, the format of the final test was cued recall, with sentences from the study phase serving as cues to recall and type in the Polish word which was presented with a given sentence (see Figure 2.1). Therefore, participants were explicitly instructed not to type in Finnish words and instead recall and type in their Polish translations. Sentences were presented in a random order and participants' time to type in their responses was not limited. Second, as this test was notably easier than the one used in Experiments 2.1a and 2.1b, we presented all 64 items in one study session followed by a single test, rather than split the procedure into two blocks as in the previous experiments. The design of the experiment was again 2 (learning condition: guess, read) x 2 (sentence type: high-constraint, low-constraint), with both factors manipulated within participants.

Results and Discussion

During the learning phase, 34 targets (25 from low-constraint and 9 from high-constraint sentences) were correctly guessed. This resulted in an average of 2.7% (4% for low-constraint and 1.4% for high-constraint sentences) correct responses at study, which were removed from all subsequent analyses.

Table 2.2. presents memory performance on the final cued-recall test. We performed a repeated-measures ANOVA with learning condition (guess vs. read) and sentence type (high- vs. low-constraint) as factors. This revealed a significant effect of sentence type, $F(1,38) = 164.23$, $p < .001$, $\eta_p^2 = .81$. High-constraint sentences ($M = .61$, $SD = .27$) resulted in better final recall performance than low-constraint ones ($M = .35$, $SD = .24$). Also, there was a main effect of learning condition, $F(1,38) = 129.7$, $p < .001$, $\eta_p^2 = .77$. Overall, guessing ($M = .61$, $SD = .27$) benefitted final recall performance over reading ($M = .35$, $SD = .24$). Crucially, the interaction was not significant, $F(1,38) = 0.67$, $p = .42$, $\eta_p^2 = .02$, with the Bayes Factor, $BF_{(inclusion)} = 0.31$, providing moderate evidence against it.

Table 2.2

Cued Recall Performance as a Function of Learning Condition and Sentence Type in Experiment 2.2.

	Learning Condition	
Sentence Type	Guess	Read
High constraint	.75 (.04)	.48 (.04)
Low constraint	.47 (.04)	.23 (.03)

Note. Standard errors of the mean are given in parentheses.

The results of the present experiment contrast starkly with the results of Experiments 2.1a and 2.1b. While previous experiments showed no benefits of guessing over reading for final associative memory performance, or sometimes even costs of guessing, the present experiment documented clear benefits of guessing. While previous experiments showed costs of learning from high-constraint sentences, the present experiment documented clear benefits of greater prediction error resulting from high- rather than low-constraint sentences. As such, the present results are consistent both with numerous studies showing that guessing improves associative memory (e.g., Kornell et al., 2009; Kornell, 2014; Richland et al., 2009), as well as studies indicating that more specific predictions augment memory for information that contravenes them (Gambi et al., 2021; Greve et al., 2017). What is, however, crucial for the present purpose is that the benefits of learning by guessing and of larger prediction error seem independent of each other. Greater prediction error clearly improved memory whether participants explicitly tried to guess the answer (a difference of .28 in the guess condition) or not (a difference of .25 in the read condition). Similarly, guessing clearly improved memory whether participants formulated their guesses based on high-constraint sentences (a difference of .27) or based on low-constraint sentences (a difference of .24). These effects contradict our initial hypothesis according to which learning via prediction error is responsible for the benefits of guessing. If prediction error was responsible for these benefits, we would not expect such robust benefits of guessing for low-constraint sentences, where prediction error should be minimal due to the nature of guesses afforded by ambiguous cues at encoding.

The fact that the present results were so different from the results of Experiments 2.1a and 2.1b indicates that it is vital to consider the aspects of memory representations affected by particular manipulations employed in the present study. We started by distinguishing between item and associative information, with both guessing and prediction error assumed to impact upon associative information, but the present results indicate that such a distinction is insufficiently specific. Associative information needs to be further divided into information that associates targets to either cues that serve to formulate guesses, or to peripheral information that is not used for the purpose of guessing. When different cues are used for guessing on the one hand (i.e., contextual sentences) and for retrieving information at the time of the final test on another (i.e., Finnish words), memory is unlikely to be improved by guessing. On the contrary, it may even be harmed by it, as in the case of high-constraint sentences generally in Experiment 2.1a, or high-constraint sentences with explicit predictions in Experiment 2.1b. It is only when the same cues are used for guessing and retrieval that associative memory benefits of these manipulations are observed.

The experiments presented thus far illuminate the impact of guessing on associative memory. So far, we have ignored the issue of item memory – memory for the target that is guessed at based on information available immediately before its presentation – as we hypothesized that it is at the level of associative memory where prediction error and guessing are likely to interact. The results of our experiments, however, indicate that while both guessing and prediction error affect associative memory in important ways, they do so independently of each other. Given this independence of the mechanisms of guessing and prediction error in terms of associative memory, it now becomes of interest whether the same independence can be documented for item memory. This was assessed in Experiment 2.3.

Experiment 2.3

The present experiment assessed the effects of both guessing and prediction error at encoding on memory for item information using an item recognition test. In general, there are two mechanisms that are likely to be involved in producing the benefits of guessing, with one of them giving rise to associative benefits when semantically related materials are used (Kornell et al., 2009), and the other giving rise to benefits to item memory both when there is a semantic relationship linking cues used to formulate guesses and the targets of those guesses (Zawadzka & Hanczakowski, 2019) and when this relationship is absent (Seabrooke et al., 2019). This latter mechanism has been argued to reflect greater attention devoted to targets that are presented as feedback after guessing compared to the same targets presented outright in the read condition (Potts et al., 2019). As we have argued here that the differences between semantically related and unrelated materials can be understood as a difference in materials affording informed guesses as opposed to materials that necessitate random guessing, it seems that target memory benefits both from informed and random guessing. But does it benefit to the same extent? Here we can assess this issue by introducing the manipulation of sentence context and directly comparing target recognition for conditions in which guesses were either informed by high-constraint sentences or were formulated at random when low-constraint sentences were used. We generally expected here better item memory for the guess condition compared to the read condition. If, however, the mechanisms of guessing are not fully independent of the mechanism of learning from prediction error, we would still expect these benefits to be larger for high- compared to low-constraint sentences.

Method

Participants. Thirty-nine students of the SWPS University, (age range 19-50 years, mean: 34.5) with fluency in Polish but no previous knowledge of Finnish participated in the experiment in exchange for course credit.

Materials, Design, and Procedure*.* The materials, design, and procedure were identical as in Experiment 2.2, bar the change in the final test format. A forced-choice recognition test was used here with the cue presented on each recognition test trial alongside the target and four novel lures, not presented earlier in the course of the experiment. Thus, there was one more lure presented compared to Experiments 2.1a and 2.1b, which also employed recognition tests. This was done to avoid ceiling effects as simple recognition is an easier task than associative recognition. Participants' task at test was to choose the word that was studied before as a translation of a Finnish word.

Results and Discussion

During the learning phase, 19 targets (11 from low-constraint and 8 from high-constraint sentences) were correctly guessed. This resulted in an average of 1.5% (1.8% for low-constraint and 1.3% for high-constraint sentences) correctly guessed items, which were removed from subsequent analyses.

Table 2.3 presents performance on the multiple-choice test across conditions in Experiment 2.3. We performed a repeated-measures ANOVA with learning condition (guess vs. read) and sentence type (high-constraint vs. low-constraint) as factors. This revealed a main effect of learning condition, $F(1,38) = 31.93$, $p < .001$, $\eta_p^2 = .46$. Overall, guessing ($M = .78$, $SD = .19$) resulted in better final recognition performance than reading ($M = .67$, $SD = .22$). There was no significant effect of sentence type, $F(1,38) = .40$, $p = .53$ $\eta_p^2 = .01$, and, crucially, there was no interaction, $F(1,38) = 0.08$, $p = .79$, $\eta_p^2 = .002$, with the Bayes Factor, $BF_{(inclusion)} = 0.24$, providing moderate evidence against it.

Table 2.3

Forced-choice Simple Recognition Performance as a Function of Learning Condition and Sentence Type in Experiment 2.3.

	Learning Condition	
Sentence Type	Guess	Read
High constraint	.78 (.03)	.67 (.03)
Low constraint	.78 (.03)	.66 (.04)

Note. Standard errors of the mean are given in parentheses.

The present results replicated the pattern of item memory benefits observed as a result of guessing (Seabrooke et al., 2019; Zawadzka & Hanczakowski, 2019). These benefits emerged whether guessing was informed by high-constraint sentences (a difference of .11) or not, i.e., when it was more random due to use of low-constraint sentences (a difference of .12). Clearly, thus, the mechanism responsible for the benefits of guessing for item memory is independent of learning via prediction error, in the same way that the mechanism responsible for the associative benefits is. In fact, while the present results confirm that guessing does affect item memory, they also suggest that the role of learning via prediction error for item memory is minimal. It seems thus that while learning via prediction error only enhances the strength of the association linking cues used to formulate guesses and targets that contradict those guesses (Experiment 2.2), the independent and positive effects of guessing are more general, affecting both associative and item memory alike.

General Discussion of Study 2

Two types of learning that involve committing errors have been postulated to benefit memory. In learning by guessing, participants are explicitly asked to formulate a guess as to the identity of the correct response, which is then revealed to be different from this guess. This process of guessing augments memory compared to reading the same information (Kornell et al., 2009). In learning via prediction error, participants are led to expect a certain stimulus, while the actual to-be-learned item defies those expectations, resulting in augmented memory (Greve et al., 2017; Quent et al., 2021). Here we assessed whether overlapping mechanisms underlie both memory benefits: that of learning by guessing and of learning via prediction error. Despite the intuitive appeal of such a common set of mechanisms operating whenever errors are committed during learning, the present results clearly indicate that memory is enhanced by explicit guesses and by prediction error due to different mechanisms.

We grounded the present study in a hypothesis that learning via prediction error was in fact the mechanism for the benefits of guessing observed when semantically related materials are studied. We reasoned that a commonly observed dissociation in the effects of guessing – that associative benefits of such learning emerge only when a clear semantic relationship links the cues used for guessing and their corresponding targets (Kornell et al., 2009), but not when such relationship is missing (Knight et al., 2012) or is opaque to participants (Seabrooke et al., 2019) – can be explained by assuming that only when a semantic relationship is present are the guesses formulated by participants informed, resulting in an appreciable prediction error when correct responses are presented. This account would suggest that guessing should improve memory whenever guesses are informed but not when they are random. While in the present study we focused first on learning foreign vocabulary, i.e., study materials for which associative benefits of guessing are absent (Butowska et al., 2022; Seabrooke et al., 2019), we hypothesized that the method of contextual sentences, previously used by Gambi et al. (2021) in their study of prediction error, would introduce informed guessing as to the meaning of foreign vocabulary when high-constraint sentences were used. If so, this could reveal the benefits of both prediction error and guessing.

Experiments 2.1a and 2.1b flatly contradicted our hypothesis, as we failed to find any benefits of guessing for participants' ability to learn the meanings of foreign words (Experiment 2.1a) – as assessed by the associative recognition task – or even found costs of guessing for high-constrained sentences (Experiment 2.1b), which constitutes a pattern directly opposite to the predicted one. For Experiment 2.2, we reasoned that the problem may not lie with the prediction error hypothesis itself but rather with the aspect of memory representation tapped by the final test. While guessing may strengthen associations between cues used to formulate a guess and the target that contradicts this guess, Experiment 2.1a and 2.1b actually assessed associations between targets and peripheral information in the form of foreign words embedded in contextual sentences. However, Experiment 2.2, which employed contextual sentences as cues in the final test, again failed to support the predicted interaction, and the benefits of guessing that emerged this time for both high- and low-constraint sentences were very similar in magnitude. While this confirms that it is important to consider which aspect of associative memory is affected by the guessing manipulation, the similarity of the benefits observed with informed and random guessing is again inconsistent with the prediction error account. Together, these three experiments seem to conclusively rule out learning from prediction error as a plausible mechanism of associative benefits of guessing. As such, the results provide empirical support to a recent suggestion by Brod (2021), who argued that learning via guessing and learning via predicting may be two largely unrelated phenomena of human memory.

Thus, in this study, we continued to fail to show any advantage of guessing at translations of foreign vocabulary in enhancing the association between the foreign word and its meaning. This refers both to more random and uninformed guessing at bare Finnish words (Experiments 1.1-1.3), and also to formulating a specific guess based on semantically rich contextual information provided by predictive sentences (Experiments 2.1a and 2.1b). Although we replicated some beneficial effects of guessing already described in the literature regarding enhanced item memory (Experiment 1.1 and Experiment 2.3), the memory aspect of most relevance in learning foreign languages – i.e., the association between the novel foreign word and its meaning – seemed to benefit more from reading than from guessing learning strategy.

At this point we decided to take a step back and review more closely the mechanisms of guessing when studying materials which were already shown to produce performance benefit. So far, the learning materials most reliably producing such benefit were weakly-related word pairs (e.g. Huelser & Metcalfe, 2012; Knight et al., 2012; Kornell et al., 2009;). However, it should be noticed that due to their simplified nature they may lack educational relevance. The pairings of the word associates are also arbitrary so that there is no actual reason as for why the correct target for the word *pond* is *frog* and not another related word such as for example *fish*. Also, word pairs may not reflect more complex dependencies which are usually intrinsic to study materials. In non-laboratory settings the learning materials are often complex and characterized by a binding semantic relationship; for example, when studying history, various events are interconnected and consequent of another. One example of study materials which show more of these quantities than weakly-related word pairs and which have already been suggested to benefit from guessing are general knowledge questions. They are semantically rich, their responses are not arbitrary but instead are closely related to the questions, which makes their ecological validity much higher compared to studying lists of word pairs.

Study 3[3]

In the last study of the current thesis we decided to explore the mechanisms underlying guessing when learning answers to trivia questions. We chose these materials for their educational applicability, such as for instance common use in exams from history (*Who was the first president of the United States?*) or geography (*What is the longest river in the world?*), as well as due to some promising results in demonstrating the effectiveness of guessing. As mentioned, Kornell (2014) showed that wrongly guessing at the answers to trivia questions does not interfere with the correct response but enhances memory instead. Interestingly and somehow contrary to the previous evidence (Hays et al., 2013; Vaughn & Rawson, 2012) – this performance benefit persisted over an increasing delay (from 6 minutes to 24 hours) introduced between the guessing attempt and feedback presentation. The results were interpreted in accord with the elaborative retrieval (Carpenter, 2009) and mediator (Pyc & Rawson, 2010) accounts, emphasizing that more semantically rich and complex materials, such as trivia questions, can produce long-term memory activation. That is because attempting to guess at meaningful question related to one's knowledge base is more engaging than attempting to guess at arbitrary word associate, which does not require such deeper activation and instead relies on short-lived priming processes (Kornell, 2014). Such differences in the duration of semantic activation have also educational relevance regarding the time delay with which the corrective information (e.g. in form of marked tests) should be delivered to students. According to Kornell (2014), teachers do not have to refrain from asking meaningful

[3] The five experiments described as Study 3 are currently under revision: Butowska, E., Hanczakowski, M., & Zawadzka, K. (2023). *Errorful learning of trivia questions and answers: The role of study time*.

questions which students fail to answer correctly because they feel that will impair learning if feedback is delayed. Contrarily, as long as the questions are rich in semantic content, unsuccessful attempts to answer them will enhance learning even when no immediate feedback is available.

However, one potential factor not taken into account in Kornell's (2014) study was the total time devoted for learning in the read and guess conditions. The guess condition involved a feedback phase, identical to the reading trial, 'enriched' by a preceding guessing attempt – in which participants were presented with the question alone. This resulted in a longer total study time (e.g. additional 12 seconds in Experiment 2 and self-timed three guessing attempts for each question in Experiments 3a and 3b) for each guessing trial when compared with reading trails. Hence, drawing valid comparisons between the two conditions could have been contaminated by the unaccounted difference in total study duration.

Interestingly, in an earlier study Kornell et al. (2009) were also looking at the effects of guessing strategy when studying fictional and non-fictional questions (Experiments 1 and 2). Crucially, though, the advantage of guessing over reading was only present in the first experiment, when the total duration of a learning trial in the guess condition (8 s for a guessing attempt + 5 s for feedback) was longer than in the read condition (5 s). The benefit disappeared in the second experiment in which the total study times were equated (13 seconds each). This finding, although observed cross-experimentally, strongly suggests a potential role of trial duration in mediating the effectiveness of guessing.

Overall, thus, the current knowledge is that the benefits of guessing at trivia questions seem readily observed when participants have additional time to formulate their guesses, but the status of this effect when learning duration is equated across guess and read conditions remains uncertain. Not only does this stand in clear contrast to studies using pairs of words as study materials, where clear benefits of guessing are found even when duration of study is equated across learning conditions, but also it potentially undermines the usefulness of the guessing strategy in many educational contexts. Does it make sense to require students to guess at answers they cannot possibly know when the same memory effect can be achieved simply by extending study time? Is there really an additional benefit resulting from engaging in futile retrieval attempt? We sought to answer these questions in the last study of the current thesis.

In short, in Study 3 we aimed to compare the effectiveness of guessing and reading strategies when studying trivia questions while accounting for the differences in total trial duration. As mentioned, Kornell et al. (2009) used two types of questions – fictional and non-fictional, which were analyzed separately. The dissociation, i.e., the guessing advantage over a shorter read condition (in which the time for presenting the question and the answer was equated with the feedback delivery phase of the guess condition), and no difference in performance in longer read condition (in which the time was equated with the full trial duration of guess condition), showed for the fictional questions, whereas for the non-fictional questions the results were inconclusive (no significant effect of learning condition in both experiments). As the results were less clear for non-fictional questions and also because these materials seem more educationally relevant and commonly used in the literature, we decided to focus on non-fictional trivia questions solely.

Study 3 consists of a series of five experiments involving learning answers to trivia questions either via guessing at correct answers or reading answers presented outright with their questions. In this examination, we employed larger samples of participants than the ones tested in the original study of Kornell et al. (2009) to ensure adequate power to detect any benefits accruing from guessing. We used true trivia questions of the sort used in the study by Kornell (2014) to ensure educational relevance of the examined learning strategy. The main aim of Experiment 3.1 was to replicate the results obtained cross-experimentally by Kornell et al. (2009) within one experimental design to examine the role of trial duration across the guess and read conditions, contrasting directly trials equating the whole duration of learning and trials equating the duration of processing answers. In Experiment 3.2, we focused more directly on trials that equated the duration of the whole learning episode. In Experiment 3.3, we attempted to isolate the effects of guessing with equated processing time for questions with familiar and unfamiliar answers. In Experiments 3.4 and 3.5, we assessed the role of semantic activation while guessing at trivia by requiring participants to guess an answer to a related question before learning the answer to the target question.

Experiment 3.1

Experiment 3.1 assessed the role of total trial duration in driving the performance differences between guessing and reading strategies. Participants were presented with various difficult trivia questions with their correct answers to study. The chosen questions were difficult because we did not want the participants to know their answers beforehand and aimed to minimize trials on which participants would correctly recall the response during the learning phase. For instance we used questions such as ‘*What nation consumes the most Coca-Cola per person?*’ or ‘*What company was the first to offer a mouse on a commercially available computer?*’. In the guess condition, participants were asked to type in within 8 seconds their best guess as to what the answer to a given question was, after which they were presented with corrective feedback for 5 seconds. In the short read condition, the response was presented outright, and the duration of the learning trial was equated with the duration of feedback presentation in the guess condition (5 seconds). In the long read condition, the response was also presented together with the question, but the duration of the learning trial was equated with the duration of the whole trial from the guess condition (13 seconds). After the learning phase, a final cued recall test followed, in which participants were presented with the trivia questions they studied earlier and asked to provide the correct response. If guessing leads to better encoding of answers to trivia questions, we would expect performance in the guess condition to be better than in both the short and long read conditions. However, if the time to process questions is vital, then we would expect performance in the guess condition to equal performance in the long read condition, with both outstripping performance in the short read condition.

Method

Participants. Forty-three participants recruited via the Prolific website (age range 24-68 years, mean: 38.4) participated in Experiment 3.1 in exchange for monetary compensation. Three participants were excluded due to close-to-zero accuracy on the final test, which gave a total of 40 participants. All participants were native English speakers. The study was approved by the Department of Psychology Ethics Committee at the SWPS University.

Materials and Design*.* Forty-eight trivia questions taken from the ones used by Kornell (2014) were used as study materials, and three additional new questions were used for practice. The question list was divided in three, with each third assigned to a different learning condition. In the long read condition, the question and the correct response were presented together for 13 seconds. In the short read condition, the question was presented together with the correct response for 5 seconds. In the guess condition, the question appeared first for 8 seconds, and after that the correct response appeared beneath the question for 5 seconds. The trial timings in Experiments 3.1-3.5 can be seen in Figure 3.1.

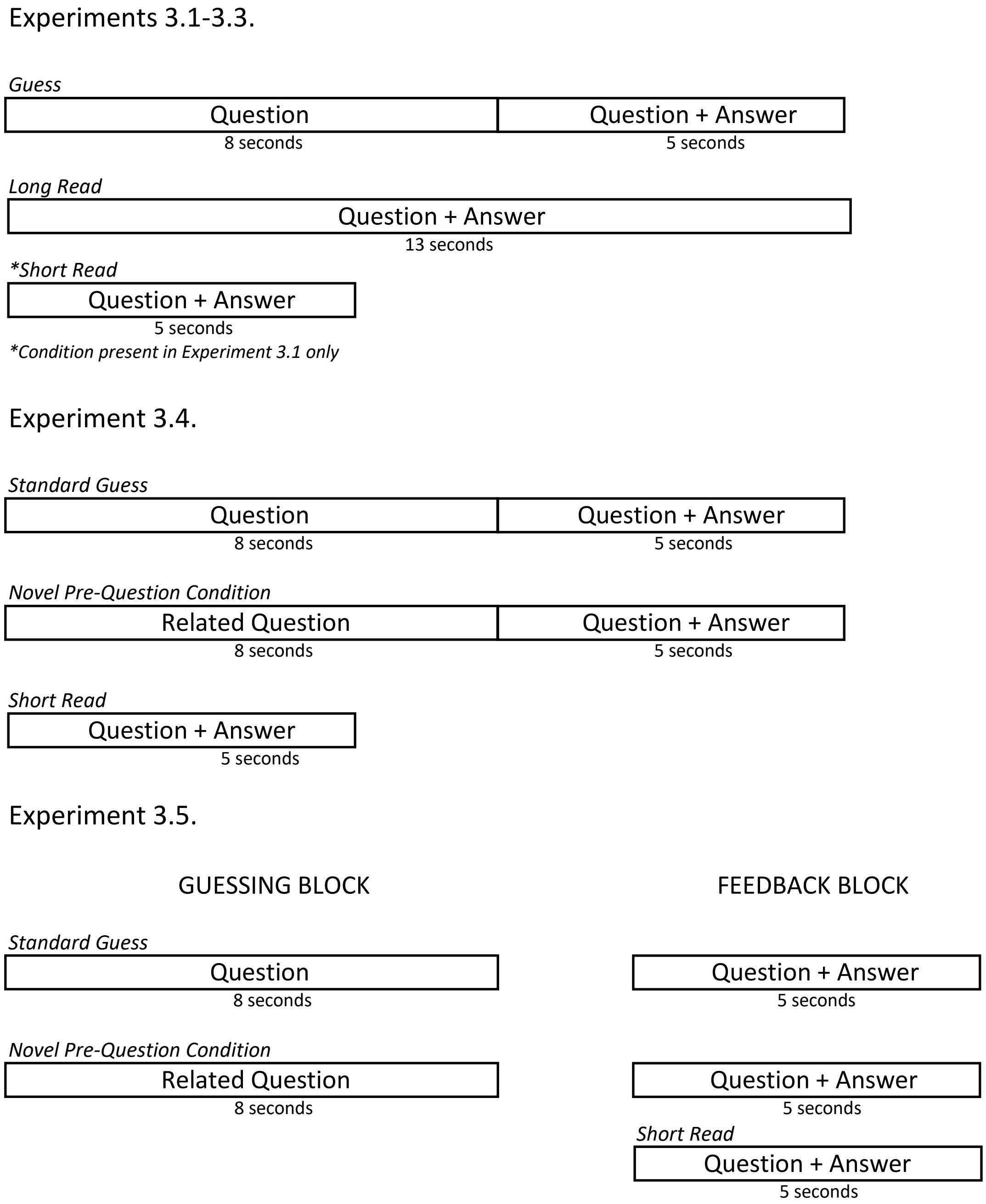

Figure 3.1. Study trial timings in Experiments 3.1-3.5.

The study phase was followed by a cued recall test for the correct responses to the studied questions. Participants were presented with the trivia questions as cues and asked to type in the response. Thus, the study had three learning conditions (long read vs. short read vs. guess) manipulated within participants. The assignment of questions to conditions was counterbalanced across participants, and the order of presentation of questions for study and test was randomized.

Procedure. Participants were tested individually online. They were instructed that their task would be to learn trivia. They were then told that they would encounter trivia questions with the answers in three learning scenarios. For some trials, the question would appear first and they would be required to type in their best guess as to what the answer might be within 8 seconds, before being provided with the correct answer for 5 seconds. For some other trials, the correct response would be presented outright with the question either for 13 or 5 seconds. Participants were also instructed not to confuse their initial guesses with the correct answers, and not to use any external resources and rely solely on their memory while doing the experiment.

After these initial instructions, participants engaged in a short training phase consisting of studying three trivia questions, one in each of the learning conditions. Then, they completed the study phase followed by a memory test which presented the trivia questions from the study phase as cues and required recalling the correct responses. The time for producing a response in the final memory test was unlimited.

Results and Discussion

During the learning phase, 53 questions were correctly answered in the guess condition, which constituted 8.3% of all trials. Since the focus of the study was on the effects of incorrect guessing on learning, these trials were removed from the analyses of final recall performance here and in all subsequent experiments, as is common practice in studies on errorful learning (Kornell et al., 2009). It is worth bearing in mind that such removal skews the recall results against any benefits in the guess condition. Questions for which correct responses are produced during learning are questions for which participants know the correct answers and thus would be highly likely to be answered correctly in the final test. Such questions are not removed from either of the read conditions, contributing to correct responding in these conditions. We return to this issue in General Discussion of Study 3.

The cued-recall performance across Experiments 3.1-3.5 is presented in Table 3.1. A one-way ANOVA on the proportion of correct responses on the final test showed a significant difference between the learning conditions, $F(2,78) = 5.62$, $p = .005$, $\eta_p^2 = .13$ This was followed up with a series of paired-sample *t*-tests to compare performance across the three learning conditions. The performance for guess and long read conditions was nearly identical and not significantly different, $t(39) = 0.013$, $p = .99$, $d < 0.01$. The performance for the short read condition was, on the other hand, significantly lower than in the guess, $t(39) = 2.74$, $p = .009$, $d = 0.43$, and in the long read, $t(39) = 2.72$, $p = .01$, $d = 0.43$, conditions.

Table 3.1

Cued Recall Performance as a Function of Learning Condition across Experiments 3.1-3.5.

	Learning Condition			
	Guess	Long Read	Short Read	Novel Pre-Question
Experiment 3.1	.83 (.03)	.83 (.03)	.77 (.04)	-
Experiment 3.2	.79 (.02)	.79 (.02)	-	-
Experiment 3.3				
Familiar Targets	.89 (.03)	.88 (.03)	-	-
Unfamiliar Targets	.33 (.04)	.39 (.04)	-	-
Experiment 3.4	.63 (.02)	-	.57 (.02)	.56 (.02)
Experiment 3.5	.61 (.02)	-	.57 (.03)	.56 (.02)

Note. Standard errors of the mean are given in parentheses.

The results of the first experiment of the series clearly show the influence of total study time on the effectiveness of learning, while at the same time again casting shadow on the role of guessing in supporting memory. When the total study time was equated, there was no difference in resulting memory from incorrect guessing and reading. As such, the present results replicate those reported by Kornell et al. (2009) in their investigation of guessing answers to fictional trivia questions. At the same time, both conditions using longer exposure to trivia questions produced better learning compared to the short read condition, where the time to process the questions was severely limited. This also remains consistent with previous studies where such benefits emerged (Kornell et al., 2009; Kornell, 2014). Still, before concluding that there is at best a limited benefit from engaging in guessing at trivia when total time to process question is held constant, we decided to conduct one more experiment, with more statistical power, and employing the Bayesian approach to quantify the evidence for a null effect.

Experiment 3.2

Experiment 3.2 included only two learning conditions – the guess and the long read condition. Thus, the sole aim of this experiment was to provide stronger evidence against learning by guessing supporting memory for answers to trivia questions once the time to process questions is equated with the time available in the read condition.

Method

Participants. Seventy-three participants recruited via Prolific website (age range 20-75 years, mean: 36.8) participated in Experiment 3.2 in exchange for monetary compensation. Two participants were excluded due to substantial (more than .50) accuracy when providing initial guesses while learning the questions, which gave a total of 71 participants. We aimed to recruit 70 participants and then assess the Bayes factor, assuming that we would continue testing if it remained uninformative, i.e., below 3 in support for either the null or the alternative hypotheses. All participants were native English speakers.

Materials, Design and Procedure*.* Materials, design, and procedure were identical to Experiment 3.1 except for one difference (see Figure 3.1). Namely, we excluded the short read condition, which resulted in a total of 24 (instead of 16) items studied and analyzed in each learning condition.

Results and Discussion

During the learning phase, 112 targets were correctly guessed. This resulted in 6.5 % of trials with a correct guess being removed from subsequent analyses.

Performance in the two learning conditions was nearly identical and not significantly different, $t(70) = 0.14$, $p = .89$, $d = 0.02$ (see Table 3.1). A Bayesian paired-samples *t*-test conducted in JASP (JASP Team, 2023) with its default priors yielded moderate evidence in support of the null hypothesis, $BF_{01} = 7.6$. Thus, the results of Experiment 3.2 are consistent with the results of Experiment 3.1 in failing to show any effect of learning condition when there is no difference in time devoted to processing questions. Whether participants formulated incorrect guesses or simply spent this time reading questions and answers together had no bearing on the effectiveness of learning. Once again, these results align with previous research on guessing in the study that used fictional trivia questions (Kornell et al., 2009) and cast shadow on the effectiveness of guessing as applied to materials other than weakly related pairs of words.

The obvious question at this point is why an effect readily observed with materials such as weakly related word pairs is absent when trivia questions and answers are studied. It is perhaps worth reiterating here that there are other situations in which guessing fails to support memory, and which involve learning the meanings of unfamiliar words such as, for example, words in a foreign language (Butowska et al., 2022, Seabrooke et al., 2019). Such conclusion also followed from the first two studies of the current thesis, which both failed to show guessing benefits for associative memory between foreign words and their translations even after modifying the learning paradigm in various theory-driven ways. In fact, in these situations guessing has been shown to result in poorer learning of associations between foreign words and their translations than simply reading those two together.

The crucial issue seems to be the transparency of the relationship between cues and their targets. When this relationship is transparent, as it is virtually always the case for weakly related pairs of words, guessing is beneficial, presumably because transparent semantic relationships allow for spreading of semantic activation caused by guessing. However, when this relationship is opaque, as necessarily occurs for all cases where one of the to-be-associated words is not known, semantic activation does not spread and guessing may even be detrimental when incorrect guesses interfere with learning or retrieval of correct answers.

When this transparency-based distinction is applied to trivia questions and answers, one could speculate that these materials may afford a mix of these two cases, with guessing facilitating learning of answers already perceived to be related to their questions but impeding learning of answers that are completely novel. A net result of this could be an overall lack of memory benefits of guessing answers to trivia questions. In Experiment 3.3, we attempted to disentangle these two situations by examining the effects of guessing separately in cases when the answers to a question were familiar concepts, presumably already related to their questions, and when these answers were novel concepts.

Experiment 3.3

In the present experiment we again assessed whether guessing improves learning of trivia questions and answers when the time of learning is equated with the control condition of reading. Here, we introduced a new factor – target familiarity. We hypothesized that while guessing may improve memory for trivia questions for which answers are familiar, it may impair memory for cases in which those answers are unfamiliar.

Method

Participants. Forty-one participants recruited via Prolific website (age range 21-85 years, mean: 38.5) participated in Experiment 3.3 in exchange for monetary compensation. All participants were native English speakers.

Materials, Design and Procedure*.* For the purpose of Experiment 3.3, we created a new set of learning materials consisting of 24 questions with familiar and 24 questions with unfamiliar answers which can be found in Appendix 3.1. Some of the questions were identical to those used in Experiments 3.1 and 3.2, while some were novel. We conducted a short pilot study on Prolific to assess the familiarity of the target answers. Participants were presented with 48 targets (half of them familiar and half unfamiliar), and were asked to rate how familiar those targets seemed to them on a scale from 1 (very unfamiliar) to 5 (very familiar). We first recruited 20 participants, and from their responses we calculated a mean familiarity score for each target and excluded those with the scores between 2 and 4. On this basis, we excluded and replaced 13 items. We then recruited another 20 participants to rate the familiarity of each target from the full set. This time all targets met our criteria. In this way, we ended up with a set of 24 unfamiliar (familiarity ratings between 1 and 2, M = 1.34) and 24 familiar (familiarity ratings between 4 and 5, M = 4.81) targets used as study materials.

The design of the experiment included two factors: learning condition (guess vs. read) and target familiarity (high vs. low), with both variables manipulated within participants. The assignment of questions to the two learning conditions was counterbalanced across participants, and the order of their presentation for study and for test was randomized. The procedure remained the same as in Experiment 3.2, with the total duration of a learning trial equated across guess and read conditions (13 seconds, with 8 seconds for guessing and 5 seconds for feedback in the guess condition, see Figure 3.1).

Results and Discussion

During the learning phase, 49 targets (41 familiar and 8 unfamiliar) were correctly provided as answers. This resulted in an average of 4.9% (8.3% for the familiar and 1.6% for the unfamiliar condition) trials with a correct answer, which were removed from subsequent analyses.

We performed a repeated-measures ANOVA with learning condition (guess vs. read) and target familiarity (high vs. low) as factors, (see Table 3.1). This yielded a significant main effect of target familiarity, $F(1,40) = 209.68$, $p < .001$, $\eta_p^2 = .84$. Overall, questions with familiar answers were answered correctly more often (M = .89, SD = .17) than questions with unfamiliar answers (M = .36, SD = .27). There was no significant main effect of learning condition, $F(1,40) = 2.62$, $p = .11$ $\eta_p^2 = .061$. Also, the interaction was not significant, $F(1,40) = 2.98$, $p = .09$, $\eta_p^2 = .07$. However, given our specific predictions formulated for the present experiment, we performed comparisons of the effectiveness of learning strategy separately for questions with familiar and unfamiliar answers. For familiar answers, there was no significant difference between the learning conditions, $t(40) = 0.24$, $p = .81$, $d = 0.04$. However, for unfamiliar answers, reading was significantly better in supporting subsequent memory than guessing, $t(40) = 2.18$, $p = .036$, $d = 0.34$.

In short, we again failed to observe the benefits of guessing over reading when learning trivia questions under conditions of equated total learning time. As could be expected, the familiarity of answers to trivia questions had a strong effect on learning – it was easier to learn familiar rather than unfamiliar answers. However, even for familiar answers – where trivia questions and their answers should have a pre-existing semantic relationship affording the way for semantic activation to spread – guessing failed to benefit memory. For unfamiliar answers, a small impairment in memory due to guessing was observed, which is consistent with studies showing a similar impairment when learning materials consist of translations of foreign vocabulary (Butowska et al., 2022; Seabrooke et al., 2019).

All experiments presented so far seem to question the effectiveness of guessing as a strategy of mastering educationally relevant materials. Once again, and notwithstanding previous studies controlling study time and pointing to a similar conclusion (Kornell et al., 2009), this is surprising inasmuch as current null results stand in contrast to results of studies employing related pairs of words as learning materials (Grimaldi & Karpicke, 2012). If spreading semantic activation resulting from guessing does support learning of such pairs of words, then why does it seem not to work for trivia questions? In Experiment 3.3 we assessed – but failed to confirm – whether the problem may lie on the side of the guess condition, where a mixture of familiar and unfamiliar answers could undermine the effectiveness of incorrect guessing. But another possibility is that the problem lies on the side of the read condition. It could be argued that trivia questions offer such semantically rich encoding opportunities that when they are processed for a substantially long time, semantic activation reaches the threshold of activation caused by incorrect guessing. In other words, under long encoding conditions – for which the guessing strategy does not supersede the reading strategy – semantic activation may plateau at the same level for both strategies. But when trivia questions are read for a relatively short time, activation has no time to accrue, resulting in a reduced effectiveness of encoding as compared to the guessing strategy. We aimed to assess this explanation in the next two experiments, where we attempted to de-confound the effects of guessing and learning time by requiring participants to guess not the answer to a question presented for learning but to a closely related question that should nevertheless activate a common semantic network.

Experiment 3.4

In the present experiment, we aimed to assess the effects of guessing an answer to a related question for subsequent learning of an answer to a target question. Thus, for example, participants could first be asked about the *tallest* US president and after formulating their guess, they would be presented with a question and the correct answer concerning the *shortest* US president. In this case, guessing an answer to a pre-question would serve to activate a semantic network common to both the pre-question and the target question, which is the heights of American presidents. This activation should in turn facilitate learning of the correct answer to the target question. The effects of such a learning strategy were compared both to the standard condition of guessing that requires guessing an answer to the target question itself, and to the short read condition. If semantic activation spreading before the target question is presented increases the effectiveness of learning, then we would expect the novel pre-question and the standard guess conditions to result in similar memory performance. At the same time, both conditions should result in better learning than the short read condition, where semantic activation should not have enough time to develop. However, if the previously observed differences across the guess and short read conditions reflect merely the difference in time of processing the target question, then we would expect the pre-question condition, in which no additional processing of the target question occurs, to result in learning effectiveness similar to that in the short read condition.

Method

Participants*.* Ninety participants recruited via Prolific website (age range 20-75 years, mean: 38.5) participated in Experiment 3.4 in exchange for monetary compensation.

Materials, Design and Procedure*.* In order to match the requirements of the novel pre-question condition we created a new set of 96 trivia questions and answers which can be found in Appendix 3.2. These were divided into pairs, with each pair consisting of two related questions activating the same response set. One question was always used as a target question for the learning task, while the other question was used in the pre-question condition as a primer preceding its related target question.

The design of Experiment 3.4 included three learning conditions: standard guess and short read, which were identical in structure to those used in Experiment 3.1, and the novel pre-question condition, which involved a guessing attempt (8 seconds) for a related pre-question, followed by an immediate presentation of the target trivia question together with its answer (5 seconds). Thus, the overall duration of the pre-question trial was equated with the standard guess condition (13 seconds), while the time to process the target trivia question and its answer was equated with the short read condition (5 seconds, see Figure 3.1).

The procedure of Experiment 3.4 was identical to previous experiments with one exception. Participants were warned that when asked to formulate their guesses during study, they will sometimes be presented not with the correct feedback to this question but rather with a related question and its answer, which they should learn in preparation for the final memory test.

Results and Discussion

During the learning phase, 73 answers to the target questions were correctly guessed in the standard guess condition, which resulted in 5% of trials with a correct guess removed from subsequent analyses. Also, 209 responses for the pre-questions were correct, constituting 14.5% of trials. Since the pre-questions were not tested themselves, these trials were not removed from the analysis.

A one-way ANOVA revealed significant differences in final test performance between learning conditions, $F(2,178) = 8.18$, $p < .001$, $\eta_p^2 = .08$, (see Table 3.1). To unpick that difference, we performed a series of paired-sample *t*-tests to compare performance across those conditions. Performance in the standard guess condition was significantly higher than in the pre-question condition, $t(89) = 3.63$, $p < .001$, $d = 0.38$ and in the short read condition, $t(89) = 3.10$, $p = .003$, $d = 0.33$. There was no significant difference between the short read and pre-question conditions ($t(89) = 0.73$, $p = .47$, $d = 0.08$).

The current results once again replicated the standard guessing benefit over the short read condition, also documented here in Experiment 3.1 and in previous studies on guessing using trivia questions and answers as the learning material (Kornell et al., 2009; Kornell, 2014). Crucially, though, the results were also quite clear in showing no memory benefit for the pre-question condition over the short read condition. These results once again argue against any specific benefit of formulating guesses in the context of learning trivia questions and answers. If guessing were to activate the semantic network, facilitating encoding of activated information, we would also expect guessing an answer to a pre-question to benefit learning of subsequent related target question, as long as answers to both questions belonged to a common semantic network. That such benefits failed to emerge suggests that the benefits observed in the standard guess condition are merely due to additional time to process the exact question for which an answer needs then to be learned.

Still, an alternative account for the present failure to find learning benefits in the pre-question condition needs to be considered. In this condition, we first asked participants to generate a response to a pre-question, only to immediately change the question to a similar one and ask to memorize it together with its response. This design generates potential challenges for participants, as they are forced to constantly switch between retrieval and encoding of various study materials. Such switches between retrieval and encoding may engender a cost to the effectiveness of the learning process. Finn and Roediger (2013) demonstrated that retrieving associations between studied faces and names hinders the incorporation of new associative information (profession) compared to restudying the same association. Also, Davis et al. (2017) showed that interpolating retrieval of old material with studying new material impaired new learning proportionally to the frequency of switching between retrieval and learning. Applied to the present paradigm, thus, switching between retrieval of potential answers to pre-questions and encoding of target questions could generate switch costs, potentially masking the beneficial effect of semantic activation resulting from guessing required for pre-questions. To assess this explanation of the apparent lack of learning benefits in the pre-question condition, in the next experiment we separated the phases of guessing and learning target questions together with their answers, in this way avoiding any switch costs in the pre-question condition.

Experiment 3.5

The main aim of Experiment 3.5 was to assess the effectiveness of learning trivia questions and their answers when preceded by a guessing attempt to a related pre-question. In this experiment, guesses for all questions – including guesses to related questions in the pre-question condition and guesses to target questions in the standard guess condition – were formulated in a separate phase of the experiment, before the proper study session in which target questions were presented together with their answers. The effectiveness of learning in the pre-question and standard guess conditions was once again compared against the baseline of the short read condition. Kornell (2014) showed that the benefits of guessing for trivia questions still emerge in a blocked design, with a delay between guessing and feedback presentation. Thus, we again expected learning to be more effective in the standard guess condition than in the short read condition. If this benefit stems from long-lasting semantic activation caused by guessing, then we would expect a similar benefit to emerge in the pre-question condition, where guesses are provided for related questions. If, however, the benefits in the guess condition reflect merely the increased time to process the target question itself, we again would expect the pre-question condition to result in no better learning than the short read condition.

Method

Materials, Design and Procedure*.* The materials and the design were identical to those from Experiment 3.4. The procedure was modified to avoid switching between retrieval and learning new information in the pre-question condition, (see Figure 3.1). In the guessing phase, 16 pre-questions and 16 target questions from the standard guess condition were presented for 8 seconds and the participants were asked to generate their best guess. In the following learning phase, all 48 target questions were presented simultaneously with their correct responses for 5 seconds each. This created an average delay of 4.75 minutes between the guessing attempt and learning phases in both the pre-question and standard guess conditions. After the learning phase, the final cued recall test followed, which was the same as in all the previous experiments.

Participants. Ninety participants recruited via the Prolific website (age range 19-64 years, mean: 45) participated in Experiment 3.5 in exchange for monetary compensation. One person was excluded due to close-to-zero final accuracy, which resulted in a total of 89 participants.

Results and Discussion

During the learning phase, 125 answers to the target questions were correctly guessed in the standard guess condition, which resulted in 8.8 % trials removed from subsequent analyses. Also, 247 responses for the pre-questions were correct, which constituted 17.3% of trials. These trials were not removed from the analysis.

A one-way ANOVA revealed that performance differed across the three learning conditions, $F(2,176) = 5.48$, $p = .005$, $\eta_p^2 = .06$, (see Table 3.1). We further performed a series of paired-sample *t*-tests to compare performance between those conditions. Performance in the standard guess condition was significantly higher than in the pre-question condition, $t(88) = 3.52$, $p < .001$, $d = 0.37$ and in the short read condition, $t(88) = 2.48$, $p = .015$, $d = 0.26$. There was no significant difference between the short read and pre-question conditions ($t(88) = 0.68$, $p = .50$, $d = 0.07$).

In short, the results of Experiment 3.5 closely resemble those of Experiment 3.4. Despite the elimination of the requirement to switch between retrieval and learning of new information in the pre-question condition, performance in this condition was still indistinguishable from the short read condition. Thus, we again failed to find any benefits of semantic activation triggered by guessing an answer to related question. At the same time, we did replicate the benefits to learning accruing from attempting to guess an answer to the target question. Since such benefits arise only when the exact same question is the target of guessing attempts and learning and these benefits are only observed when assessed against short reading times, it ultimately seems that the bulk of them stem from additional time to process the target question itself.

General Discussion of Study 3

In the last series of experiments we assessed the role of presentation timing in mediating the benefits of guessing over reading when learning trivia questions and their answers. We replicated the pattern of benefits of guessing over reading when additional time for a guessing attempt was provided (Kornell et al., 2009) in Experiments 3.1, 3.4, and 3.5. More importantly, however, guessing did not outperform reading when learning times were equated in Experiments 3.1, 3.2, and 3.3. Also, answering a related question did not facilitate the encoding of the following target question with its response, compared to studying the target question alone (Experiments 3.4 and 3.5). Thus, the results of Study 3 suggest that the benefits of guessing for studying trivia are elusive – they show that when additional time is provided for a guessing attempt and are gone when the same amount of time is allowed for the read condition. They also suggest that guessing has limited educational applicability, i.e., its effectiveness is comparable to a more simple reading strategy and so it can be used with no harm but also with no additional benefit for the memory of studied information and possibly only when one has a lot of time for practice.

One issue that merits additional discussion is the role of item-selection artifacts. Because learning by guessing is defined as learning after formulating incorrect guesses, we excluded from the analysis those questions for which correct responses were formulated when the question was first asked. If these questions are not eliminated, then it is unclear to what extent subsequent memory performance reflects the effects of guessing or strengthening of already known answers via the mechanism of the testing effect (Kornell, 2014). However, as noted by Kornell et al. (2009), excluding correctly answered questions from the guess condition puts this condition at a disadvantage compared to the read condition. This is because correctly responded questions that are excluded from the guess condition are necessarily the easiest questions that still contribute to correct responding in the read condition. Could our failure to document benefits of guessing reflect merely a disadvantage of this strategy due to an item-selection artifact? We consider this possibility unlikely for two reasons. First, Kornell et al. (2009) addressed the issue of item-selection artifacts by examining learning by guessing in the context of fictional trivia questions for which no correct answer exists. In this case, no exclusions were necessary, yet the patterns of results – which inspired the present investigation – were the same: no benefits of guessing over reading with equated study time.

Second, we also re-analyzed the results of Experiments 3.1-3.3, where guessing was contrasted with the long read condition, but without excluding questions correctly answered initially in the guess condition. Such an analysis now puts the guess condition at an advantage due to a small possible contribution of the testing effect to learning. For Experiment 3.3 only the familiar target condition was included. A 2 (condition: guess vs. long read) x 3 (Experiment: 3.1, 3.2, 3.3) mixed ANOVA yielded a significant main effect of experiment, $F(2,149) = 5.07$, $p = .007$, $\eta_p^2 = .064$, which reflected differences in final test performance across experiments, while the main effect of learning condition and the interaction were not significant, $F(1,149) = 1.27$, $p = .26$, $\eta_p^2 = .008$, and $F(2,149) = .01$, $p = .99$, $\eta_p^2 < .001$, respectively. Thus, even without excluding any responses the guess condition does not seem to facilitate subsequent memory performance over and above a simple reading strategy.

In short, Study 3 clearly underlined that the differences in learning times are crucial for final performance. A question which follows is how exactly this additional time is used by students. A more parsimonious account would suggest that these additional 8 seconds in the guess condition were used to understand the question, which would later help to associate it with the feedback. Alternatively, it could be that guessing process activates a specific mechanism responsible for the later performance increase – one of semantic activation of potential responses, which may also include the target answer, which later helps to encode the correct feedback. However, Experiments 3.4 and 3.5 failed to find any evidence for the role of spreading semantic activation when guessing at trivia questions. They both showed that engaging in a guessing attempt at a related question (which should result in the activation of a similar response set) did not facilitate learning of the following answers to target questions – both when these were presented immediately after a related pre-question in Experiment 3.4 and when a delay was introduced between pre-questions and their related target questions in Experiment 3.5. These results present a strong limitation to the effectiveness of guessing, namely that it aids learning only when the exact same questions are used during the guessing attempt and feedback presentation stage (i.e., in the standard guess condition). When a different, yet related information is quired in the guessing attempt, no performance increase for later related materials is observed.

At the same time we appreciate that the most direct test of the effects of semantic activation during a guessing attempt would be to compare a condition in which participants are either to guess at the question or are instructed to read it without performing any elaboration on the potential answers. In this scenario, participants would not experience any task switching costs and also would not need to generate a set of responses to a pre-question, which may require some existing knowledge base. However, in such scenario there would be no way of verifying whether people actually follow the instructions and are able to effectively refrain from guessing at a presented question or automatically formulate guesses, despite what the instruction said. This is why we did not introduce such instruction manipulation in our experiments.

General Discussion

Guessing has recently been proposed as an effective, readily implemented and educationally relevant learning strategy (Kornell et al., 2009, Potts & Shanks, 2014, Richland et al., 2009). This was quite revolutionary, especially when considering the tradition of errorless approach to learning, which postulates that increasing the potential to err in the learning process can only harm performance, mainly due to increased interference from the incorrect responses generated at the guessing attempt.

Following this novel direction of research into learning effectiveness, we aimed to establish whether learning by guessing can be effective focusing on scenarios in which its benefits have so far been shown to be limited. However, as evident throughout the present study, we failed to document much benefits of guessing over reading, suggesting that guessing is a strategy suited only to learn familiar words linked by a semantic relationship (Experiment 1.4). When studying other materials, however, we have provided evidence that guessing at best does not harm performance compared to reading, but at worst it actually produces costs to learning. For foreign words, that was the case both when the translations were presented alone (Experiments 1.1-1.3) and also when the foreign words were presented within a cueing sentence (Experiment 2.1). What is more, after changing materials to trivia questions in our last project, we also did not observe any guessing benefits over those afforded by longer time spent studying the question, which clearly does not advocate for promoting learning by guessing as a more demanding yet effective way of studying. Below, I will analyze our results from the perspectives of various accounts of guessing effectiveness offered by the existing literature.

As mentioned throughout the manuscript, the accounts of guessing benefits for the final performance can be divided into two main classes. The first one underlines the role of semantic processes operating while formulating a guess and applies specifically to studying materials with an inherent semantic relationship binding the cues with the targets such as e.g. weakly-related word pairs. Accounts focusing on the semantic relationship between the cue and the target include the search set account (Grimaldi & Karpicke, 2012) and the elaborative retrieval account (Carpenter, 2009, 2011).

The second class of explanations does not rely on any semantic processes and can be applied to learning both semantically related and unrelated sets of materials. In general, it postulates the attentional up-regulation of feedback encoding following a guessing attempt (Overman et al., 2021) and can account for the item memory benefits of guessing (Zawadzka & Hanczakowski, 2019). Therefore, it can for instance readily explain better target memory for studied translations when guessing at a foreign word's meaning rather than just reading the translation outright (Seabrooke et al., 2019; Seabrooke et al., 2022).

Starting with the first group of accounts, the explanations here rely on some sort of semantic elaboration (see Zawadzka et al., 2023) of the presented cue, which later benefits feedback encoding. As mentioned, these accounts refer only to semantically related materials and do not explain the guessing benefits when learning unrelated materials as the novel unfamiliar targets cannot be, by definition, activated upon elaboration on their cues. Accounts underling the significance of the semantic relationship within the study material include the already described search set theory (Grimaldi & Karpicke, 2012) which states that the guessing attempt results in a spreading activation of semantically relevant concepts, including the correct response. A similar idea has also been described by Carpenter (2009, 2011), who offered an elaborative retrieval hypothesis of the testing effect. It refers to the activation of information related to the cue during retrieval, which also may spread to the target and explains the benefits of testing over restudying. As these hypotheses focus on the strength of the semantic relationship between the cues and targets, I will mainly analyze how they fit the results of our experiments which used semantically related learning stimuli, i.e., Experiment 1.4, which used weakly-related word pairs and also the last series of Experiments 3.1-3.5 which used general knowledge questions, whose responses were inherently related to their cues.

In short, the semantic elaboration accounts correctly predict the pattern of results in Experiment 1.4 where we observed the guessing advantage over reading in the final cued-recall test of related word pairs. This replicates the results of many previous studies (Huelser & Metcalfe, 2012; Knight et al., 2012; Kornell et al., 2009) which adds to an already solid base of studies showing the associative guessing benefit when studying weakly-related word pairs. Also, the account does not make any predictions regarding the interaction of guessing with restudying, which stays in accord with a no observable interaction of learning condition with restudying in Experiment 1.4.

Moving onto the experiments employing trivia questions, the semantic elaboration accounts would predict that participants, while attempting to guess at the answer to the presented question, should generate a pool of responses, possibly including the target, which should enhance target encoding once the corrective feedback is presented. This activation should show as a memory benefit over the read condition. We did observe the guessing benefit over our short read condition, however, it disappeared after accounting for the differences in study time with the introduction of the long read condition. Experiments 3.4 and 3.5 directly tested the search set account by introducing related pre-questions which were supposed to activate a response set relevant also to the test question and so enhance encoding of its answer. However, it turned out that elaboration on the related question did not increase final accuracy, which was similar to simply studying the test question outright without any additional elaboration of the response set.

The questions arises, why did our participants failed to use the semantic activation to increase final performance when learning trivia, and succeeded in using it when studying related word pairs? In order to explain this discrepancy we postulate that the associative benefits of guessing when studying semantically related materials are a function of the strength of this relationship as perceived by the participants. Weakly-related word pairs, used in Experiment 1.4, were chosen from associative norms which are rich in pre-existing semantic associations apparent to most learners (Brainerd et al., 2023). Thus, this type of learning material seems uniquely suited to revealing the role of spreading semantic activation that can result from guessing. By contrast, materials such as trivia questions and answers constitute a mix of elements of various levels of semantic transparency, ranging from completely novel arrangements, where nothing in the question points to a correct response, to examples of marginal knowledge (Cantor et al., 2015), where the semantic association is already established but is not strong enough to allow for correct responding without another presentation of the correct answer. It can be that the low strength of semantic relationship may results in costs to guessing as evident in Experiment 3.3 in the unfamiliar targets condition. One can argue that Experiments 1.1-1.3 and 2.1 which investigated associative memory for foreign words and their translations also used semantically related materials, because the Finnish and Polish words had the very same meaning. This relationship, however, was clearly opaque to our participants and so the perceived strength of the relationship between Finnish and Polish words was very low and resulted in reading outperforming guessing on the final associative test. In other words, it seems that the conditions which are automatically in place when studying related word pairs (high probability of generating a response closely related to the target) are much less likely to be met when guessing at a general knowledge question, not to mention unfamiliar Finnish words. We may easily image that when people are guessing at a question: '*How do you call a raised curved part at the back of a saddle?'* – they may struggle with generating any relevant responses (unless they are equestrians) and so their elaboration on the saddle parts is insufficient to increase and may even decrease final performance.

There is one more account which can be used to explain the associative benefits of guessing without necessarily restricting study materials to those which are semantically related – namely the episodic mediator hypothesis (Metcalfe & Huelser, 2020). According to this explanation, the guessing attempt may serve to create an additional route by which targets can be retrieved during test. Cues used to generate guesses become associated with the guesses, which in turn become associated with correct targets. When these cues are then presented at retrieval, participants may retrieve first the guesses, and then use the additional link from their guess to the target to improve their performance compared to the read condition, where no additional mediation is likely to occur. This account, at least in theory, may apply to studying both related and unrelated materials as the generated guess, if retrieved at test, can serve to recall even unrelated target.

To test this account directly, one should obtain a measure of guess recollection at test and see how this affects final accuracy. In the current thesis, we included such measure in Experiments 1.2 and 1.3 where we asked our participants to recall their initial guess at final test. On the one hand, our results seem to align with the episodic mediator account – as successfully recalling one's own guess at a given Finnish word helped to recall the target. At the same time, this performance increase only matched and was not higher than the performance afforded by simply reading the translations. Also, probably because the cues and the targets (Finnish words and their translations) seemed unrelated to participants, they struggled to recall their initial guesses on the final test as their success in doing so stayed well below 50% even after providing additional cues to increase the recall. This suggests that for a successful episodic mediation to take place, there should be some more apparent link between the cue, the guess, and the target, which is clearly missing when studying foreign translations. From this perspective, the episodic mediator hypothesis account also chimes with the results of Experiment 2.1a and 2.1b, which showed no guessing benefit on the final associative test. This may be because the guesses generated in the learning phase, even if informed by the cueing sentences, were hard to associate with the novel Finnish words and therefore could not serve as effective mediators on the final test.

Although we did not include a direct test of the episodic mediator hypothesis in the second project, we believe that this hypothesis may be well suited for explaining also some other results obtained there. In Experiment 2.2 we asked the participants to recall the endings of the sentences they saw in the learning phase and guessing showed a performance increase over reading for both high- and low-constraint sentence types. Importantly, the sentences we used were composed deliberately so that the correct ending was not likely to be guessed neither in the high nor in the low constrain condition. The semantic activation accounts struggle to explain these results for two reasons. Firstly, they do not predict any guessing benefits as the target was unlikely to be activated by the guessing attempt. Secondly, they predict some differences driven by the manipulation of sentence type, which were also absent from our results as guessing benefitted performance for high- and low-constraint sentences similarly. However, it seems plausible that the participants, when presented with the same sentence on final test, recalled first their incorrect guess, which then helped them to access the correct response and increased the accuracy, as suggested by the episodic mediation. Also, the episodic mediator account aligns with the lack of a significant main effect of sentence type, as both high- and low- constraint sentences can by hypothesized to be equally easy to associate with one’s guesses, which in turn become useful mediators on the final cued-recall test. Of note, Experiment 2.2 was the only one in the current series which showed the associative benefits of guessing after accounting for the timing differences (except for Experiment 1.4, which replicated the already known guessing benefit when studying related word pairs). This is notable, especially minding a pattern of null findings from our third project. One may suspect that the trivia questions were harder to associate with the generated guesses compared to cueing sentences for which producing any sensible guess was probably more automatic and less reliant on participants’ factual knowledge base. This may be why we did not observe any associative guessing benefits when studying trivia and manage to find some when studying sentences and their endings. From this perspective, the episodic mediator hypothesis seems to capture some genuine benefits of guessing, which lead to this strategy outperforming the reading strategy even when the learning times are the same.

As already mentioned, the main aim of our second project was to investigate whether a mechanism of prediction error can explain the associative benefits of guessing. One of the appeals of the prediction error account is that it can be potentially used when investigating the associative performance benefits when learning materials which do not have a clear semantic relationship between the cueing information and the target (similarly to the episodic mediator account described above). This is because the account underlines both the semantic process active during guess generation and also the role of surprise at unexpected outcome in mediating performance benefits. The idea is related to the hypercorrection effect (Butterfield & Metcalfe, 2001) which reveals better memory for corrective feedback when the incorrect response was given with high rather than low confidence. Such confident although incorrect responses can result in a feeling of surprise upon delivery of the corrective feedback which, in turn, translates to its preferable encoding and higher performance (Brod & Breitwieser, 2019). Foremost, this account refers only to specific learning materials which allow participants to form a confident, yet incorrect guess, which can later be disconfirmed by a surprising target. Also, the more the actual outcome differs from the predicted one, the higher the surprise response which can fuel later memory performance (Brod et al., 2018; Theobald & Brod, 2021).

In the second project we found that the benefits of prediction error (i.e. surprise) and of guessing were independent of each other. We observed that the associative benefits of surprise can show only when the same information which led to formulating an incorrect guess is repeated on the final test (Experiment 2.2), and do not transfer to the associations between items peripheral to the cueing information (Experiment 2.1). Of note, in Experiment 2.3, which focused on item memory, we found only one main effect of learning condition, with guessing outperforming reading regardless of sentence type. Thus, it seems that guessing enhances both item (Experiment 2.3) and associative memory, the latter when semantically related materials are used. The effects of surprise, however, are more limited and specifically aid the associations between cues used to formulate the predictions and the surprising targets (Experiment 2.2) and do not transfer to the item memory benefit as we observed no effect of sentence type in Experiment 2.3.

As the surprise account refers to study materials which allow to formulate an incorrect yet confident guess, it cannot be applied to our first project, where guessing was rather random, or the last one – in which our participants were also assumed neither to know the answers nor to respond with high confidence. However, materials which seem naturally suited to test if making confident mistakes when studying foreign words are the so-called false cognates, which are words similar to another word in different language but have in fact different meaning, e.g. Spanish CARPETA means *folder* and apart from superficial resemblance has little to do with the meaning of English CARPET. Cyr & Anderson (2018) found that guessing at foreign words which match the expectations (e.g. PARIENTE for PARENT) is more helpful in terms of final test performance than wrongly guessing at the false cognates. This directly contradicts the surprise account of learning benefits and instead suggests that guessing at foreign translations can benefit memory only to the extent to which the guesses semantically overlap with the targets, which by the way resembles studies on related word pairs. When the guesses do not tap into this semantic proximity of correct meaning, they seem to hinder foreign vocabulary learning, no matter if they were informed and committed with high confidence or uninformed and committed with low confidence.

Now I move onto the second class of accounts, which focuses on the item memory only and explains guessing benefits when studying both semantically related and unrelated materials. These accounts underline the role of curiosity, which precedes feedback presentation and is assumed to be resulting in performance increase. In other words, trying to answer a question can cause a state of higher curiosity, which is postulated to enhance the encoding of the answer and also of some peripheral information (Gruber & Ranganath, 2019, but see Hollins et al., 2023). In general, guessing is assumed to afford higher curiosity than reading, which was confirmed by a set of studies which compared the effects of predictions (attempt to generate a response before it was presented) with the so-called post-dictions (attempt to generate a response after the correct one was presented, Potts et al., 2019). Beneficial performance was observed only when guessing preceded the presentation of the correct answer, but not when participants were asked to state what their guess would have been only after the correct answer is presented.

However, as underlined throughout the manuscript, our main interest was the associative benefit of guessing and we did not focus on its potential to improve item memory, which is thought to benefit from higher curiosity. The experiments which included a recognition test sensitive to target memory were Experiment 1.1 and Experiment 2.3, and they both showed a guessing advantage over reading for the memory of studied native words, which chimes with the prediction of the curiosity account. Also, in our second project we included a direct measure of participants' curiosity. As the issue of participants' curiosity was not of our primary research interest, we decided not to include the following results in the direct description of

Study 2 and instead report them here, where we discuss the results from different theoretical viewpoints. We included the direct measure of participants curiosity, because on the one hand, participants could be more curious whether their answers were correct when the cueing sentence allowed for more specific guesses, or, on the other hand, they could be more curious when the sentences were ambiguous. Thus, we tested a group of 20 participants who saw the Polish sentences ending with the Finnish words and were asked to rate their curiosity as to what the translation of the Finnish word could be, without having to guess or being given any feedback until the end of the experiment. Half of the words were presented with high-constraint sentences, and half with low-constraint sentences, which was counterbalanced across participants. It turned out that participants were more curious as to the meanings of Finnish words presented at the end of low-constraint (M = 4.43, SD = 0.89) rather than high-constraint sentences (M = 3.49, SD = 1.07), $t(19) = 5.11$, $p < .001$, d = 1.14.

Therefore, apart for the guessing benefit over reading, the curiosity account predicts that low-constraint sentences should afford higher performance than high-constraint ones on the final recognition measure. This is not what was observed in Experiment 2.3. which yielded only one main effect of learning condition, with guessing outperforming reading similarly regardless of sentence type. What is more, we failed to observe an interaction of learning condition with sentence type, which was also predicted by the curiosity account and would show as the more pronounced benefits of guessing for the low-compared to high-constraint sentences.

Another aspect of our results which is of note from a theoretical viewpoint is the presence of dissociations obtained between the learning conditions and other manipulations such as restudying materials (Experiments 1.1-1.4) and prediction error (Experiments 2.1-2.3). Starting with the former, despite research showing that greater familiarity of individual components of to-be-learned materials should aid in associating these components (Gagnepain et al., 2008), we have failed to observe such a pattern in our own study. While in Experiment 1.1, using foreign vocabulary, we confirmed – by assessing recognition performance – better target memory for individual targets studied in the guessing with feedback condition (see also Seabrooke et al., 2019), we can also assume similar benefits when related pairs of words were studied in Experiment 1.4 based on previous work (Zawadzka & Hanczakowski, 2019). Still, in both cases it did not appear to be easier for participants to associate stronger targets with cues at restudy. It is worth noting that recent years saw a surge in studies assessing the role of item memory in encoding associations, and conclusions from this literature are not clear, with some studies showing better encoding of associations for stronger items (e.g., Greve et al., 2017; Poppenk & Norman, 2012), while other showing the opposite pattern (Kim et al., 2012). Recent work by Lee et al. (2020) suggested a modulating role of pre-experimental familiarity of study materials, with novel stimuli yielding a benefit of increased item strength for association formation and familiar stimuli yielding a cost. However, in Experiments 1.1-1.3 we used stimuli that were completely novel to participants and in Experiment 1.4 we used stimuli – words – that were familiar, and the results were the same, with equal associative memory across conditions differing in item strength before restudy. This consistent null pattern does not fit any theories that try to account for either benefits or costs observed in previous studies. This issue clearly awaits further research.

Moving to the second series of experiments focusing on learning by prediction error, as already mentioned throughout the manuscript – we found that the manipulations of guessing and prediction error led to independent effects. In Experiment 2.1a, greater prediction error stemming from contextual sentences directing participants towards specific but incorrect meanings of foreign words embedded in these sentences negatively affected learning of these words' meanings. At the same time, the manipulation of explicit guessing and reading failed to impact memory for those meanings. In Experiment 2.2, when contextual sentences were used as cues on the final test, the magnitude of prediction error was positively related to memory for the target words and so was the requirement to guess these final words at encoding. However, these two effects were independent of each other.

Finally, Experiment 2.3, which assessed memory for translations of foreign language words via a simple recognition test, yielded results that were directly opposite to the results of Experiment 2.1a. This time guessing affected item memory, producing the expected benefits, but the magnitude of prediction error had no effect. The only interactive effect between guessing and the magnitude of prediction error was observed in Experiment 2.1b, but even there it was not related to the benefits of committing errors during learning, as it actually reflected *worse* associative memory for foreign language translations when prediction error was larger and guesses were explicitly formulated. In short, while both guessing and learning via prediction error have the capacity of augmenting at least some aspects of memory representations, they do it via different psychological mechanisms. Notably, we were not the first to suggest that the mechanisms of guessing and forming a specific, although incorrect prediction are independent of each other. Helpful in understanding why that can be the case is a distinction made by Brod (2021) between a less informed and more random *guessing,* and *predicting,* which can be operationalized as a special instance of guessing when participants can base their speculations on some relatively solid premises, increasing their confidence in predicted responses. According to Brod (2021), it is possible that these two approaches to learning rely on different mechanisms, and thus should not be confused with one another.

In the Introduction I mentioned that the main challenge in designing desirable difficulties is to find a balance between adding too much or too little hardship to the learning process. Also, increasing the difficulty of a study session can only benefit learners when they already possess some basic pre-experimental knowledge and cognitive ability. From this perspective our lack of positive findings for studying unfamiliar foreign words or trivia questions with difficult answers seems understandable. In other words, our learning paradigm did not include any pre-study session with a list of translations or established a necessary knowledge structure our participants could refer to so that the benefit of added challenges could follow.

An interesting theory which may prove useful when designing the optimal level of difficulty added for effective learning is the New Theory of Disuse (Bjork & Bjork, 1992). It distinguishes between two strengths which together characterize each item stored in memory: storage strength which reflects how well item is learnt, and retrieval strength which determines present access to that item in memory. Retrieval strength solely determines the probability that the item will be recalled given a particular cue, thus the test performance is fully determined by it, and storage strength has no effect on current memory performance. As these two strengths are independent it may be that some items have low retrieval and high storage strength (e.g. memories from childhood) and others may have high retrieval and low storage strength (a location of a meeting later that day).

The way both strengths interact can be counterintuitive – in general storage strength has an infinite capacity for increase, however the higher the current retrieval strength of given item, the lower increase in its storage strength. As for retrieval strength – the higher it currently is, the lower its increase upon additional learning. What follows in terms of guidelines for effective learning is that benefits of successful retrieval would be most profound when it was achieved in challenging conditions because those would be characterized by a low retrieval strength. This chimes with the idea common to of all desirable difficulties which aim to make the learning process more challenging rather than fluent and easy, as the latter will be associated with high retrieval strength, which makes the learning process redundant.

A good example which can illustrate how to design possibly effective desirable difficulties is combining retrieval practice with the spacing effect. Bjork and Landauer, (1978) showed that increasing the lags between study sessions (the so-called 'expanding retrieval practice') serves to keep retrievals successful and at the same time increases their learning potential. Referring this to the New Theory of Disuse – longer lags lowered the current retrieval strength of cued information which increased learning upon retrieval success.

As already mentioned, guessing can be seen as a special instance of testing in which almost all responses given in the learning phase are incorrect because participants are not offered a pre-study session and the answers are usually very hard to correctly guess by chance. What is more, those few correct responses obtained during learning phase are usually removed from the analysis as its scope refers to *errorful* learning. However, from the New Theory of Disuse perspective, those very excluded items should generate most learning – as they entertain a successful retrieval at low retrieval strength. Although a scenario in which learning materials are designed to promote correct guessing lies beyond the focus of the current project which refers to errorful learning – it may show some positive results.

For instance, the second project described here could include another experiment in which the cueing sentences were both predictive and also not entirely misleading with regards to the meaning of the Finnish word: e.g., *He was scratching his arm because he was bitten by a* MEHILÄINEN (fin. bee). In this instance participants would be likely to correctly guess the semantic category of the Finnish word (in this case – insects that can potentially bite) even if the generated exemplar would not match the target (for instance, they would guess a word *mosquito* instead of the correct response – *bee*). It is possible that such partially correct guessing would finally transform to the associative memory and help to bind the Finnish word with its correctly generated Polish counterpart. This, however, remains to be tested because, as we have already stated, guessing tends to aid associative memory performance only when the same cues are used for guessing in learning session and final retrieval.

Although we clearly failed in designing another difficult yet desired approach to learning, we would like to expand on these null findings and analyze them in terms of educationally relevant directions for self-study which may have some real utility for learners. The first and most direct advice would be to discourage learners from guessing at novel translations and reading the correct meaning instead. We can reason that as four of our experiments (Experiment 1.1, 1.2, 1.3, and 2.1) consistently showed inferior (or at best – comparable) associative performance for guessing at foreign translations compared to reading them outright. Also, the combined Bayesian analysis of Experiments 1.1-1.3 provided extreme evidence for a difference between reading and guessing. In other words, guessing at a translation of a foreign word can actually impair memory compared to reading the foreign word-translation pair outright. What is more, in Experiment 3.3, we demonstrated that guessing was also disadvantageous when learning unfamiliar responses to trivia, which can be compared to novel foreign vocabulary. Thus, we can safely conclude, that guessing is not advisable, particularly when studying unfamiliar, new material, such as, e.g., foreign vocabulary. Interestingly, this same general conclusion was reached almost a hundred years ago by Forlano and Hoffman (1937), who found that presenting full Hebrew-English translations to primary schoolers was a better method of vocabulary teaching than asking the pupils to guess the meaning first. Our results generalize this pattern of results to adults, despite us doing our best to reverse it.

Given our futile attempts to demonstrate the memory benefits of guessing (for materials other than related word pairs), it is worth reminding here that our efforts were motivated by a plethora of recent research which portrays guessing as an effective, easily implementable study technique (e.g., Grimaldi & Karpicke, 2012; Knight et al., 2012; Kornell et al., 2009, Potts & Shanks, 2014; Richland et al., 2009; Wong & Lim, 2022). What is more, claims of the benefits of incorrect guessing continue to be made. For instance Pan and Rivers (2023) wrote that "Taking a practice test on information that has yet to be learned (…) can improve memory substantially relative to nontesting methods (e.g., reading), provided that the correct answers are studied afterwards" (p. 1461). Also, Kliegl et al. (2023) very recently suggested that incorrect guessing (at related word pairs) benefits final recall over reading to an even greater extent when participants are presented with interfering material before the final test.

Thus, our findings should be interpreted as going somewhat far in tempering the optimism with which guessing is often treated in the literature on learning strategies. At the same time we acknowledge that there are other learning techniques with a higher potential for improved performance when studying foreign vocabulary such as, e.g., keyword method (Atkinson, 1975) in which the foreign word is connected to its translation by an acoustic link and a mental image of an interaction between the studied word and its meaning. Also, we underline that foreign vocabulary acquisition has been shown to benefit from the testing effect (e.g. Toppino & Cohen, 2009). Nevertheless, guessing should not be considered as successful when studying foreign words, especially when learners possess no previous background in the foreign language.

A second direction comes from the first series of our experiments which was designed to investigate whether introducing a restudy phase will reverse the unfavorable pattern of results and deliver guessing advantage over reading. However, it consistently showed that it is the initial encounter with the to-be-learned material that determines the overall effectiveness of learning and it is not significantly changed by adding one restudy session. Naturally, the restudy opportunity further aids learning and does so to a large extent, but – as shown across Experiments 1.1-1.4 – does not modify the patterns obtained during the initial learning phase. This is not to say that relearning is never effective in modifying patterns of initial learning. For example, Rawson and Dunlosky (2014) documented how relearning negates the benefits initially accruing from spaced learning. However, these authors used initial learning to criterion and multiple relearning sessions. It is likely that limited relearning in a single study session has markedly less pronounced effects, as shown here. This observation is of high practical importance given the limited time and effort people often spend during the learning process. The fact that the patterns of performance resulting from this first study opportunity are not modifiable by additional study in a single session speaks to the importance of choosing an appropriate strategy for the initial learning. Given that the first encounter with study materials is likely to happen in organized educational settings such as classrooms, our findings underscore the role of appropriate strategy choice on the part of educators.

Another educationally relevant direction follows from the last series of our experiments and states that asking questions before the study material is presented can benefit performance on the test, this benefit however will be restricted only to the exact same questions that were asked. This conclusion agrees with the research on the so-called pretesting effect (Carpenter et al., 2018) where a series of questions is asked before the presentation of study material. These studies have generally found that pretesting facilitates subsequent learning of answers to these questions but does not improve learning of other information contained in the same lecture (James & Storm, 2019, but see Sana & Carpenter, 2023). At the same time it should be noticed that these study benefits can be easily matched by extending time for a simpler approach to learning, like, e.g., reading. Applying this to the lecture setting, we can equally recommend asking pre-questions before the lecture or devoting longer time to cover the relevant topics during its course.

Also, it should be noted that when studying more familiar materials like trivia and not like foreign translations, guessing was no *worse* than reading – in fact we obtained a Bayesian evidence against any difference between the two learning conditions in our last project. This null finding alone clearly is insufficient to promote guessing as an effective learning strategy but, at the same time – does not necessarily discourages from using it. The fact that in our third project the guess condition presented the correct response for eight seconds shorter compared to the long read condition and showed similar performance can be of some significance. At least it can eliminate the concern of teachers and other educators who are reluctant to test their students in appreciation of the fact that incorrect responses may interfere with correct responses. According to these results – such concerns are unwarranted.

Though the vast majority of our findings was negative or null, we did obtain some positive effects of guessing, which may prove useful also from an applied point of view. Firstly, we replicated the guessing advantage for item memory when learning foreign translations in Experiments 1.1 and 2.3. Still, as we have already mentioned, this type of memory has a very restricted utility when studying foreign vocabulary as it translates to better remembering which native words were studied. Moving on to the associative benefit of guessing, we obtained some positive results in Experiment 1.4, when studying weakly-related word pairs, and in Experiment 2.2, in which people were tested from the associations between cueing sentences and surprising endings. As we believe that studying word associates has little relevance in more applied settings, we will focus on the positive results of Experiment 2.2. As already mentioned, the advantage of guessing over reading documented here may be of particular relevance especially minding the findings of the last series of our experiments which failed to produce any performance increase of guessing over reading after equating the timing of the learning trials. In Experiment 2.2, contrarily, we documented the guessing advantage over reading while both learning trials lasted 13 seconds. It seems, thus, that for guessing to be effective, some semantic relationship between the cue and the target should be present and the same cues should be used at test as were during learning. In other words, guessing could be successfully used in educational settings as long as the learning outcome is also consistent with those requirements, i.e., learning some semantically rich information which is identically cued while learning and while testing.

All in all, the current thesis succeeded in consistently showing why it is better not to guess when learning unfamiliar novel materials. This consistency was disrupted by some positive findings which, however, are of little educational or practical utility and did not transform to more applicable benefits such as enhanced associations between Finnish and Polish translations or better learnt trivia responses. To summarize, we rather discourage from using guessing as an effective learning strategy as it seems to be so within a narrow spectrum of materials which have little relevance to those students usually want to learn.

References

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In *Memory* (pp. 237-313). Academic Press.

Arnold, K. M., & McDermott, K. B. (2013a). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 940-945.

Arnold, K. M., & McDermott, K. B. (2013b). Free recall enhances subsequent learning. *Psychonomic Bulletin & Review*, *20*, 507-513.

Atkinson, R. C. (1975). Mnemotechnics in second-language learning. *American Psychologist, 30*(8), 821-828.

Atkinson, R. C. (1972). Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology, 96*(1), 124-129.

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective?. *Cognitive Psychology*, *61*(3), 228-247.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society, 2*(59-68).

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. *From learning processes to cognitive processes: Essays in honor of William K. Estes, 2,* 35-67.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417-444.

Bjork, R. A., & Landauer, T. K. (1978). On keeping track of the present status of people and things. *Practical Aspects of Memory*, 52-60.

Brainerd, C. J., Liu, X., Bialer, D. M., & Chang, M. (2023). The big three: Accuracy, organization, and retrieval effects of latent semantic attributes. *Journal of Experimental Psychology: General*, *152*(6), 1768-1786.

Bridger, E. K., & Mecklinger, A. (2014). Errorful and errorless learning: The impact of cue–target constraint in learning from errors. *Memory & Cognition, 42*(6), 898-911.

Brod, G. (2021). Predicting as a learning strategy. *Psychonomic Bulletin & Review, 28*, 1839-1847.

Brod, G., & Breitwieser, J. (2019). Lighting the wick in the candle of learning: generating a prediction stimulates curiosity. *NPJ Science of Learning*, *4*(1), 1-7.

Brod, G., Greve, A., Jolles, D., Theobald, M., & Galeano-Keiner, E. M. (2022). Explicitly predicting outcomes enhances learning of expectancy-violating information. *Psychonomic Bulletin & Review*, *29*(6), 2192-2201.

Brod, G., Hasselhorn, M., & Bunge, S. A. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction*, *55*, 22-31.

Burt, J. S., Leggett, J. M., Chalmers, K. A., & Boulton, P. A. (2021). Retrieval practice via corrective feedback: is learning better for targets in an expected or surprising sense? *Memory, 29*(10), 1396-1410.

Butowska, E., Hanczakowski, M., & Zawadzka, K. (2022). You won't guess that: On the limited benefits of guessing when learning a foreign language. *Memory & Cognition, 50*, 1033-1047.

Butterfield, B., & Metcalfe, J. (2001). Errors committed with high confidence are hypercorrected. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1491-1494.

Cantor, A. D., Eslick, A. N., Marsh, E. J., Bjork, R. A., & Bjork, E. L. (2015). Multiple-choice tests stabilize access to marginal knowledge. *Memory & Cognition*, *43*, 193-205.

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.

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.

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review, 24*, 369-378.

Carpenter, S. K., Rahman, S., & Perkins, K. (2018). The effects of prequestions on classroom learning. *Journal of Experimental Psychology: Applied*, *24*(1), 34-42.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology*, *56*(4), 236-246.

Chan, J. C., Manley, K. D., Davis, S. D., & Szpunar, K. K. (2018). Testing potentiates new learning across a retention interval and a lag: A strategy change perspective. *Journal of Memory and Language*, *102*, 83-96.

Cyr, A. A., & Anderson, N. D. (2018). Learning from your mistakes: Does it matter if you're out in left foot, I mean field?. *Memory, 26*(9), 1281-1290.

Davis, S. D., Chan, J. C., & Wilford, M. M. (2017). The dark side of interpolated testing: Frequent switching between retrieval and encoding impairs new learning. *Journal of Applied Research in Memory and Cognition*, *6*(4), 434-441.

Di Vesta, F. J., & Peverly, S. T. (1984). The effects of encoding variability, processing activity, and rule–examples sequence on the transfer of conceptual rules. *Journal of Educational Psychology, 76*(1), 108-119.

Fazio, L. K., & Marsh, E. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin & Review, 16*(1), 88-92.

Fillingham, J. K., Hodgson, C., Sage, K., & Lambon Ralph, M. A. (2003). The application of errorless learning to aphasic disorders: A review of theory and practice. *Neuropsychological Rehabilitation*, *13*(3), 337-363.

Finn, B., & Roediger III, H. L. (2013). Interfering effects of retrieval in learning new information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1665-1681.

Finn, B., & Tauber, S. K. (2015). When confidence is not a signal of knowing: How students’ experiences and beliefs about processing fluency can lead to miscalibrated confidence. *Educational Psychology Review, 27*, 567-586.

Flavell, J. H., & Wellman, H. M. (1975). Metamemory.

Forlano, G., & Hoffman, M. N. (1937). Guessing and telling methods in learning words of a foreign language. *Journal of Educational Psychology, 28*(8), 632-636.

Gagnepain, P., Lebreton, K., Desgranges, B., & Eustache, F. (2008). Perceptual priming enhances the creation of new episodic memories. *Consciousness and Cognition, 17*(1), 276-287.

Gambi, C., Pickering, M. J., & Rabagliati, H. (2021). Prediction error boosts retention of novel words in adults but not children. *Cognition, 211*, 104650.

Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language, 94,* 149-165.

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding?. *Memory & Cognition, 40*(4), 505-513.

Gruber, M. J., & Ranganath, C. (2019). How curiosity enhances hippocampus-dependent memory: The prediction, appraisal, curiosity, and exploration (PACE) framework. *Trends in Cognitive Sciences, 23*(12), 1014-1025.

Haeuser, K. I., & Kray, J. (2021). Effects of prediction error on episodic memory retrieval: evidence from sentence reading and word recognition. *Language, Cognition and Neuroscience,* 1-17.

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(1), 290-296.

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus, 20*, 1315-1326.

Hollins, T. J., Seabrooke, T., Inkster, A., Wills, A., & Mitchell, C. J. (2023). Pre-testing effects are target-specific and are not driven by a generalised state of curiosity. *Memory*, *31*(2), 282-296.

Huelser, B. J., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition, 40*(4), 514-527.

Huppert, F. A., & Kopelman, M. D. (1989). Rates of forgetting in normal ageing: A comparison with dementia. *Neuropsychologia*, *27*(6), 849-860.

Izawa, C. (1971). The test trial potentiating model. *Journal of Mathematical Psychology*, *8*(2), 200-224.

Jacoby, L.L., & Wahlheim, C.N. (2013). On the importance of looking back: The role of recursive remindings in recency judgments and cued recall. *Memory & Cognition 41,* 625-637.

James, K. K., & Storm, B. C. (2019). Beyond the pretesting effect: What happens to the information that is not pretested?. *Journal of Experimental Psychology: Applied*, *25*(4), 576-587.

JASP Team (2023). JASP (Version 0.17.2) [Computer software].

Jones, R. S., & Eayrs, C. B. (1992). The use of errorless learning procedures in teaching people with a learning disability: A critical review. *Mental Handicap Research, 5*(2), 204-212.

Kafkas, A., & Montaldi, D. (2018). Expectation affects learning and modulates memory experience at retrieval. *Cognition, 180*, 123-134.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In *Psychology of Learning and Motivation* (Vol. 61, pp. 237-284). Academic Press.

Karpicke, J. D., & Roediger III, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*(2), 151-162.

Karpicke, J. D., & Roediger III, H. L. (2008). The critical importance of retrieval for learning. *Science, 319,* 966-968.

Kerr, R., & Booth, B. (1978). Specific and varied practice of motor skill. *Perceptual and Motor Skills*, *46*(2), 395-401.

Kim, K., Yi, D. J., Raye, C. L., & Johnson, M. K. (2012). Negative effects of item repetition on source memory. *Memory & Cognition, 40*(6), 889-901.

Kliegl, O., Bartl, J., & Bäuml, K. H. T. (2023). The pretesting effect thrives in the presence of competing information. *Memory*, *31*(5), 705-714.

Knight, J. B., Ball, B. H., Brewer, G. A., DeWitt, M. R., & Marsh, R. L. (2012). Testing unsuccessfully: A specification of the underlying mechanisms supporting its influence on retention. *Journal of Memory and Language, 66*(4), 731-746.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349-370.

Kornell, N. (2014). Attempting to answer a meaningful question enhances subsequent learning even when feedback is delayed. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(1), 106-114.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”?. *Psychological Science*, *19*(6), 585-592.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 989-998.

Kornell, N., Klein, P. J., & Rawson, K. A. (2015). Retrieval attempts enhance learning, but retrieval success (versus failure) does not matter. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(1), 283-294.

Lawson, M. J., & Hogben, D. (1998). Learning and recall of foreign-language vocabulary: Effects of a keyword strategy for immediate and delayed recall. *Learning and Instruction, 8*(2), 179-194.

Lee, H., Kim, K., & Yi, D.-J. (2020). Preexperimental stimulus familiarity modulates the effects of item repetition on source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(3), 539-548.

Leggett, J. M. I., & Burt, J. S. (2021). Errors may not cue recall of corrective feedback: Evidence against the mediation hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47,* 65-74.

Mayer, R. E., & Anderson, R. B. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, *84*(4), 444-452.

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition, 39*, 462-476.

McDaniel, M. A., & Donnelly, C. M. (1996). Learning with analogy and elaborative interrogation. *Journal of Educational Psychology*, *88*(3), 508-519.

Metcalfe, J., & Huelser, B. J. (2020). Learning from errors is attributable to episodic recollection rather than semantic mediation. *Neuropsychologia*, 138, 107296.

Middleton, E. L., & Schwartz, M. F. (2012). Errorless learning in cognitive rehabilitation: A critical review. *Neuropsychological Rehabilitation, 22*(2), 138-168.

Nelson, D.L., McEvoy, C.L. & Schreiber, T.A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers 36*, 402-407.

Overman, A. A., Stephens, J. D. W., & Bernhardt, M. F. (2021). Enhanced memory for context associated with corrective feedback: *Evidence for episodic processes in errorful learning. Memory,* 29, 1017-1042.

Pan, S. C., Lovelett, J., Stoeckenius, D., & Rickard, T. C. (2019). Conditions of highly specific learning through cued recall. *Psychonomic Bulletin & Review, 26*, 634-640.

Pan, S. C., & Rivers, M. L. (2023). Metacognitive awareness of the pretesting effect improves with self-regulation support. *Memory & Cognition*, 1-20.

Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological Review, 127*(1), 1-46.

Poppenk, J., & Norman, K. A. (2012). Mechanisms supporting superior source memory for familiar items: A multi-voxel pattern analysis study. *Neuropsychologia, 50*(13), 3015-3026.

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, *143*(2), 644-667.

Potts, R., Davies, G., & Shanks, D. R. (2019). The benefit of generating errors during learning: What is the locus of the effect?. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(6), 1023-1041.

Pressley, M., McDaniel, M. A., Turnure, J. E., Wood, E., & Ahmad, M. (1987). Generation and precision of elaboration: Effects on intentional and incidental learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(2), 291-300.

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330*(6002), 335-335.

Quent, J. A., Henson, R. N., & Greve, A. (2021). A predictive account of how novelty influences declarative memory. *Neurobiology of Learning and Memory, 179*, 107382.

Rawson, K.A. & Dunlosky, J. (2014). Bang for the buck: Supporting durable and efficient student learning through successive relearning. In M. McDaniel, R. Frey, S. Fitzpatrick, & H.L. Roediger (Eds), Integrating cognitive science with innovative teaching in STEM disciplines.

Rawson, K. A., Thomas, R. C., & Jacoby, L. L. (2015). The power of examples: Illustrative examples enhance conceptual learning of declarative concepts. *Educational Psychology Review*, *27*, 483-504.

Reder, L. M., Liu, X. L., Keinath, A., & Popov, V. (2016). Building knowledge requires bricks, not sand: The critical role of familiar constituents in learning. *Psychonomic Bulletin & Review, 23*(1), 271-277.

Reder, L. M., Victoria, L. W., Manelis, A., Oates, J. M., Dutcher, J. M., Bates, J. T., & Gyulai, F. (2013). Why it’s easier to remember seeing a face we already know than one we don’t: Preexisting memory representations facilitate memory formation. *Psychological Science*, *24*(3), 363-372.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F.

Prokasy (Eds.), Classical conditioning, 2: Current research and theory (pp. 64-69). Appleton-Century-Crofts.

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15(*3), 243-257.

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249-255.

Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. Ross (Eds.), Psychology of learning and motivation: Cognition in education, (pp. 1-36). Oxford: Elsevier

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*, 481-498.

Rommers, J., & Federmeier, K. D. (2018). Lingering expectations: A pseudo-repetition effect for words previously expected but not presented. *NeuroImage, 183*, 263-272.

Rumelhart, D. E., & McClelland, J. L. (1986). Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations. Cambridge, MA: MIT Press.

Sana, F., & Carpenter, S. K. (2023). Broader benefits of the pretesting effect: Placement matters. *Psychonomic Bulletin & Review*, 1-9.

Schomaker, J., & Meeter, M. (2015). Short-and long-lasting consequences of novelty, deviance and surprise on brain and cognition. *Neuroscience & Biobehavioral Reviews, 55*, 268-279.

Seabrooke, T., Hollins, T. J., Kent, C., Wills, A. J., & Mitchell, C. J. (2019). Learning from failure: Errorful generation improves memory for items, not associations. *Journal of Memory and Language, 104*, 70-82.

Seabrooke, T., Mitchell, C. J., Wills, A. J., & Hollins, T. J. (2021). Pretesting boosts recognition, but not cued recall, of targets from unrelated word pairs. *Psychonomic Bulletin & Review, 28*(1), 268-273.

Seabrooke, T., Mitchell, C. J., Wills, A. J., Inkster, A. B., & Hollins, T. J. (2022). The benefits of impossible tests: Assessing the role of error-correction in the pretesting effect. *Memory & Cognition, 50*, 296-311.

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental psychology: Human Learning and memory*, *5*(2), 179-187.

Slamecka, N. J., & Fevreiski, J. (1983). The generation effect when generation fails. *Journal of Verbal Learning and Verbal Behavior, 22(*2), 153-163.

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*(4), 342-353.

Theobald, M., & Brod, G. (2021). Tackling scientific misconceptions: The element of surprise. *Child Development*, *92*(5), 2128-2141.

Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, *56*(4), 252-257.

Tullis, J. G., Benjamin, A. S., & Ross, B. H. (2014). The reminding effect: presentation of associates enhances memory for related words in a list. *Journal of Experimental Psychology: General*, *143*(4), 1526-1540.

Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition, 41*, 429-442.

Vaughn, K. E., Dunlosky, J., & Rawson, K. A. (2016). Effects of successive relearning on recall: Does relearning override the effects of initial learning criterion?. *Memory & Cognition, 44*(6), 897-909.

Vaughn, K. E., & Rawson, K. A. (2012). When is guessing incorrectly better than studying for enhancing memory? *Psychonomic Bulletin & Review, 19*(5), 899-905.

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition, 41*(1), 1-15.

Warmington, M., Hitch, G. J., & Gathercole, S. E. (2013). Improving word learning in children using an errorless technique. *Journal of Experimental Child Psychology, 114*(3), 456-465.

Wong, B. Y. (1985). Self-questioning instructional research: A review. *Review of Educational Research*, *55*(2), 227-268.

Wong, S. S. H., & Lim, S. W. H. (2022). The derring effect: Deliberate errors enhance learning. *Journal of Experimental Psychology: General, 151,* 25-40.

Yan, V. X., Yu, Y., Garcia, M. A., & Bjork, R. A. (2014). Why does guessing incorrectly enhance, rather than impair, retention? *Memory & Cognition, 42*, 1373-1383.

Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: a review of the forward testing effect. *NPJ science of learning*, *3*(1), 1-9.

Zawadzka, K., Baloro, S., Wells, J., Wilding, E. L., & Hanczakowski, M. (2021). On the memory benefits of repeated study with variable tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(7), 1067-1082.

Zawadzka, K., & Hanczakowski, M. (2019). Two routes to memory benefits of guessing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(10), 1748-1760.

Zawadzka, K., Zaborowska, O., Butowska, E., Piątkowski, K., & Hanczakowski, M. (2023). Guessing can benefit memory for related word pairs even when feedback is delayed. *Memory & Cognition*, 1-14.

Appendices

Appendix 1.1 Translations used in Experiments 1.1 and 1.2.

Finnish word	Polish translation	English Translation
aika	czas	time
alku	początek	beginning
amme	wanna	bath
asema	stacja	station
aurinko	słońce	sun
avain	klucz	key
eilen	wczoraj	yesterday
elokuva	film	film
hame	spódnica	skirt
hissi	winda	lift
housut	spodnie	trousers
huone	pokój	room
hylly	półka	shelf
ikkuna	okno	window
ilma	pogoda	weather
jalka	stopa	foot
joki	rzeka	river
juna	pociąg	train
kaappi	postać	person
kassi	torba	bag
kaula	szyja	neck
keitto	zupa	soup
kerros	podłoga	floor
kirja	książka	book
laiva	statek	ship
laki	prawo	law
lapsi	dziecko	child
laulaja	śpiewak	singer
lintu	kino	cinema
lopakko	portfel	wallet
maito	mleko	milk
makkara	kiełbasa	sausage
naapuri	sąsiad	neighbor
nappi	guzik	button
nopeus	prędkość	speed
olut	piwo	beer
omena	jabłko	apple
ongelma	problem	problem
opinnot	studia	studies
osoite	adres	address
parturi	fryzjer	hairdresser
peili	lustro	mirror

peruna	ziemniak	potato
piste	kropka	dot
puhelin	telefon	phone
rakennus	budynek	building
ratikka	tramwaj	tram
rinta	pierś	breast
roskat	śmieci	rubbish
sana	słowo	word
sisko	siostra	sister
sormi	palec	finger
sota	wojna	war
taide	sztuka	art
tavara	rzecz	thing
tukka	włosy	hair
tulosti	drukarka	printer
urheilu	sport	sport
uskonto	religia	religion
vaimo	żona	wife
valokuva	zdjęcie	photo
vasta	brzuch	belly
veli	brat	brother
virhe	błąd	error

Appendix 1.2 Translations used in Experiment 1.3

Finnish word	Polish translation	English translation
alku	początek	beginning
amme	wanna	bath
apu	pomoc	help
asema	stacja	station
ateria	posiłek	meal
aurinko	słońce	sun
elokuva	film	film
esitys	pokaz	show
etana	ślimak	snail
farkut	jeansy	jeans
flunssa	grypa	flu
hame	spódnica	skirt
hella	piekarnik	oven
huone	pokój	room
ikkuna	okno	window
ilma	pogoda	weather
jalka	stopa	foot
joki	rzeka	river
juna	pociąg	train
kassi	torba	bag
keitto	zupa	soup
kirje	list	letter
korva	ucho	ear
kuva	obrazek	picture
laiva	statek	ship
lentokone	samolot	airplane
lintu	ptak	bird
lopakko	portfel	wallet
luonto	natura	nature
makkara	kiełbasa	sausage
mehu	sok	juice
musta	czerń	black
nappi	guzik	button
niemi	przylądek	cape
nopeus	prędkość	speed
olut	piwo	beer
omena	jabłko	apple
ongelma	problem	problem
opinnot	studia	studies
orava	wiewiórka	squirrel
osoite	adres	address
otsa	czoło	forehead
parturi	fryzjer	hairdresser
peili	lustro	mirror

piste	kropka	dot
pokia	chłopiec	boy
puhelin	telefon	phone
ratikka	tramwaj	tramway
retki	podróż	travel
rinta	pierś	chest
roskat	śmieci	rubbish
ruskea	brąz	bronze
saari	wyspa	island
seura	firma	company
sormi	palec	finger
suklaa	czekolada	chocolate
syy	powód	cause
tavara	rzecz	thing
tentti	egzamin	exam
tieto	dane	data
toimisto	biuro	office
tukka	włosy	hair
urheilu	sport	sport
uskonto	religia	religion

Appendix 1.3 Word pairs used in Experiment 1.4

Cue	Target	Cue – English	Target – English
naklejka	znak	sticker	sign
kawa	ziarno	coffee	grain
randka	chłopak	date	boy
prezent	pudełko	gift	box
szyja	obroża	neck	collar
kanapka	lunch	sandwich	lunch
morze	krab	sea	crab
brodawka	pieprzyk	wart	mole
ciężar	kotwica	weight	anchor
wypadek	wrak	accident	wreck
młody	niewinny	young	innocent
sklep	ubrania	store	clothes
gra	wideo	game	video
portfel	kieszeń	wallet	pocket
rzeka	tama	river	dam
zapach	perfumy	smell	perfume
absurd	dziwny	absurd	weird
chleb	bochenek	bread	loaf
sałatka	oliwka	salad	olive
wojsko	mundur	army	uniform
ciemny	strych	dark	attic
półka	klif	shelf	cliff
gumka	kreda	rubber	chalk
instalacja	fabryka	installation	factory
woda	naczynie	water	dish
bęben	zespół	drum	team
laser	promień	laser	ray
ściek	zlew	sewage	sink
park	piknik	park	picnic
taniec	impreza	dance	party
aktor	obsada	actor	cast
klon	roślina	maple	plant
wakacje	wycieczka	holidays	trip
wino	kolacja	wine	dinner
urzędnik	biuro	clerk	office
karty	talia	cards	deck

szkoła	semestr	school	term
droga	most	way	bridge
pszenica	mąka	wheat	flour
dolina	wzgórze	valley	hill
test	stres	test	stress
Afryka	plemię	Africa	tribe
sekret	kłamstwo	secret	lie
liść	zioło	leaf	herb
ptak	śpiew	bird	singing
sok	limonka	juice	lime
głowa	guz	head	bump
telefon	dzwonek	phone	bell
herbata	cukier	tea	sugar
rączka	miotła	handle	broom
królowa	pałac	queen	palace
opowieść	mit	story	myth
energia	aktywny	energy	active
jogurt	mleko	yogurt	milk
uprawy	farma	crops	farm
wschód	orient	east	orient
but	noga	shoe	leg
gęś	łabędź	goose	swan
okno	widok	window	view
klocki	zabawki	bricks	toys
uraza	zdrada	resentment	betrayal
strażnik	tarcza	guard	shield
trawnik	ogród	lawn	garden
włosy	cięcie	hair	cut

Appendix 2.1 Materials used in Experiments 2.1-2.3

High-constraint Polish	High-constraint English	English meaning	Finnish meaning	Polish meaning
Do kuchni chodził	He would go to the kitchen to	sleep	nukkua	spać
Piłkarz kopnął	A footballer kicked a	stick	keppi	patyk
Panna młoda miała na sobie piękną	The bride was wearing a beautiful	cape	viitta	peleryna
Pocałował ją w	He kissed her on her	knee	polvi	kolano
Do herbaty dodawał plaster	He put in his tea a slice of	strawberry	mansikka	truskawka
Miał dwie siostry i jednego	He had two sisters and one	cousin	serkku	kuzyn
Na karmniku siedziała	On a birdhouse there sat a	bee	mehilainen	pszczoła
W piwnicy trzymała wiele słoików pełnych	In the cellar she kept many jars full of	honey	humaja	miód
Lis polował na	A fox was hunting	sparrows	varpuset	wróble
Powiesiła obrazek na	She put a picture on the	fireplace	takka	kominek
Lubiła świeże	She liked fresh	meat	liha	mięso
W stawie pływało dużo łabędzi i	In the pond there were many swans and	fish	kalastaa	ryby
Uwielbiała zwiedzać	She liked visiting the	castle	linna	zamek
Na festiwalu filmowym pojawił się znany	At the film festival there arrived a famous	scientist	tutkija	naukowiec
Chodziła na basen uczyć się	She went to a swimming pool to learn how to	read	lukea	czytać
Uwielbiał biegać słuchając	He loved to run listening to the	noise	melua	szum
Smok zabił dzielnego	A dragon killed a brave	shoemaker	suutari	szewc
Kot cierpliwie czekał, żeby upolować	A cat was waiting patiently to catch a	rabbit	pupu	królik
Uczniowie grali na boisku przed	Students were playing on the pitch in front of the	town hall	raastupa	ratusz
Po przepracowaniu całej nocy chciał iść do	After working the night shift he wanted to go to a	restaurant	ravintola	restauracja
Wykopał dół za pomocą	He dug a hole with his	legs	jalkat	nogi
W odstraszaniu wampirów pomóc może	Helpful when scaring vampires away is	joy	riemu	radość
Gorąca zupa zaczynała mieć dobry	The hot soup started to have a good	smell	haju	zapach
Na pastwisku pasły się	On the pasture one could see grazing	donkeys	aasit	osły
Na sawannie głośno zaryczał	On the savannah they heard a	car horn	sarvi	klakson
Po ciężkim dniu zapalił	After a long day he lit a	light	valo	światło
Policjant łapał	A policeman was catching	butterflies	perhoset	motyle
Często jadła kanapki z serem i	She often ate sandwiches with cheese and	jam	hillo	dżem

Na wiosnę zakwitło wiele	In spring there were many blooming	trees	puut	drzewa
Magik pokazał im nową	A magician showed them a new	wand	sauva	różdżka
Opowiedziała im bajkę o starej, strasznej	She told them a story about an old wicked	sheep	lammas	owca
Na środku oceanu dryfowała pusta	In the middle of an ocean there was a drifting	bath	kylpy	wanna
Drapał się całą noc, tam gdzie ugryzł go	All night, he scratched himself where he was bitten by a	tortoise	kilpikonna	żółw
Popełnił wiele	He wrote many	articles	kappale	artykuł
Usiadła na wygodnym	She sat on a comfortable	tree trunk	runko	pień
Po wypadku samochodowym był cały w	After the car crash he was covered in	mud	muta	błoto
Pociąg powoli wjechał na	A train slowly ran over a	bridge	silta	most
W oddali zobaczyli jeźdźca na	In the distance they saw a rider on a/an	elephant	norsu	słoń
Pan umył podłogę	The man washed the floor with	soap	saippua	mydło
W budynku został tylko sprzątający	In the building there was only a cleaning	boss	esimies	szef
Kadra narodowa musiała zmienić	The national team had to change their	attitude	asenne	nastawienie
Kibice zamówili trzy kufle	The fans ordered three pints of	water	vesi	woda
Wiele osób lubi zbierać grzyby w	Many people like to pick mushrooms in	October	lokakuu	październik
Wieczorem, poszli potańczyć na	In the evening, they went dancing in the	street	katu	ulica
Nauczyciel pisał na	The teacher was writing on a	bench	penkki	ławka
Cały dzień kosił	All day, he was mowing	cereal	vilja	zboże
Beduin jechał na swoim	A Bedouin rode his	scooter	potkulauta	skuter
Na parkingu stał	In the car park there was a	truck	rekka	tir
Wybrał się w podróż dookoła	He took a trip around the	apartment	kerrostalo	blok
Na boisko wyszły dwie	Onto the pitch there came out two	women	naiset	kobiety
Dziewczynka skakała na	The girl jumped on the	pavement	ajotie	chodnik
Bokser często chodził na	A boxer would often go to the	market	basaari	targ
Zeszłej nocy miała piękny	Last night she had a lovely	idea	ajatus	pomysł
Co by się nie działo, zawsze miał dobry	Whatever happened, he always had a good	result	tulos	wynik
To był jej najskrytszy	It was her most private	success	menestys	sukces
W takim upale każdemu chciało się	In such heat everyone wanted to	hoover	imuroida	odkurzać
Na ławce siedzieli dziadek z	On the bench sat a grandpa with a/an	uncle	eno	wujek
Małpka jadła	The monkey was eating a	carrot	porkkana	marchewka
Pirat miał gadającą	The pirate had a talking	crew	joukkue	załoga
Zagrał wieczorem piękny	In the evening he played a beautiful	match	ottelu	mecz
Zamówili duże burgery, colę i	They ordered big burgers, coke and a/an	casserole	laatikko	zapiekanka

Grupa harcerzy pojechała na	A group of scouts went	shopping	ostokset	zakupy
Zakochała się i oddała mu swoje	She fell in love and gave him her	presents	lahjat	prezenty
Siedzieli przy ognisku i piekli	They sat around the campfire and roasted	vegetables	vihannekset	warzywa

Low-constraint Polish	**Low-constraint English**	**English meaning**	**Finnish meaning**	**Polish meaning**
Nigdy nie lubił dużo	He never liked to	sleep	nukkua	spać
W parku znalazł	In the park he found a	stick	keppi	patyk
Zapomniał spakować do plecaka	He forgot to put in his backpack a	cape	viitta	peleryna
Jedną z części ciała jest	One of body parts is the	knee	polvi	kolano
Zerwał dojrzałą	He picked a ripe	strawberry	mansikka	truskawka
Na mieście spotkał swojego	Downtown he met his	cousin	serkku	kuzyn
Spójrz, tam lata	Look, there flies a	bee	mehilainen	pszczoła
Zwykle po południu jadła chleb z	Usually in the afternoon she ate bread with	honey	humaja	miód
W lesie mieszkają	In the forest there live	sparrows	varpuset	wróble
Co tydzień czyści	Every week he cleans the	fireplace	takka	kominek
Poszła do miasta kupić	She went to town to buy	meat	liha	mięso
W lodówce miała kilka kawałków	In the fridge she had a few pieces of	fish	kalastaa	ryby
To był bardzo ładny	It was a very nice	castle	linna	zamek
Zawsze chciała być	She has always wanted to be a	scientist	tutkija	naukowiec
Zawsze lubiła	She has always liked to	read	lukea	czytać
Gdy czytał przeszkadzał mu	When he was reading he was disturbed by a	noise	melua	szum
Ona jest żoną	She is a wife of a	shoemaker	suutari	szewc
Dziewczynka bawiła się z	The girl played with a	rabbit	pupu	królik
Grupa zwiedzała	The group visited the	town hall	raastupa	ratusz
Dziś mieli ochotę na wycieczkę do	Today they felt like taking a trip to a	restaurant	ravintola	restauracja
Trzeba mieć silne	You have to have strong	legs	jalkat	nogi
Miał w sobie mnóstwo	He had plenty of	joy	riemu	radość
Uważała, że miał dobry	She thought he had a good	smell	haju	zapach
Ze wszystkich zwierząt najbardziej lubił	Of all the animals his favorite were	donkeys	aasit	osły
W samochodzie znajduje się	In the car there is a	car horn	sarvi	klakson
W oddali pojawiło się	In the distance there was a	light	valo	światło
Lubiła książki o	She liked books about	butterflies	perhoset	motyle

Musiał się wrócić do sklepu po	He had to go back to the shop to buy	jam	hillo	dżem
Ciekawiły go różne gatunki	He was interested in different species of	trees	puut	drzewa
Marzyła o tym, żeby mieć	She dreamt of having a	wand	sauva	różdżka
Na wsi można spotkać	In the countryside you could see	sheep	lammas	owca
Jej ulubionym miejscem była	Her favorite place was a	bath	kylpy	wanna
To był bardzo ładny	It was a very pretty	tortoise	kilpikonna	żółw
Byli zadowoleni ze swojego	They were happy with their	articles	kappale	artykuł
W lesie można zobaczyć	In the forest you can see a	tree trunk	runko	pień
Chłopiec bawił się w	The boy was playing in the	mud	muta	błoto
W mieście znajdował się znany	In the city there was a famous	bridge	silta	most
Na wybiegu stał	In the enclosure there was an	elephant	norsu	słoń
W drogerii można było kupić	In the drugstore you could buy	soap	saippua	mydło
Na konferencji obecny był	At the conference there was a	boss	esimies	szef
Na wyprawie ważne jest	On a trip it is important to have a good	attitude	asenne	nastawienie
W sklepie można był kupić kilka rodzajów	In the shop you could buy several types of	water	vesi	woda
Jej ulubiony okres w roku to	Her favorite time of year was	October	lokakuu	październik
To była ładna	It was a pretty	street	katu	ulica
Po prawej stronie stała	On the right there was a	bench	penkki	ławka
Istnieje wiele odmian	There are many varieties of	cereal	vilja	zboże
Na urodziny dostała	For her birthday she got a	scooter	potkulauta	skuter
Obok samochodu stał	Next to the car there was a	truck	rekka	tir
Pomyślała, że to fajny	She thought it was a cool	apartment	kerrostalo	blok
Nie potrafili zrozumieć	They couldn't understand	women	naiset	kobiety
Znaleźli to czego szukali na środku	They found what they were looking for in the middle of the	pavement	ajotie	chodnik
W tej miejscowości jest	In this town there is a	market	basaari	targ
Podobał się im ten	They liked this	idea	ajatus	pomysł
Zależało mu na	He cared about the	result	tulos	wynik
Długo czekała na	She waited a long time for a	success	menestys	sukces
Nie każdy lubi	Not everyone likes to	hoover	imuroida	odkurzać
Często widywał się za swoim	He often saw his	uncle	eno	wujek
Kroiła w kuchni	In the kitchen she was cutting a	carrot	porkkana	marchewka
Bardzo szanował swoją	He respected very much his	crew	joukkue	załoga
Bardzo jej zależało na tym	She cared a lot about the	match	ottelu	mecz

Dzisiaj miał ochotę na	Today he fancied a	casserole	laatikko	zapiekanka
Zapomniała o	She forgot about	shopping	ostokset	zakupy
Każdy lubi	Everyone likes	presents	lahjat	prezenty
W kuchni było dużo	In the kitchen there was a lot of	vegetables	vihannekset	warzywa

Appendix 3.1 Questions used in Experiment 3.3

Question	Familiar Resposne	Question	Unfamiliar Response
What is the nickname of Melanie Chisholm from Spice Girls?	Sporty	What was the name of the world's first commercially produced birth-control pill?	Enovid
What company was the first to offer commercially available laptops?	Toshiba	In Star Wars, what planet was Yoda's home in his final years?	Dagobah
What was the third name of Queen Elisabeth II?	Mary	What's the name of the mythological poem core to Finnish language and identity?	Kalevala
What bird's eye is bigger than its brain?	Ostrich	What is the name of the famous art gallery in Florence?	Uffizi
Which country's name means 'abundance of fish'?	Panama	What is the name of the sacred Aboriginal rock in Australia?	Uluru
What is the most common mammal in the United States?	Mouse	What is the oldest Egyptian city	Shedet
What is the name of the constellation that looks like a flying horse?	Pegasus	What is the longest river in Spain called?	Ebro
What is the name of the country in which the game of dominoes was invented?	China	Which horse breed is the smallest?	Falabella
Which country has the most national parks?	Australia	What's the name of the music genre from Colombia which translates 'born in the valley'?	Vallenato
What is the only planet in the Solar System to rotate clockwise?	Venus	What is the name of a yellow herbal liqueur from Benevento, Italy?	Strega
Which city has the most Rolls-Royce per capita?	Hong Kong	Which volcano caused the year without a summer in 1816?	Tambora
Which tree species is the most common in Poland?	Pine	What is the name of the Norse god of archery, skiing and hunting?	Ullr
What part of the world was referred to as West Indies?	The Caribbean	Which body part attaches intestines to the abdominal wall?	Mesentery
What was the first song to be sung in outer space?	Happy Birthday	What was the name of the famous Napoleon's horse?	Marengo
What was the first trademarked product?	Beer	Where does Conan the Barbarian come from?	Cimmeria
In which European city can you see Charlottenburg?	Berlin	What is the capital of Burkina Faso?	Wagadugu
In which city was Pepsi-Cola first bottled?	Baltimore	How is the study of growing of fruit called?	Pomology
Which ancient philosopher was the author of The Republic?	Plato	What is the middle name of Uma Thurman?	Karuna
Which Disney movie features a cat named Lucifer?	Cinderella	In Christianity, what is another term for the Second Coming of Jesus Christ?	Parousia
Which element has the symbol Sn?	Tin	How do you call a raised curved part at the back of a saddle?	Cantle
Which flower is the national symbol of Italy?	Lily	What philosophical movement was founded by the Chinese philosopher Mo Di?	Mohism
Which team sport causes the most injuries?	Basketball	In the Lord of the Rings what was the name of Arwen's father?	Elrond
Which U.S. state is known as the "Beehive State"?	Utah	What is the name of Saturn's smallest moon?	Mimas
Who was the shortest US president?	Madison	Westerplatte, attacked by the Germans on the first day of World War II, is in which city?	Gdansk

Appendix 3.2 Questions used in Experiments 3.4 and 3.5

Pre-Question	Related Target	Test Question	Target
What is the name of the country in which the game of dominoes was invented?	China	What is the name of the country in which chess was invented?	India
What kind of poison did Socrates take at his execution?	hemlock	Which poison killed Russian ex-spy Litvinenko?	polonium
What was the first capital of ancient Egypt?	Memphis	What is the oldest Egyptian city	Shedet
What is a group of owls called?	parliament	What is a group of crows called?	murder
What is the name of the constellation that looks like a flying horse?	Pegasus	What is the name of the M-shaped star constellation?	Cassiopeia
What is the name of the brightest star in the sky excluding the Sun?	Sirius	Which star is closest to the Earth after the Sun?	Proxima centauri
What was the first state to allow women to vote?	Wyoming	Which was the last US state to allow women to vote?	Mississippi
What company was the first to offer a mouse on a commercially available computer?	Xerox	What company was the first to sell laptops?	Toshiba
What is the only planet in the Solar System to rotate clockwise?	Venus	Which planet spins the fastest?	Jupiter
Who was the shortest US president?	Madison	Who was the tallest US president?	Lincoln
Who won the ladies' final at Roland Garros in 2022?	Swiatek	Who won the ladies' Wimbledon final in 2022?	Rybakina
In Greek mythology, who was the god of blacksmiths, metals, and fire?	Hephaestus	In Greek mythology, who was the god of war?	Ares
Who discovered morphine?	Serturner	Who discovered penicillin?	Fleming
Which NBA player is the only one to have scored a 100 points in a single game?	Chamberlain	Which NBA player was married to Eva Longoria?	Parker
In Star Wars, what planet was Yoda's home in his final years?	Dagobah	In Star Wars, what planet was Padme's home?	Naboo
Who composed Erlkönig and Ave Maria?	Schubert	Who composed The Swan Lake ballet?	Tchaikovsky
Who was the first European explorer to reach India by sea?	de Gama	Which famous European explorer was killed on the Mactan island?	Magellan
Which horse breed is the smallest?	falabella	Which horse breed is the largest?	shire
What was the name of the longest serving Pope in history?	Pius	Who was the most recently canonised pope?	John Paul II
Who was the first woman to receive the Nobel Prize in Literature in 1909?	Lagerlof	Who was the first woman to receive a Nobel Prize in any field?	Skłodowska Curie
In which era did humans evolve?	cenozoic	At the end of which era did dinosaurs become extinct?	mesozoic
Which wartime German general was nicknamed 'The Desert fox'?	Rommel	Which wartime German official was the chief architect of the 'Final Solution'?	Himmler
Who won a Grammy for the best reggae album in 1995?	Shaggy	Which reggae artist's real name was Miguel Orlando Collins?	Sizzla
What is the name of the uppermost region of Earth's atmosphere?	exosphere	Which layer of the atmosphere is the hottest?	thermosphere
What is the name of a cold, strong wind that blows onto the Adriatic region of Croatia?	bora	What is the name of the warm, humid wind which blows from North Africa towards southern Europe?	sirocco

What kind of pistol was used to kill Archduke Franz Ferdinand?	Browning	What kind of pistol did James Bond traditionally use?	Walther
What is the SI unit of electric current?	ampere	What is the SI unit of temperature?	kelvin
Which NFL player was nicknamed 'Broadway'?	Namath	Which NFL player was nicknamed 'Hollywood'?	Henderson
Who was famous for creating a series of paintings of ballerinas?	Degas	Who was famous for creating a series of paintings of water lilies?	Monet
Which spell was used to light up the wand in Harry Potter?	lumos	Which spell was used to summon things in Harry Potter?	accio
Who won the first edition of America's got talent?	Ryan	Who won the 2021 edition of America's got talent?	Tavella
Which ancient philosopher proposed the floatation law?	Archimedes	Which ancient philosopher was the author of The Republic?	Plato
What is the name of the sacred Aboriginal rock in Australia?	Uluru	What is the name of the highest mountain of Japan, also its cultural symbol?	Fuji
What was the most bombed German city in WWII?	Dresden	What German city is famous for housing the trials of WWII criminals?	Nuremberg
Which Disney movie features a cat named Lucifer?	Cinderella	Which Disney movie features a cat named Figaro?	Pinocchio
Which Hollywood actress has won the most Oscars?	Hepburn	Which Hollywood actress married the Prince of Monaco?	Grace Kelly
Which UFC fighter created his own whisky brand?	McGregor	Which UFC fighter is famous for his 'Legendary Polish Power'?	Błachowicz
What was the name of the East Coast rapper killed in the '90s?	Biggy	What was the name of the West Coast rapper killed in the '90s?	Tupac
Which flower is the national symbol of Scotland?	thistle	Which flower is the national symbol of England?	rose
What is the largest tree-grown fruit?	jackfruit	Which fruit is famous for its bad smell?	durian
Who directed Ikiru and Drunken Angel?	Kurosawa	Who directed Wild Strawberries and Persona?	Bergman
Near which French city would you find The If Castle?	Marseille	In which French city would you find the largest cathedral?	Amiens
What is the name of the famous art gallery in Florence?	Uffizi	What is the name of the famous art gallery in Madrid?	Prado
What was the name of the deadliest Atlantic hurricane, which killed around 500 people in 1954?	Hazel	What was the name of the 2005 hurricane which devastated New Orleans?	Katrina
Which volcano caused the year without a summer in 1816?	Tambora	Which volcano destroyed the ancient city of Pompeii?	Vesuvius
Which Hollywood actor starred in the one cast member movie All is Lost?	Redford	Which Hollywood actor's real name was Marion Robert Morrison?	Wayne
Which element has the symbol Sn?	tin	Which element has the symbol Au?	gold
Which Australian bird is known for its laughter?	kookaburra	What is the only type of bird which can fly backwards?	hummingbird