Take a guess! Why generating predictions is more effective than generating examples in primary school children

Janiek Huijser
S2686201

12 November 2020

Supervisor: dr. Dietsje Jolles
Master Education and Child Studies
Applied Neuroscience in Human Development
Leiden University

Abstract

There has been a long-standing debate about the effectiveness of active versus passive learning. Proponents of active learning present generative learning strategies (GLS) as an effective way of learning, whereas others reject this constructivist idea, saying children's brains are not developed enough for it. This study investigated this claim by looking at the underlying mechanisms of two GLS, prediction generation and example generation. It specifically looked at the way prediction errors can contribute to learning. Existing data from a study by Breitwieser and Brod (2020) were used, including 25 children (M = 9.84) and 25 young adults (M = 21.24). The experimental task consisted of a study phase with two conditions, in which participants had to predict a numerical quantity in an incomplete numerical fact (e.g., "X out of 10 animal species are insects") or had to give an example related to the theme of the numerical fact ("butterflies are insects"). After every prediction or example, the correct answer was shown. Participants completed both conditions. In the following test phase, they had to complete all the incomplete numerical facts, without receiving feedback. Results were that children seemed to benefit more from prediction generation, compared to example generation, than adults. Specifically, making a prediction error seemed to stimulate learning, with a large error leading to a greater effect. Limitations and implications for future research and educational practice are discussed.

Introduction

Recently, the annual report of the Dutch Inspectorate of Education (Inspectie van het Onderwijs, 2020) was published, analyzing the current state of education in The Netherlands. It found that the best performing primary schools had a few matching features. Three teacher-related factors that repeatedly seemed to explain their success, were: responding to the individual needs of students, involving them in the instruction and letting them take responsibility for their learning process. Students indicated that this increased their motivation and self-confidence (Inspectie van het Onderwijs, 2020).

Stimulating students to take ownership of their learning process is something that has received increasingly more attention over the last two decades (e.g. Bada & Olusegun, 2015; O'Shea & Leavy, 2013; Stefanou, Perencevich, Dicintio, & Turner, 2004; Terhart, 2003). Students are encouraged to play an active role in their learning process and to learn from their own mistakes (Stefanou et al., 2004). Recent studies show that this can have positive learning effects: children generate their own answers to questions, to find out their answers are not always correct, which then leads to surprise and subsequently to learning (Breitwieser & Brod, 2020; Brod, Breitwieser, Hasselhorn, & Bunge, 2019).

This way of learning seems consistent with the theory of constructivism, which states that learning is a process in which the child plays an active role. Understanding is achieved by actively connecting prior knowledge to new knowledge in a meaningful way (Piaget, 1926; Vygotsky, 1978). In other words: learning is a generative activity (Fiorella & Mayer, 2016) where the learner is responsible for actively constructing knowledge, by mentally integrating the information that is to be learned with prior knowledge (Wittrock, 1974). Inspired by this theory are generative learning strategies

(GLS) that aim to help the learner integrate new knowledge with prior knowledge (Fiorella & Mayer, 2016). Although they are not mentioned in the report of the Dutch Inspectorate of Education, there is evidence that these GLS result in better learning outcomes than passive learning strategies (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Fiorella & Mayer, 2016; Lee, Lim, & Grabowski, 2008).

The current study will examine this more closely by focusing on two GLS: prediction generation and example generation. These two were selected because it is expected they lead to different effects; different executive functions are needed (e.g., working memory vs. analogical reasoning) and different underlying mechanisms are assumed (for example, surprise plays an important role when making predictions, but not when generating examples; Brod, Hasselhorn, & Bunge, 2018). These processes will be discussed in detail later.

This study will reanalyze data from a study by Breitwieser and Brod (2020), who looked at the effects of prediction and example generation on numerical fact learning. They compared 25 children (aged 9-11 years) and 25 adults (aged 17-29 years) who had to give a prediction or an example of an unfinished fact before seeing the right answer. An unfinished fact could be “X out of 10 historical statues are female figures”. A prediction that subsequently could be generated, was “4”. An example could be “There is a statue of Joan of Arc in Paris”. Results were that both GLS were equally effective in adults, whereas children benefited more from prediction generation. They remembered more facts on the final test after generating a prediction than after generating an example. In other words, there seemed to be a significant interaction between age and learning strategy (Breitwieser & Brod, 2020). The current study will examine the mechanisms behind prediction generation.

Experiential learning versus direct instruction

One way to improve the learning performances of young students might be to adjust the instruction to their cognitive development. However, what kind of instruction this should be, remains a discussion. There is a long-standing debate around the question how much guidance a student needs during learning, with direct instruction on one side of the continuum, and the model of ‘minimal guidance’ on the other side (Brunstein, Betts, & Anderson, 2009; Kirschner, Sweller, & Clark, 2006).

The model of minimal guidance has been called by different names, varying from experiential learning (Kolb, 1984) and inquiry learning (Rutherford, 1964) to constructivist learning (Steffe & Gale, 1995) and discovery learning (Anthony, 1973). The basic idea of this model is to let students take responsibility for their own learning process, with minimal instructional guidance. Advocates of direct instruction, on the other hand, state that learners need a teacher to provide clear instructions and information about the concepts they have to study. Students should not discover this all by themselves (Kirschner et al., 2006).

A large base of research states that direct instruction has more beneficial effects on learning than constructivist learning (Klahr & Nigam, 2004; Mayer, 2004; Moreno, 2004; Sweller, 2004). Kirschner et al. (2006) argue that pure constructivist learning does not work based on our ‘cognitive architecture': it creates a working memory load that is too heavy. Indeed, self-directed learning appears to place high demands on executive functioning and self-regulation; skills that have not fully matured yet in primary schoolers (Diamond, 2013). Lastly, Kirschner et al. (2006) warn for the negative effects of constructivist learning, saying that students may form misconceptions or incomplete knowledge about a topic.

Generating predictions

Still, some *aspects* of constructivist learning might be beneficial for learning in combination with direct instruction and might potentially be even more effective than pure instruction. GLS like prediction generation and example generation are among these aspects. Evidence suggests that generation is a powerful tool to enhance learning and retention (Roediger & Pyc, 2012). This involves generating information from one's own mind, instead of being presented with external information (Slamecka & Graf, 1978). In the classic generation model, participants are asked to learn a list of certain items in two conditions: a read condition, in which the items are presented in their complete form (for example, GARBAGE-WASTE; HOT-COLD; et cetera) and a generate condition, in which participants have to generate the correct item by using semantical or phonological cues, like synonyms, antonyms or rhymes (for example, GARBAGE-W_ST_; HOT-____; Bertsch, Pesta, Wiscott, & McDaniel, 2007; Rosner, Elman, & Shimamura, 2013). The benefits of generation over reading have been demonstrated in many different generation tasks, using a wide range of material, and in both healthy adults and patient populations (for an overview, see Bertsch et al. (2007)).

A specific form of generation is generating examples; one of the GLS. Generating examples appears to improve concept learning, for example in the field of psychology (Rawson & Dunlosky, 2016) and mathematics (Sağlam & Dost, 2016). To select appropriate examples, learners need to activate relevant prior knowledge, generate multiple examples, compare them with the question and with each other, and eventually decide which one is most suitable (Breitwieser & Brod, 2020). Hence, this strategy demands analogical reasoning (Duit, 1991); a skill that is still developing during childhood and adolescence. Therefore, according to Breitwieser and Brod (2020) it is plausible that generating examples is less effective in children than generating predictions.

Generating predictions, another GLS, seems to stimulate learning in many different fields, including physics (Crouch, Fagen, Callan, & Mazur, 2004; Inagaki & Hatano, 1977), biology (Schmidt, De Voider, De Grave, Moust, & Patel, 1989), geography (Brod et al., 2019; Brod, Hasselhorn, & Bunge, 2018) and trivia fact learning (Brod & Breitwieser, 2019). Schmidt et al. (1989) suggest that generating predictions requires learners to access their prior knowledge. This is known to increase recollection and apprehension of new information (Bransford & Johnson, 1972). This way, generating predictions could stimulate learning. An example of a prediction generation task can be found in the study of Brod et al. (2018). The goal was to improve participants' geography learning by asking them either to predict which of two countries had the largest population, or to make post-hoc evaluations, i.e. to indicate what they would have expected after they had seen the correct answer. Results were that making predictions was most effective for learning geography facts (Brod et al., 2018).

One aspect of generating predictions is making prediction errors. Brod et al. (2018) observed in their geography experiment that expectancy-violating information enhanced memory, compared to post-hoc evaluations. In other words: when adults are exposed to new information that contradicts their prior knowledge (i.e. prediction errors), this can increase their learning performance. The authors describe that answers that violated the participants' expectations after predicting the possible outcome, led to a surprise response (measured with the pupil diameter). In turn, the strength of this surprise response correlated positively with the extent of learning (Brod et al., 2018). These prediction errors seem especially beneficial when accompanied with corrective feedback (Henson & Gagnepain, 2010; Huelser & Metcalfe, 2012; Potts, Davies, & Shanks, 2019). Potts and Shanks (2014) demonstrated that prediction errors, in combination with feedback, improved memory much better than just reading the information, or selecting one answer from a few possible answers and getting feedback. An answer to the question why prediction generation, and prediction errors specifically, seem to benefit learning, can be found in literature about the brain.

Prediction in the brain

The human brain is constantly generating predictions (Bar, 2007). This is a universal mechanism to adapt most effectively to the ever-changing environment (Greve, Cooper, Kaula, Anderson, & Henson, 2017). Bar (2007) demonstrated that memory systems use associations to produce predictions. Objects and events that often emerge together, are linked to each other in a scheme. This facilitates the detection of trends and the anticipation on future events – in other words, prediction (Bar, 2007). The brain areas that seem important for these associations, have a significant overlap with brain areas that are part of the default mode network (Bar, Aminoff, Mason, & Fenske, 2007). This network consists of regions (including the medial temporal lobe, medial prefrontal cortex, and medial parietal cortex) that are active when the individual is resting, i.e. not doing a specific task (Raichle et al., 2001). Since these regions also seem to be involved in forming associations, this supports the idea that the brain is constantly generating predictions, instead of passively waiting to be activated by stimuli (Bar, Aminoff, Mason, & Fenske, 2007). In other words, the specific brain regions for making predictions are still active when there is no specific reason to be active (i.e., a particular task). Therefore, it is unlikely that these brain regions are passively waiting to be activated. Hence, it is assumed that generating predictions is a continuous activity in the brain.

Memory systems. Research regarding the brain mechanisms underlying prediction could give insight into why prediction seems to work so well during learning. In general, there are two important memory systems: working memory and long-term memory. Working memory holds information in mind for a short period of time (Vallar, 2015). This is necessary to successfully carry out a task that one is currently working on (Pinel & Barnes, 2014). For example, when making predictions it is important to keep in mind the initial answer until the correct one is given, so that the two can be compared. Working memory is thought to be consolidated by the hippocampus into long-term memories (Squire & Bayley, 2007).

The long-term memory system can be divided in multiple subsystems, each with their own participating brain regions. Long-term explicit memory includes declarative memory and is important for the conscious remembrance of events and objects (Kolb & Wishaw, 2015). This seems to play a role when making an 'episodic prediction error': having violated expectations about an event (Lisman & Grace, 2005). Long-term implicit memory (also called nondeclarative memory) plays a role in conditioning. When making a prediction error, a condition process takes place that minimizes the chance of making a similar error in the future (Van den Bos, Cohen, Kahnt, & Crone, 2012). Altogether, these memory systems work together to generate the best possible predictions and to learn from prediction errors (Henson & Gagnepain, 2010), thereby contributing to the effectiveness of GLS.

Interacting memory systems. When predictions do not match the correct answer, this introduces the prediction error. Many brain regions seem to be involved in the processing of this prediction error (for an overview, see Schultz & Dickinson, 2000). Henson and Gagnepain (2010) link the prediction error to learning via the PIMMS model, where PIMMS is an abbreviation for "predictive interactive multiple memory systems". Memory systems interact with each other to minimize prediction errors in the future, by continuously comparing incoming information with top-down predictions. The synaptic changes following a prediction error increase memory; the larger the prediction error, the greater these synaptic changes and the better the encoding of memories (Henson & Gagnepain, 2010). This suggests that the prediction error is a signal that something needs to be changed, which could explain why generating predictions is more effective than generating examples or only giving the correct answer.

When taking a closer look at the different memory systems that are interacting as part of the PIMMS network, there seem to be two important structures involved: the medial temporal lobe network (MTL) and the striatal network. The latter seems to be involved in nondeclarative learning: the acquisition of skills and habits (Doyon et al., 2009; Filoteo, Maddox, & Davis, 2001; Myers et al., 2003). The striatum itself is believed to be involved in the storage of persistent relationships between stimuli and responses (Laubach, 2005), procedural learning (Poldrack, Prabhakaran, Seger, & Gabrieli, 1999) and reward learning (O'Doherty, 2004). Specifically, the dorsomedial striatum appears to be involved in goal-directed learning (Yin & Knowlton, 2006) and instrumental conditioning (Yin, Ostlund, Knowlton, & Balleine, 2005). The dorsolateral striatum seems involved in habit learning (Voorn, Vanderschuren, Groenewegen, Robbins, & Pennartz, 2004; Yin, Knowlton, & Balleine, 2004).

Prediction errors are often associated with reward learning in the striatum (O'Doherty, 2004; O'Doherty et al., 2004). Here, prediction errors are encoded by dopaminergic neurons (McClure, Berns, & Montague, 2003; Schultz, Dayan, & Montague, 1997). When the prediction error is either positive or negative (meaning that the outcomes are respectively better or worse than expected), the striatum sends a signal to the prefrontal cortex, which, in turn, learns to adapt to this outcome in the future (Van den Bos et al., 2012). According to Doll, Jacobs, Sanfey, and Frank (2009), the mature prefrontal cortex also *controls* this striatal activity, meaning it can suppress the prediction error when it is inconsistent with the (direct) instruction or their goals. In other words: in adults, the prefrontal cortex strengthens the influence of outcomes that are consistent with the instruction. The effects of instruction-inconsistent outcomes, on the other hand, are reduced. This creates a confirmation bias: it stimulates adults to appraise actions or ideas that were recommended through direct instruction more highly than actions or ideas that were learned through experience only. Even when the recommendations turn out to be inaccurate, this remains the case (Doll et al., 2009).

The MTL, including the hippocampus, seems to be responsible for declarative learning, i.e. semantic and episodic memories (Knowlton, Mangels, & Squire, 1996; Ranganath & D'Esposito, 2001). The hippocampus is also involved the processing of prediction errors, although these errors seem different from the prediction errors the striatum is involved in. The hippocampus is activated by events that are linked to an existing memory representation (meaning there is a learnt association), but differ in some way from that representation (Lisman & Grace, 2005). Or, more easily put: it detects mismatches between contemporary sensory input and past experiences (Kumaran & Maguire, 2007). Lisman and Grace (2005) call this the *episodic* prediction error. Their hippocampal-VTA loop model states that when the hippocampus detects a mismatch, it sends a signal to the substantia nigra/ventral tegmental area (SN/VTA), where neurons synthesize dopamine. In turn, the SN/VTA stimulates the release of dopamine into the hippocampus. Shohamy and Wagner (2008) later added to this model that this release of dopamine enhances the encoding of both past and current events into an integrated representation.

As already stated in the PIMMS model, there seems to be an interaction between the striatal network and the MTL (including the hippocampus). A growing base of evidence states that the interaction between these memory systems can be both competitive (activation of one system decreases the activation of the other) and cooperative (both systems make parallel contributions to the learning process; Freedberg, Toader, Wassermann, & Voss, 2020; Sadeh, Shohamy, Levy, Reggev, & Maril, 2011; Wimmer, Braun, Daw, & Shohamy, 2014). Whether the interaction is competitive or cooperative, could depend on the specific learning context (Delgado & Dickerson, 2012). During trial-and-error learning, for example, the interaction seems mostly cooperative (Dickerson, Li, & Delgado, 2011). In this context of feedback learning, activity in the hippocampus could reflect attempts to identify differences between positive and negative outcomes (Li, Delgado, & Phelps, 2011) alongside the corticostriatal systems, possibly encoding a feedback-based model-derived reward prediction error signal (Dickerson et al., 2011).

Surprise and curiosity as underlying mechanisms. There are more suggestions for processes that play a role when making predictions or prediction errors. Brod et al. (2018) argued that surprise might play a role in learning from the prediction of facts. Expectancy-violating outcomes, i.e. prediction errors, lead to an emotional reaction: the surprise response (D'Mello et al., 2014; Ekman, 1992).This response has been linked to the release of noradrenalin in the locus coeruleus (Aston-Jones & Cohen, 2005), which in turn stimulates the development of long-term memory (McGaugh & Roozendaal, 2009). In addition, enhanced arousal may elevate attention to the surprising results, which also increases memory for these facts (Fazio & Marsh, 2009; Stahl & Feigenson, 2015). Then, the surprise response is followed by confusion: a cognitive disequilibrium that can be beneficial for learning as well (D'Mello, Lehman, Pekrun, & Graesser, 2014). Indeed, Brod et al. (2018) found that prediction errors led to a surprise response, and that the intensity of this response was positively related to memory performance. This is in line with other research, saying that the more surprising the information is, the better it will be remembered (Fazio & Marsh, 2009; Greve, Cooper, Kaula, Anderson, & Henson, 2017). Butterfield and Metcalfe (2001) called this phenomenon 'the hypercorrection effect'. Accordingly, both the surprise response and the prediction error seem to send certain signals when outcomes do not match expectations. The difference is that the surprise response seems to be an *effect* of the prediction error; an emotional reaction that would not exist without the prediction error being made. Possibly the surprise response strengthens the effects of the prediction error, depending on its intensity. Another difference is that metacognitive skills are needed for surprise; previous beliefs need to be changed in the light of new information (Brod et al., 2019). The prediction error, on the other hand, seems to happen almost automatically and unconsciously in the PIMMS network (Henson & Gagnepain, 2010).

Another mechanism that could underly learning from prediction, is curiosity: a process that emerges before the prediction error takes place. In their experiment, Brod and Breitwieser (2019) found that generating predictions stimulated curiosity, which in turn was related to a better memory for the correct answer. They explain this by saying that curiosity, just like surprise, stimulates the release of noradrenalin from the locus coeruleus (Kang et al., 2009). Indeed, they could demonstrate that higher levels of curiosity were accompanied with a larger pupil dilation. Generating examples did not lead to such high levels of curiosity as generating predictions and was related to less correct recalls (Brod & Breitwieser, 2019).

Perseverative errors. A risk of making prediction errors could be that students remember their incorrect guess instead of the correct answer. Marsh, Roediger, Bjork and Bjork (2007) found that students can make such perseverative errors during a multiple-choice test, remembering the false lures instead of the correct option. In addition, Vaughn and Rawson (2012) did three experiments in which they compared an associative guessing condition with a study condition (participants could study the associations before the final test). In their first experiment, they found that studying outperformed incorrect guessing. The second experiment revealed that timing of feedback functioned as a moderator: when the incorrect guessing was followed by immediate feedback, participants performed better in the feedback condition than in the studying condition. But when feedback was delayed, incorrect guessing lead to a worse recall. According to Vaughn and Rawson (2012), this could have to do with perseverative errors, in the sense that participants confused their previous guess with the correct answer.

However, their third experiment gave evidence against this suggestion. In the final test, they asked participants to indicate if they made a guess for this specific stimulus in the practice phase (with the guessing condition), and if so, whether they remembered what they guessed. Participants from both groups (delayed-study and immediate-study) were equally likely to remember their original guesses correctly (Vaughn & Rawson, 2012). Still, other studies demonstrate that errors made with high confidence persevered when they were not tested immediately after the participant realized it was an error (Butler, Fazio, & Marsh, 2011; Metcalfe & Miele, 2014).

To summarize, brain research can give some possible explanations for the effectiveness of generation predictions on learning. These explanations are, respectively: synaptic changes resulting from predictive interactive multiple memory systems, the surprise response, and curiosity. All three of them increase memory, but this could be a memory of the incorrect answer. In other words, there may be perseverative errors when generating predictions, meaning that one remembers the incorrect prediction instead of the correct answer. However, these studies did not investigate the underlying (brain) mechanisms in children during prediction generation. Research in these age group is scarce; still, there are some studies that shed light on the possible brain systems of children that play a role during prediction.

Effects of prediction in children

In infants and children, violated expectations or prediction errors seem to stimulate learning (Reuter, Lev-Williams, & Borovsky, 2019; Stahl & Feigenson, 2015; Stahl & Feigenson, 2017). In children, the striatal reward signals look quite mature already (Cohen et al., 2010). However, the connection between the striatum and the prefrontal cortex is still developing into adulthood (Liston et al., 2006). This relative immaturity of the prefrontal-striatal connection seems to have an influence on children’s way of learning. Whether this is a positive or a negative influence, depends on the situation. Immaturity of the prefrontal cortex seems to be related to a weaker level of cognitive control, meaning that children are less able to inhibit inappropriate responses and more easily distracted by the environment (Bunge, Dudukovic, Thomason, & Vaidya, 2002). In addition, because the connection between the prefrontal cortex and the striatum is not fully developed yet, there is less top-down control of behavior under motivational demands (e.g. rewards; Barrett, Fox, Morgan, Fidler, & Daunhauer, 2013).

However, the immaturity of the prefrontal-striatal connection could have a positive influence on the process of generating predictions. A study by Decker, Lourenco, Doll and Hartley (2015) compared the effectiveness of feedback learning to direct instruction in children. These children appeared to benefit more from feedback learning than adults; in fact, the effects of feedback learning seemed to be intensified in children, compared to adults. The explanation Decker et al. (2015) gave for these results, is the immaturity of the prefrontal-striatal connectivity. As discussed, adults have a confirmation bias for actions or ideas that were recommended through direct instruction. It strengthens the effect of outcomes that were consistent with instruction and diminishes the influence of instruction-inconsistent outcomes, or ideas that were learned through experience only. This seems to stem from the influence of the rule-following prefrontal cortex on the feedback-evaluative striatum (Doll et al., 2009). Children, however, seem to learn more from experience than direct instruction, because the connection between the prefrontal cortex and the striatum is not yet mature enough to create this confirmation bias (Decker et al., 2015). In conclusion, the immaturity of the prefrontal-striatal connectivity can be beneficial for learning.

Age could influence the effectiveness of GLS in general, for several reasons. First, age influences the amount of prior knowledge they have. Gurlitt and Renkel (2008) found that GLS are important learning support for students with a lower prior knowledge, suggesting that younger learners might benefit most from GLS. Second, age could affect GLS by influencing the level of executive functioning. It could be that different GLS have different developmental pathways, depending on the different executive functions that are involved (see Li et al., 2004). Brod et al. (2019) discovered that generating predictions only benefited learning in children with better inhibitory control skills. And, as was already explained before, example generation requires the executive function of analogical reasoning (Duit, 1991), an ability that undergoes strong developmental changes from childhood to adolescence (Richland, Kornell, & Kao, 2009). The same goes for inhibitory control, and specifically the error-processing circuit (Ordaz, Foran, Velanova, & Luna, 2013; Velanova, Wheeler, & Luna, 2008). Consequently, Breitwieser and Brod (2020) also looked at the influence of executive functions on GLS in their study. Indeed, there was an effect of analogical reasoning on example generation: children with better analogical reasoning skills performed more equally in the two GLS conditions. There was no effect of inhibitory skills on prediction, however (Breitwieser & Brod, 2020). Altogether, it seems that especially the GLS of generating predictions has a positive influence on children's learning – even though the underlying executive functions are still developing.

Depth of processing

Apart from age, there is something else that could influence prediction generation specifically: depth of processing. In a study of Benjamin, Bjork and Schwarz (1998), participants had to make two tests: one in which they had to answer trivia questions, and one in which they received a blank sheet of paper and were asked to free-recall their previous answers. During the first test, after every question participants were asked to rate the possibility that they would recall the answer to this fact. Results were that the more quickly participants responded during this first test, the more confident they were to recall their answer and, surprisingly, the *less* likely they were to recall it during the second test. Thus, participants were more likely to remember the answers they thought about for a longer time (Benjamin et al., 1998). Accordingly, depth of processing and retrieval success seem to be related. When applying this to prediction generation, it could be the case that the longer it takes participants to predict an answer, the more likely it is they will recall it.

The current study

In this study, the data of Breitwieser and Brod (2020) were used for a secondary analysis. In short, they asked children and adults to generate either a prediction or an example about an incomplete fact (e.g., “X out of 10 historical statues are female”). When they had given their answer, the true answer appeared, and the next fact was shown. After this study phase, participants were tested with the same unfinished facts to see how much they had remembered.

First, the current study repeated analyses by Breitwieser and Brod, focusing on the different effects of prediction and example generation in children and adults. Next, these data were used for additional, new analyses. It is still not known whether it is a large or a small difference between prediction and correct answer that leads to the best recall. One could argue that a large difference between prediction and correct answer leads to a larger neural prediction error and therefore to a better remembrance. On the other hand, a large difference could indicate that the participant does not have any idea what the answer could be. Then, the neural prediction error and the learning effect will probably be smaller. Therefore, this study wanted to find out if it is a large or a small behavioral prediction error that is responsible for a better recall of the correct answers. In order to do so, it was examined whether there were differences in performance between the facts with a high prediction error and the facts with a small prediction error. Furthermore, to examine the possibility that prediction errors could persevere (remembering the incorrect prediction instead of the correct answer), a second additional analysis focused on the presence of perseverative errors. Because knowledge and executive functions differ across development and between individuals, main effects and interactions with age, inhibitory skills and cognitive flexibility were also examined. Lastly, it is not known yet whether depth of processing is related to retrieval success in the context of prediction generation and example generation. Therefore, this study investigated the role of reaction times when making predictions.

Previous literature states that the more surprising the information is, the better it will be remembered (e.g. Brod et al., 2018; Greve et al., 2017; Butterfield & Metcalfe, 2001). Accordingly, it was hypothesized that the facts with a larger difference between prediction and true answer (i.e., a high prediction error) would be remembered correctly more often than the facts with a smaller difference (i.e., a small prediction error). Regarding perseverative errors, it was hypothesized that persons with lower levels of inhibitory skills and cognitive flexibility would make more perseverative errors than persons with higher levels of these executive functions. Building on that, it was also hypothesized that children would make more perseverative errors than adults. The last hypothesis, based on Benjamin et al. (1998), was that the longer the reaction time in the study phase, the better the memory performance in the test phase, and the other way around.

Method

For this study, data from the study of Breitwieser and Brod (2020) was used. It was preregistered on the Open Science Framework.

Participants

Data were collected from 25 children of 9-11 years old (10 females; *M* = 9.84, *SD* = 0.47) and 25 young adults of 17-29 years old (17 females; *M* = 21.24, *SD* = 3.68). All participants came from middle-class households. The children were recruited through a database of children that had participated in previous studies of the research group in Frankfurt, and via an email that regularly distributes information to parents of fourth graders in Frankfurt. Adults were acquired via the informal network of the researchers, advertisements on the bulletin boards in the Goethe-University of Frankfurt and announcements in student groups on social media. All participants were asked to fill out informed consent forms. The children received a toy worth five euros for their participation, their parents received five euros to compensate for their travel costs, and the adults received ten euros or ECTS (course credit). The ethics committee of the DIPF | Leibniz Institute for Research and Information in Education gave permission for this research.

Experimental task

A task of 60 numerical facts was designed. They were written in such a way that they were incomplete. The answer needed to be a number (ranging from 1 to 9) divided by 10. For example, a fact could be: “X out of 10 animal species are insects”. The experimenters ensured all children knew X represented a number. In both the prediction condition and the final test, a visual analog scale (VAS) was used to support participants with low numeracy skills. PsychoPy v1.8 was used to present all the facts (Peirce, 2007). In the prediction condition, participants had to predict a number (e.g., the number of animal species that are insects). The answer would light up for one second. Then, after a brief delay, the correct answer to the fact appeared on the screen for three seconds. Subsequently, the next incomplete fact would be presented. In the example condition, the participants had to come up with an example for the fact that was shown (e.g. “butterflies are insects”). As soon as they had found an example, they had to click on a smiley button. If they could not think of an example, the participants could click on a red button. Before each condition, there were three practice trials. For a schematic overview, see figure 1.

For each condition, there was a study phase and a test phase. Participants completed both conditions successively. The order of the conditions was counterbalanced across the participants. After the prediction and the example condition, there would be a test phase in which the participants received the same 30 facts as in the study phase. This time, all the facts were incomplete, and the participants needed to remember the right answer. Their response was again highlighted for one second. The right answer was not shown.

From all the 60 numerical facts, two subtests of 30 facts were created. The facts were assigned randomly to each subtest. After 30 trials there was a short break. In total, the sessions took 60 minutes for young adults and 70 for children.

Procedure

Before the experimental task, participants performed two executive functions tests that measured their inhibitory skills and cognitive flexibility, one of them being the Hearts and Flowers task (HFT; Wright & Diamond, 2014). This task has been validated with both children (4-13 years old) and adults (Davidson, Amso, Anderson, & Diamond, 2006; Diamond, Barnett, Thomas, & Munro, 2007) and has been shown to have concurrent validity with other tasks measuring inhibitory skills (Brocki & Tillman, 2014). Task parameters and test procedure were the same as in the study of Brod, Bunge and Shing (2017). The task consisted of three blocks, each containing 20 trials, that increased in level of executive functioning needed to complete the task successfully. The first block was the congruent condition (and baseline condition), in which participants had to press a button at the side where the heart appeared on the screen. With the second block, the incongruent condition, inhibitory skills were assessed. A flower appeared on the screen and participants had to press a button at the opposite side relative to the location of the flower. The third block, the mixed condition, measured cognitive flexibility. Both hearts and flowers appeared, and participants had to switch between the heart rule and the flower rule.

Another task that measured executive functions, and specifically their perseverative errors, was the Berg Card Sorting Test (BCST; Fox, Mueller, Gray, Raber, & Piper, 2013): an abbreviated, low-cost, open-source version of the better-known Wisconsin Card Sorting Task. Participants had to match cards from a deck to 4 cards at the top of the screen, according to a rule they had to discover (for example, “sort on color”). This sorting rule changed over time (to, for example, “sort on shape”). With moderate intra-test and low inter-test correlations, the convergent and divergent validity were sufficient (Piper et al., 2015).The (test-retest) reliability of this task was judged to be intermediate (Pearson’s r = 4.5, Spearman’s rho = 3.5; Piper et al., 2015).

The experimental task was carried out individually. Participants were instructed to remember the correct answers for the final test. In the study phase, they predicted 30 numerical facts (prediction condition) and had to give an example for 30 numerical facts (example condition). Participants were instructed to remember the facts for the later memory test. During both study phases, eye tracking data was collected, which focused on pupil size changes in response to the correct answer. After the study phase, there was a short task of approximately one minute, to clear the participants’ short-term memory from any numerical facts. This was done with the de Digit Span Backward test (DSB; Wechsler, 2008) in which participants had to recall a series of numbers in the reversed order of how they were presented. This intermediate task was followed by the test phase. During both study and test phase, reaction time of the participants was measured.

Finally, participants were asked to fill out a questionnaire. They were asked to indicate the facts they had already known prior to the experiment, the condition they thought was the most educational, and the condition they experienced as more enjoyable (ranging from 1, clearly prediction, to 6, clearly example).

Design

A design was used that included both a within-subjects component (the generative learning strategies) consisting of two levels (the two conditions: generating predictions and generating examples) and a between-subjects component (age). Before the actual analyses started, it was investigated whether this study could produce the same results as Breitwieser and Brod (2020) – an interaction between condition and age on memory performance – using a slightly different analysis procedure. Memory performance was operationalized into proportion of answers correct on the final test. Then, it was examined if there was an effect of the size of the prediction error in the study phase on memory performance. A prediction error was defined as any answer but the correct one in the prediction condition of the study phase. The size of the prediction error was measured using the distance between the incorrect prediction and the correct answer. Here, memory performance was operationalized as the difference between prediction error sizes of correct and incorrect items on the test.

The second question was whether there were perseverative errors and if there was an effect of age and level of executive functioning on the number of perseverative errors. A difference was made between the number of corrected errors, i.e., facts that were predicted incorrectly, but answered correctly in the test phase, and the number of perseverative errors, i.e., facts that were predicted incorrectly during the study phase for which the exact same incorrect response was given during the test. Level of executive functioning was operationalized into inhibitory skills and cognitive flexibility, two variables that were in turn operationalized into reaction times on the different conditions of the HFT. A faster reaction time in the incongruent condition meant a higher level of inhibitory skills, and a faster reaction time in the mixed condition meant a higher level of cognitive flexibility. In addition, cognitive flexibility was also operationalized into the proportion of perseverative errors on the BCST, with lower proportions reflecting a higher flexibility.

Lastly, it was investigated whether depth of processing in the study phase influenced memory performance (which is again about the difference between correct and incorrect items in the test phase), and if this differed by age. Depth of processing was operationalized into reaction times during the study phase, based on the study by Benjamin et al. (1998).

Statistical analyses

To replicate the findings of Breitwieser and Brod (2019), a repeated measures ANOVA was used. For the second research goal, the facts from the experimental task were categorized in ‘answered correctly’ and ‘answered incorrectly’ in the test phase. Prediction errors were analyzed in the prediction condition only. The means of these categories were compared by using a repeated measures ANOVA as well. For the third research goal about perseverative errors and executive functioning, hierarchical linear regression analyses were carried out. For the fourth research goal regarding depth of processing, reaction times during the prediction and example condition were compared with a repeated measures ANOVA.

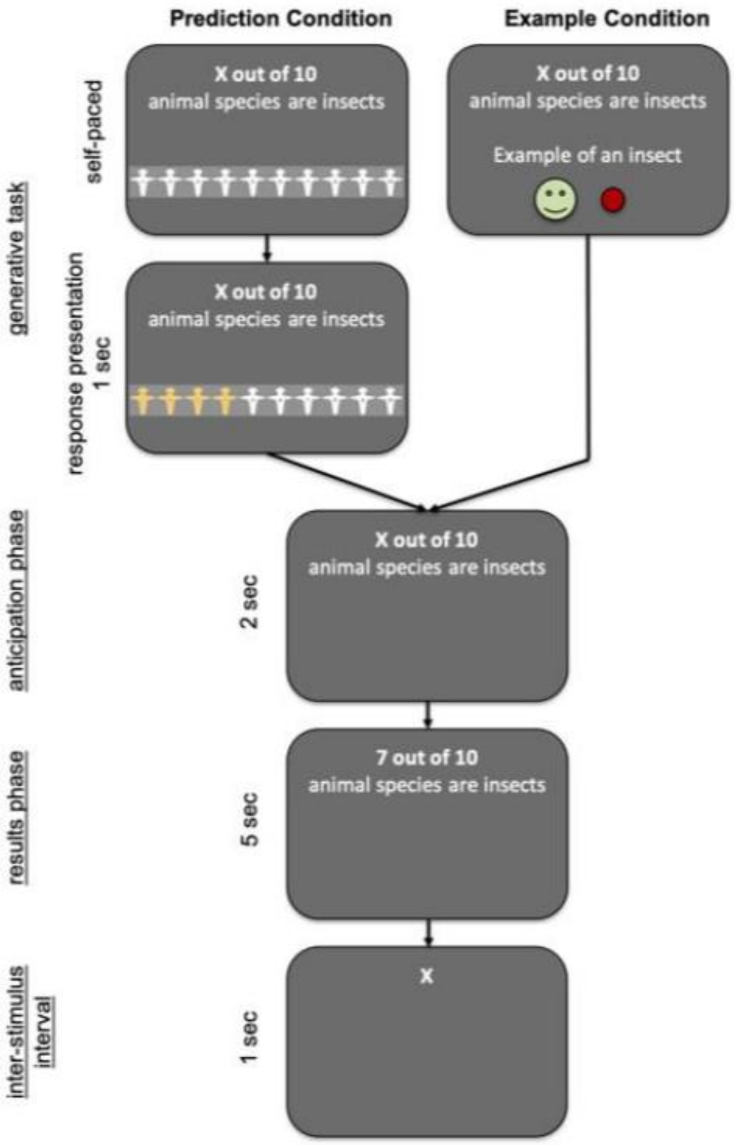

Figure 1. Schematic overview of the study phase of the experimental task, in which participants had to generate a prediction about the value of X (in this case, "X out of 10 animal species are insects") or generate an example that was related to the unfinished fact.
Note. Reprinted from "Cognitive prerequisites for generative learning: Why some learning strategies are more effective than others" by Breitwieser, J., and Brod, G. (in press), 2020, *Child Development*.

Results

Data exploration

Facts known prior to the experiment were deleted per participant. Also deleted were reaction times (RTs) of zero, that were found once or twice in three participants. One participant had thirty RTs of zero, which were all deleted. Contrary to Breitwieser and Brod (2019/2020), facts with no examples were not deleted from the data. The reason for that is that participants seemed to have thought of their answers before deciding not to give an example. The average RT for these facts was 12.12 seconds (*SD* = 7.06), which is significantly longer than the average RT of 6.74 seconds (*SD* = 2.22) for the facts where an example was given (t (22) = -4.225, $p < 0.05$, 95% CI [-9.06;-3.09]). Having thought of examples could still lead to a better memory for the facts in the test phase. In addition, the example condition was more similar to the prediction condition this way. After all, it is highly likely that the prediction condition also consisted of facts participants did not know anything about. Interestingly, an independent *t* test demonstrated that adults (M = 0.998, SD = 0.013) and children (M = 0.925, SD = 0.007) differed significantly in how many facts they left blank on average in the example condition (t (37.019) = -20.980, $p < 0.001$, 95% CI [-0.069;0.056]).

Histograms and normality plots were inspected to gather insight in the distributions of the variables (for the descriptive statistics, see table 1.). Proportion of facts correct in the test phase was relatively normally distributed for both the prediction condition and the example condition from the study phase. Though when looking at the age groups separately, the distributions appeared to be more positively or negatively skewed. The distributions of RT were positively skewed, and this was also the case in both age groups. The same went for the distributions for prediction errors, perseverative errors and corrected errors (i.e., number and proportion of facts that were incorrect in the prediction condition of the study phase and correct in the test phase). However, these skewed distributions were not highly problematic since the repeated measures ANOVA is relative robust to violations of normality.

Replication analyses

Before performing the planned analyses, it was important to establish that the findings of Breitwieser and Brod (2020) could be replicated using a slightly different analysis method (i.e., repeated measures ANOVA instead of logistic/linear mixed-effects regression). A repeated measures ANOVA was carried out for the proportion of facts that were correct in the test phase and were learned previously in either the prediction condition or the example condition in the study phase. Condition was significant (F (1, 48) = 9.401, p = 0.004), meaning that there was a significantly higher proportion of correct facts in the test phase when those facts were learned in the prediction condition of the study phase, compared to facts learned in the example condition. Also, there was a main effect of age (F (1, 48) = 39.277, p < 0.001), indicating that adults had a significantly higher proportion of correct facts in the test phase than children, for both conditions of the study phase. The interaction of age with condition was not significant (F (1, 48) = 2.186, p = 0.146). These results are not in line with those of Breitwieser and

Brod (2020), who found a significant interaction effect, i.e. a stronger effect for generating predictions than for generating examples in children compared to adults. Therefore, post-hoc analyses were carried out by looking at the age groups separately. The difference between the conditions was significant for children (F (1, 24) = 11.629, p = 0.002), but not for adults (F (1, 24) = 1.133, p = 0.298). So, although the age x condition interaction was not significant, there did seem to be a trend in which children benefited more from making predictions than adults.

To analyze this further, this study looked at the average absolute difference between test answer and true answer, for facts that were learned in either the prediction condition or the example condition. Repeated measures ANOVAs were carried out again; condition was significant (F (1, 48) = 25.642, p < 0.001), indicating that the difference between the test answer and the true answer was significantly smaller for facts learned in the prediction condition of the study phase, compared to the example condition. There was a main effect of age (F (1, 48) = 43.550, p < 0.001). This means that on average, adults had a significantly smaller difference between test answer and true answer in the test phase than children, for both conditions of the study phase. Lastly, the interaction between condition and age was significant as well (F (1, 48) = 4.842, p = 0.033; see figure 2). When analyzing the age groups separately, the average differences were significant for both children (F (1, 24) = 16.701, p < 0.001) and adults (F (1, 24) = 9.756, p = 0.005). Facts that were learned in the prediction condition thus resulted in less incorrect answers during the test phase, in comparison to facts learned in the example condition. For children, this difference was significantly larger than for adults. In conclusion, the previous analysis demonstrated that the prediction condition resulted in more correct facts in the test phase, whereas this second analysis found that the prediction condition resulted in less incorrect answers in the test phase. The difference is that the second analysis looked at the distance between the test answer and the correct answer. These results provided useful information for the following analysis about prediction errors.

Table 1

Descriptive statistics of all variables except age

	Children (n = 25)		Adults (n = 25)	
	M	*SD*	*M*	*SD*
Replication analyses				
Test phase: correct (proportion)				
Prediction condition	0.47	0.14	0.66	0.12
Example condition	0.36	0.17	0.62	0.16
Test phase: Difference between test answer and true answer				
Prediction condition	1.12	0.38	0.53	0.22
Example condition	1.68	0.80	0.75	0.36
Additional analyses				
Study phase: Difference between prediction and true answer for test-correct	0.52	0.47	1.16	0.32
Study phase: Difference between prediction and true answer for test-incorrect	1.13	0.41	1.00	0.37
Proportion corrected errors[a]	0.216	0.071	0.030	0.086
Proportion perseverative errors[b]	0.046	0.028	0.046	0.031
Study phase: RT for test-correct (prediction)	7.34	1.80	5.47	0.94
Study phase: RT for test-incorrect (prediction)	7.69	1.62	5.92	2.33
Study phase: RT for test-correct (example)	8.31	3.50	5.55	2.07
Study phase: RT for test-incorrect (example)	8.50	2.92	5.56	2.10
BCST				
Proportion perseverative errors	0.146	0.037	0.102	0.018
HFT				
RT incongruent block	543.23	92.70	367.10	81.11
RT mixed block	784.84	151.10	554.79	107.93

Note. [a] Proportion corrected errors is defined as the proportion of facts incorrect in the prediction condition and correct in test phase.

[b] Proportion perseverative errors is defined as the proportion of facts with the same incorrect answer in both the prediction condition and the test phase.

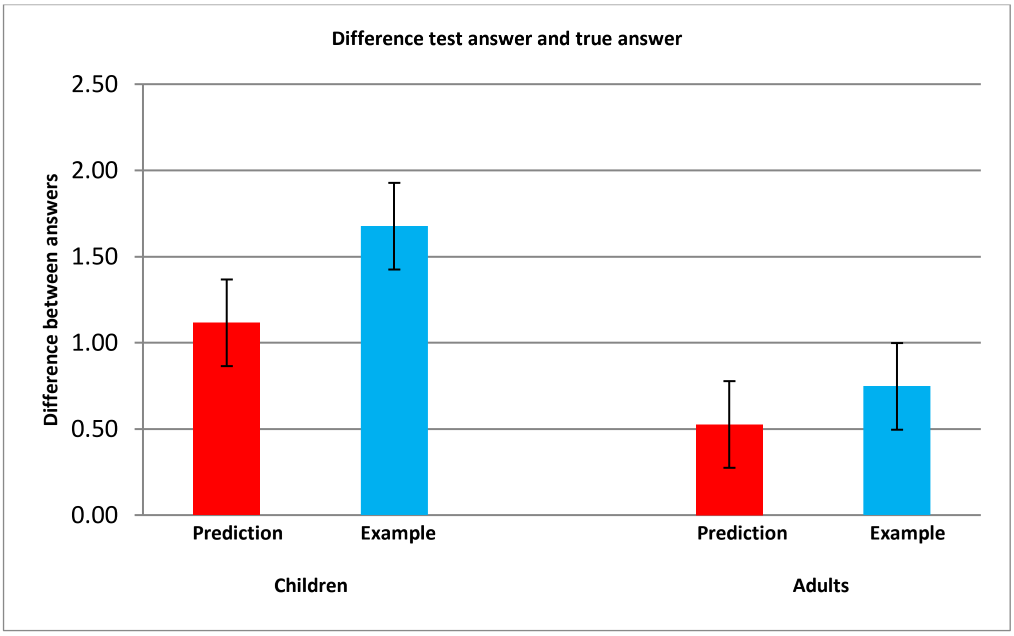

Figure 2. Graph of participants' performance in the test phase. The error bars represent standard errors of the means. This figure indicates that facts learned in the prediction condition resulted in fewer incorrect answers during the test phase, than facts learned in the example condition. For children, this difference was significantly larger than for adults (hence, there was an interaction effect).

Additional analyses

Prediction errors. As a result of the replication analyses, it was now known that facts learned in the prediction condition lead to more correct answers in the test phase than facts learned in the example condition. To test whether this was related to the size of the prediction error, another repeated measures ANOVA was performed. This analysis looked at the average absolute difference between the prediction and the true answer in the prediction condition, for facts that were either correct or incorrect in the test phase. Condition was significant (F (1, 48) = 11.925, p = 0.001; see figure 3), meaning that the facts answered correctly in the test phase were characterized by a *larger* difference between prediction and true answer in the study phase (compared to facts answered incorrectly in the test phase). Therefore, a larger prediction error seemed to lead to a better memory for the correct answer, which is in line with our hypotheses. There was also a main effect of age (F (1, 48) = 9.566, p = 0.033), indicating that adults had a significantly smaller difference between prediction and true answer in the study phase than children, for both correct and incorrect facts in the test phase. However, interaction with age was not significant (F (1, 48) = 1.980, p = 0.166), which suggests that children and adults did not significantly differ in how much they benefited from generating incorrect predictions.

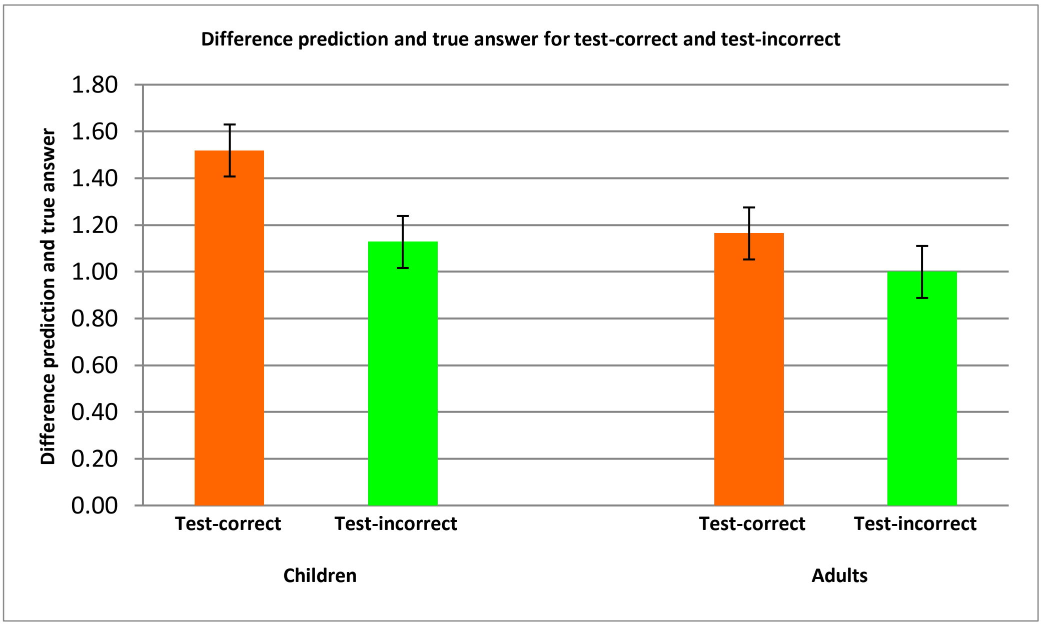

Figure 3. Graph of participants' performance in the test phase. The orange bar represents the difference between the prediction and the true answer in the study phase, for facts that were answered correctly in the test phase. The green bar represents the difference between the prediction and the true answer, for facts that were answered incorrectly in the test phase. Together, they indicate that facts answered correctly in the test phase, were characterized by a larger difference between the prediction and the correct answer in the prediction condition (i.e. a larger prediction error). This was the case for both age groups.

Perseverative errors. Next, it was investigated whether prediction errors from the study phase persevered in the test phase (i.e., remembering the incorrect prediction instead of the true answer on the final test), and if there was an effect of age and level of executive functioning on the number of perseverative errors. The average proportion of perseverative errors on the experimental task was very low (M = 0.05, SD = 0.293 for all participants).

To see if age and the proportion of perseverative errors on the BCST were predictors of the proportion of corrected errors on the experimental task (i.e. facts that were answered incorrectly in the prediction condition of the study phase, but were answered correctly in the subsequent test phase), a hierarchical regression analysis was carried out. Age was significant (t = 3.426, p = 0.001, 95% CI [0.261;1.001]). When adding the executive function of cognitive flexibility (measured by the BCST as proportion of perseverative errors), age remained significant (t = 2.468, p = 0.017, 95% CI [0.097;0.953]). Proportion of perseverative errors on the BCST was not significant, however (t = -0.998, p = 0.323. 95% CI [-0.946;0.319]). As a result, the R squared change from 0.196 to 0.017 was not significant either (p = 0.323).

Another hierarchical regression analysis was done with age and the two executive functions of inhibitory skills and cognitive flexibility (this time measured by the HFT), to see if age and level of executive functioning influenced the proportion of corrected errors (i.e., how much a person had learned) on the experimental task. When adding inhibitory skills and cognitive flexibility to the model with age, age did not remain significant (t = 1.879, p = 0.067, 95% CI [-0.034; 1.002]). Both executive functions were not significant either (t = -0.524, p = 0.603, 95% CI [-0.042; 0.025] for inhibitory skills, and t = -0.172, p = 0.865, 95% CI [-0.024; 0.021] for cognitive flexibility). The R squared change from 0.196 to 0.012 was also not significant (p = 0.707). Interestingly, though, when looking at inhibitory skills and cognitive flexibility separately, both were significant predictors of the proportion of corrected errors (respectively t = -2.822, p = 0.007, 95% CI [-0.027;-0.005] and t = -2.520, p = 0.015, 95% CI [-0.018;-0.002]). A plausible explanation is that the independent variables were highly correlated with each other (r = 0.785, p < 0.001). However, the VIF of 2.605 remained under 10 for both variables. When carrying out regression analyses with age and one of the executive functions, both remained insignificant (first analysis: t = 1.974, p = 0.054 for age and t = -0.827, p = 0.412 for inhibitory skills; second analysis: t = 2.266, p = 0.028 for age and t = 0.657, p = 0.515 for cognitive flexibility). Another explanation is that both inhibitory skills and cognitive flexibility were highly correlated with age (respectively r = 0.672, p < 0.001 and r = 0.620, p < 0.001). Still, the VIFs remained under 10 (3.047 for age and inhibitory skills, and 2.715 for age and cognitive flexibility). Furthermore, since age was significant or almost significant in both analyses, it seems that this mostly explained the variance in the proportion of corrected errors on the experimental task.

Subsequently, it was investigated whether age and level of executive functioning could predict the proportion of perseverative errors on the experimental task. First, a hierarchical linear regression analysis looked at the question whether age and cognitive flexibility (measured by the proportion of perseverative errors on the BCST) could predict the proportion of perseverative errors on the experimental task. Age was not a significant predictor (t = -0.187, p = 0.852, 95% CI [-0.122; -0.147]). When adding proportion of perseverative errors on the BCST as a predictor to this model, the results remained insignificant (for age, t = 0.203, p = 0.840, 95% CI [-0.141;0.173]; for proportion of perseverative errors on the BCST, t = 0.086, p = 0.932, 95% CI [-0.222; 0.242]). Logically, the R squared change from 0.001 to 0.000 was insignificant as well (p = 0.932). In a second hierarchical regression analysis, inhibitory skills and cognitive flexibility (as measured by the HFT) were added to the model with age as a predictor of the proportion of perseverative errors on the experimental task. Age remained insignificant (t = 0.699, p = 0.488, 95% CI [-0.123; 0.253]). Inhibitory skills and cognitive flexibility were not significant either (respectively t = 0.743, p = 0.461, 95% CI [-0.008;0.017] and t = -0.091, p = 0.928, 95% CI [-0.009;0.008]). The R squared change from 0.001 to 0.017 was also not significant (p = 0.674). Therefore, according to these analyses, it seems that neither executive functions nor age had a significant influence on the proportion of perseverative errors a person made in the test phase of the experimental task. Interestingly, however, age *was* a significant predictor of the proportion of perseverative errors on the BCST (t = -3.997, p < 0.001, 95% CI [-0.508; -0.168]), with a higher age relating to fewer perseverative errors. It was also a significant predictor of inhibitory skills and cognitive flexibility on the HFT (respectively t = -6.280, p <0.001, 95% CI [-17.393;-8.957] and: t = 5.417, p < 0.001, 95% CI [-23.392; -10.819]), with a higher age relating to faster RTs.

Depth of processing. To examine whether it is more likely that participants recall the correct answer in the test phase when it takes them longer to predict this answer in the study phase (meaning their depth of processing is larger), a repeated measures ANOVA was carried out. First, it looked at the average RT in the prediction condition of the study phase for the facts that were correct or incorrect in the test phase. RT was significantly longer for facts that were incorrect during the test phase than for facts that were correct during the test phase (F (1, 48) = 6.634, p = 0.015). In addition, there was a main effect of age (F (1, 48) = 11.425, p = 0.003): adults had faster RTs than children. However, there was no significant interaction with age (F (1, 48) = 0.084, p = 0.773). Subsequently, the analysis looked at the average RT in the example condition of the study phase for facts that were correct or incorrect in the test phase. A repeated measures ANOVA demonstrated that there was no significant effect of RT on number of facts correct (F (1,48) = 0.115, p = 0.736). Still, there was a main effect of age (F (1, 48) = 16.328, p < 0.001), with – again – adults having faster RTs than children. Yet, the interaction with age was not significant (F (1,48) = 0.096, p = 0.758). So, for generating examples, depth of processing in the study phase did not seem to influence memory performance in the study phase.

Transformation of variables. To check whether the findings did not result from the non-normal distributions of the variables, logistic transformations were carried out. These transformations yielded no different results.

Discussion

Recently a report from the Inspectorate of Education (2020) was released, identifying features of constructivism as contributing to the strength of schools. This study sought to investigate the underlying mechanisms of two constructivism based GLS, prediction generation and example generation, thereby contributing to the debate about the most effective kind of instruction. In order to do so, this study used data from Breitwieser and Brod (2020), who found an interaction effect between age and GLS. Prediction generation was more beneficial than example generation for primary school children, whereas both GLS were equally effective in adults. The current study examined whether memory performance on the final test was affected by prediction error size (as indicated by the difference between the predicted answer and the true answer) and depth of processing (as indicated by the RT) in the study phase, and if this differed by age. In addition, it investigated whether there was an effect of age and level of executive functioning on the number of perseverative errors.

Conclusion and discussion of results

Replication analyses. Surprisingly, this study did not find the interaction between condition and age that Breitwieser and Brod (2020) found when looking at the proportion of correct facts, even though the same data set was used. There was a main effect of condition, but no interaction with age, meaning that generating predictions was more beneficial in both children and adults. Still, post-hoc analyses revealed that the benefits of generating predictions were significant in children but not in adults. In addition, a second analysis *did* find a significant interaction, now looking at the average absolute difference between test answer and true answer. Both age and condition were significant as well. This means that children seem to benefit relatively more from prediction generation, compared to example generation, than adults.

An explanation for why the initial results differed from Breitwieser and Brod (2020) is that the current study used a repeated measures ANOVA for the analyses, whereas they used a logistic mixed-effects model. The latter is more sensitive and therefore could detect the differences between adults and children more easily. Another difference is that in this study, it was decided to leave the facts with no given example in the dataset, whereas Breitwieser and Brod (2020) deleted them. Since children and adults significantly varied in how many facts they left blank in the example condition, this could have lead to different results.

Prediction errors. Conform the hypothesis, the current study found a larger difference between prediction and true answer in the study phase, for facts that were answered correctly in the test phase. In other words, a larger prediction error seemed to stimulate memory. There was no interaction with age, suggesting that both groups benefited more from larger prediction errors. All in all, prediction errors seem to lead to a better memory; something that is in line with previous literature (Breitwieser & Brod, 2020; Brod et al., 2018; Henson & Gagnepain, 2010; Huelser & Metcalfe, 2012; Potts, Davies, & Shanks, 2019; Potts & Shanks, 2014). Apparently, memory systems – most likely the MTL with the hippocampus, and the striatum – are triggered by prediction errors, therefore remembering more than when learning in a passive situation. The reason for why larger prediction errors lead to greater memory effects than smaller prediction errors, could be that larger prediction errors cause a larger surprise response. After all, the given prediction was far more different than the correct answer. The surprise response, in turn, is known to trigger memory systems, with a stronger response leading to greater learning effects (Brod et al., 2018). Future neuroscientific research could investigate how exactly the different memory networks of the MTL and the striatum play a role when making a prediction error, and how they respond to a large prediction error compared to a small one.

It is important to note that there is a difference between the prediction error as this study measured it, and the physiological prediction error as described by Henson and Gagnepain (2010). Whereas this study defined the prediction error as the distance between the incorrect answer and the true answer, Henson and Gagnepain (2010) operationalized it as the difference between top-down predictions from, for example, the hippocampus, and bottom-up sensory input. For future research, it is interesting to investigate whether this neural prediction error and the behavioral prediction error as measured in this study, are comparable.

Perseverative errors. With respect to perseverative errors, the current study found a relatively small amount of perseverative errors, and quite a large amount of corrected errors. As discussed, frequently heard criticism on prediction generation – or constructivist learning in general – is that children will remember their incorrect guess instead of the correct answer. This study demonstrates the opposite: children more often corrected their initial incorrect guess, than that they kept making the same mistake.

It was hypothesized that participants with lower levels of executive functioning would make more perseverative errors and fewer corrected errors than persons with higher levels. This effect was not found for the experimental task. It was also hypothesized that children would make more perseverative errors than adults, because their executive functions had not fully developed yet. Age was a significant predictor of the proportion of perseverative errors on the BCST, and it also significantly predicted the level of inhibitory skills and cognitive flexibility. Yet, age was not a significant predictor of the proportion of perseverative errors on the experimental task.

Regarding corrected errors, on the other hand, age was a significant predictor: the higher age, the higher the proportion of corrected errors on the experimental task. Consequently, it was also found that inhibitory skills and cognitive flexibility had a significant effect on the proportion of corrected errors, with lower levels of these executive functions leading to fewer corrected errors on the experimental task. Be that as it may, when inhibitory skills and cognitive flexibility were put together in one model, neither was a significant predictor for the proportion of corrected errors on the experimental task. This did not seem to be due to the great overlap between the two predictors, so the reason for these results remains an issue for further research.

Depth of processing. It was hypothesized that memory performance on the final test would improve when RTs in the study phase were longer. The results demonstrated the opposite: facts that were incorrect in the test phase were characterized by a significantly longer RT in the prediction condition of the study phase. When looking at RTs in the example condition of the study phase, there was no significant correlation with final test performance.

These results are not in line with Benjamin et al. (1998), who found that participants with a longer RT remembered more facts correctly than participants with a shorter RT. One factor that could play a role in these differences is that Benjamin et al. (1998) used a free-recall task, whereas this study asked participants to remember every answer to every fact. That means the current participants had a higher chance of answering a question incorrectly in the test phase, since they needed to answer all the facts; even the facts they did not remember anymore. Participants from Benjamin et al. (1998) only needed to answer the facts they – correctly or incorrectly – remembered. It is possible they mostly remembered the facts they had thought about for a longer time, but that does not imply that the facts they did not remember (i.e. the facts with a shorter RT), were answered incorrectly. In addition, Benjamin et al. (1998) used a between-subjects design, looking at participants with a short or long RT, whereas this study had a within-subjects design, looking at individual facts with a short or long RT. Lastly, an explanation could be that a deeper processing (i.e. a longer RT) did not lead to more incorrect answers, but that a deeper processing still took place, *despite* more incorrect answers. After all, it is quite likely that the facts participants thought about for a longer time, were also less familiar. After a deep processing of this fact, the possibility of an incorrect answer would still be high.

The results of the current study partly correspond with the study of Huelser and Metcalfe (2012), who used a classic generation model with three conditions: read-short (study cue and target for

5 seconds), read-long (study cue and target for 10 seconds) and error-generation (generate a response before receiving the true answer). After this study condition, there was a final test. Results were that for all conditions, RTs for correct responses were shorter than RTs for incorrect responses on the final test. In their second experiment, they found the same results specifically for the items that were learned in the error-generation condition (Huelser & Metcalfe, 2012). One must keep in mind, however, that this study did not look at RTs during the test phase, but at RTs during the study phase. Still, the study of Huelser and Metcalfe (2012) underlines the possibility of a shorter RT leading to a correct response.

Limitations

There are some factors that have not been measured in this study but are likely to have an effect on the process of generating predictions. First of all, when predicting numerical facts, it could make a difference whether somebody is able to generate numerical predictions. In other words, the level of ‘number sense’ somebody possesses could influence the quality of the predictions and the amount of learning that takes place. Number sense is defined in many ways (Berch, 2005), one of the more broader definitions being “the ability to use numbers and quantitative methods as a means of communicating, processing and interpreting information” (McIntosh, Reys, & Reys, 1992). Estimation is one aspect of number sense and was found to be influenced by the ability to use number relations and understand the relative sizes of large numbers (Pike & Forrester, 1997). Although estimation and prediction are not the same, their processes of guessing a certain number seem alike. Therefore, it is possible that number sense affects the generation of numerical predictions. To expand our understanding about the effect of generating predictions on learning, future research should look into the effects of number sense.

Working memory also could have had an influence on the results. As is known, working memory holds information in mind for a short period of time. It is, as Morrison and Chein (2011) describe, a “flexible, capacity limited, mental workspace that is used to store and process information in the service of ongoing cognition” (p. 47). Participants’ different working memory capacities could have influenced the degree to which participants succeeded in comparing their own prediction with the true answer. After all, Kirschner et al. (2006) argue that constructivist learning poses a load that is too heavy on children’s working memory. Future research should investigate whether this is the case.

In addition, Breitwieser and Brod (2020) mentioned a few limitations of their research that are relevant for the current study, too. First, there was no control condition, meaning that the measurements could not be compared with a baseline performance of participants not being stimulated to use certain learning strategies. Therefore, one could question whether learning from prediction or example generation is more effective than baseline performance. However, Breitwieser and Brod (2020) explain why they explicitly decided to not incorporate a control condition: it would be too difficult to compare baseline performances between the age groups, since adults – or university students – are more likely to use GLS without being prompted to (Justice & Weaver-McDougall, 1989). Children, on the other hand, are not expected to use these GLS spontaneously, and are characterized by large inter-individual differences (Bjorklund, 2010). Second, Breitwieser and Brod (2020) mention that the sample size was limited; large enough for studying the age x condition interaction, but too small for reliable between-subject analyses like the regression analyses regarding executive functions. Their study and the current replication can be seen as exploratory analyses that pave the way for further research into the different age-related mechanisms. A final limitation is that the long-term effects of prediction and example generation were not investigated (Brod & Breitwieser, 2020). Consequently, is also not known whether prediction errors persevere in the long term. Metcalfe and Miele (2014) stated that high-confidence errors persevered more often when they were not tested immediately after participants realized their answer was wrong. In other words: participants reverted more often to their incorrect prediction (made with high confidence) as the time gap between their initial answer with corrective feedback (i.e. the study phase) and the final test (i.e. the test phase) became larger. This study did not test the prediction error immediately after participants realized they gave a wrong answer; instead, a next fact was presented. Still, the final test was a relatively quick follow-up of the study phase. Perhaps, when the test would have been delayed for a longer period, high-confidence errors would have persevered after all.

Implications and future directions

The present study adds a new perspective to the debate that centres around the question which kind of instruction is the most appropriate for children in primary school. It gives evidence for the benefits of some aspects of constructivist learning, by indicating that generating information from one's own mind is an effective learning strategy. Specifically, generating incorrect predictions and receiving corrective feedback on them seems to be beneficial for children's learning. In fact, it did not seem to lead to perseverative errors, something that is often mentioned as criticism on constructivist learning. This could indicate that the best instruction includes segments from both constructivist learning (generating predictions) and direct instruction (receiving corrective feedback).

Furthermore, the current study provides new, useful insights for future brain research. Since it has found new evidence for the effectiveness of prediction errors in children, it would be interesting to investigate what happens in their brains when generating predictions, and if and in which way this differs from adults. It is conceivable that children's brain activity regarding prediction errors differs from adults' activity, since the connection between the striatum and the prefrontal cortex has not yet fully matured in children (Decker et al., 2015). This would mean that when making predictions and the accessory errors, the striatum of children would be more active without the prefrontal cortex being involved, whereas in adults, the prefrontal cortex would be actively controlling the striatum. Understanding these underlying neural mechanisms of prediction generation in children makes it possible to tie the educational instruction more appropriately to their needs.

Another direction for future brain research could be the question whether the behavioral and the neural prediction error are related. It is plausible that in some situations, they differ; for example, the behavioral prediction error could be quite large (i.e. there is large distance between the prediction and the correct answer), while at the same time, the neural prediction error could be much smaller, because participants did not know anything about the subject so they randomly guessed something. Random guessing could lead to a smaller neural prediction error, because there was no expectation about the possible answer. The striatum responds to prediction errors in terms of whether the outcome was better or worse than expected (Van den Bos et al., 2012). When there are hardly any expectations, the subsequent striatal reward signals could be much smaller. In addition, one might ask whether participants would even remember their randomly guessed answer long enough to compare it to the correct answer. Working memory has to select information in order to maintain effective, and in doing so it is influenced by stimulus salience and reward expectations (Klink, Jeurissen, Theeuwes, Denys, & Roelfsema, 2017). During random guessing, presumably neither stimulus salience nor reward expectations are high. As a consequence, the neural prediction error will probably not be high either. In conclusion, future brain research should investigate this specific prediction error in relation to the behavioral prediction error.

Finally, it is important to note that this study does not implicate that generating predictions is the best way of learning. Fiorella and Mayer (2016) sum up more GLS than prediction (and example) generation alone, that should be investigated before such a claim could be made. Still, there are already successful curricula that implement prediction generation with receiving corrective feedback in physics (Champagne, Klopfer, & Anderson, 1980; Gunstone & White, 1981; Liew & Treagust, 1995). In addition, there is an interactive computer program that uses predictions with corrective feedback to teach students about genetics (Tsui & Treagust, 2003). The current study could stimulate the implementation of the prediction-feedback cycle in more areas, and specifically in subjects relevant for primary school. Since prediction has already been proven to be successful in geography (Brod et al., 2018) and trivia fact learning (Brod & Breitwieser, 2019), here lay opportunities to incorporate generating predictions in the instruction. However, it remains a question in which subjects generating predictions is effective in children. The studies of Brod et al. (2018) and Brod and Breitwieser (2019) were done with adults and although the present study was done with children, it did not teach them a specific subject. Therefore, future research should focus on the effects of prediction generation in more specific subjects of primary school.

Conclusion

In conclusion, the current study demonstrates the effectiveness of generating predictions in primary school children. Breitwieser and Brod (2020) found that generating predictions was more effective than generating examples in children, whereas both were equally effective in adults. This study found a similar trend: generating predictions lead to more correct answers on the final test than generating examples, and this effect was significantly stronger in children than in adults. In addition, this study discovered a possible reason for this effect: generating predictions involves making prediction errors, and these seem beneficial for learning. That is, a larger prediction error was related to a better memory for the correct answer. These results provide exciting opportunities for education and research in the future.

References

Anthony, W. (1973). Learning to discover rules by discovery. *Journal of Educational Psychology, 64*(3), 325-328.

Aston-Jones, G., & Cohen, J. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive Gain and optimal performance. *Annual Review of Neuroscience, 28*(1), 403-450. doi:http://doi.org/10.1146/annurev.neuro.28.061604.135709

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11*(7), 280-289. doi:https://doi-org.ezproxy.leidenuniv.nl:2443/10.1016/j.tics.2007.05.005

Bar, M., Aminoff, E., Mason, M., & Fenske, M. (2007). The Units Of Thought. *Hippocampus*, 420-428.

Barrett, K., Fox, N., Morgan, G., Fidler, D., & Daunhauer, L. (2013). *Handbook of Self-Regulatory Processes in Development: New Directions and International Perspectives.* London: Psychology Press.

Benjamin, A., Bjork, R., & Schwarz, B. (1998). The mismeasure of memory: when retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*(1), 55-68.

Berch, D. (2005). Making Sense of Number Sense. Implications for Children With Mathematical Disabilities. *Journal of Learning Disabilities, 38*(4), 333-339.

Bertsch, S., Pesta, B., Wiscott, R., & McDaniel, M. (2007). The generation effect: A meta-analytic review. *Memory & Cognition, 35*(2), 201-210.

Bjorklund, D. (2010). Remembering on their own: The development of strategic memory. In H. Trolldenier, W. Lenhard, & P. & Marx, *Das Gedächtnis entwickeln: Entwicklungs- und pädagogisch-psychologische Forschungen zum Gedächtnis (The development of memory: Research on memory in the fields of developmental and educational psychology)* (pp. 171-189). Göttingen, Germany: Hogrefe Verlag.

Bransford, B., & Johnson, M. (1972). Contextual Prerequisites for Understanding: Some Investigations of Comprehension and Recall. *Journal of Verbal Learning and Verbal Memory, 11*(6), 717-726. doi:https://doi.org/10.1016/S0022-5371(72)80006-9

Breitwieser, J., & Brod, G. (in press). Cognitive prerequisites for generative learning: Why some learning strategies are more effective than others. *Child Development*.

Brocki, K., & Tillman, C. (2014). Mental Set Shifting in Childhood: The Role of Working Memory and Inhibitory Control. *Infant and Child Development, 23,* 588-604. doi:10.1002/icd.1871

Brod, G., & Breitwieser, J. (2019). Lighting the wick in the candle of learning: generating a prediction stimulates curiosity. *NPJ Science of Learning, 4*(1). doi:https://doi.org/10.1038/s41539-019-0056-y

Brod, G., Breitwieser, J., Hasselhorn, M., & Bunge, S. (2019). Being proven wrong elicits learning in children – but only in those with higher executive function skills. *Developmental Science, 23*(3), 1-14. doi:10.1111/desc.12916

Brod, G., Bunge, S. A., & Shing, Y. L. (2017). Does one year of schooling improve children's cognitive control and alter associated brain activation? *Psychological Science, 28*, 967-978. doi:https://doi.org/10.1177/0956797617699838

Brod, G., Hasselhorn, M., & Bunge, S. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction, 55*, 22-31. doi:https://doi.org/10.1016/j.learninstruc.2018.01.013

Brunstein, A., Betts, S., & Anderson, J. (2009). Practice Enables Successful Learning Under Minimal Guidance. *Journal of Educational Psychology, 101*(4), 790-802. doi:10.1037/a0016656

Bunge, S., Dudukovic, N., Thomason, M., & Vaidya, C. G. (2002). Immature Frontal Lobe Contributions to Cognitive Control in Children: Evidence from fMRI. *Neuron, 33*(2), 301-311.

Butler, A., Fazio, L., & Marsh, E. (2011). The hypercorrection effect persists over a week, but high-confidence errors return. *Psychonomic Bulletin & Review, 18*, 1238-1244. doi:https://doi.org/10.3758/s13423-011-0173-y

Butterfield, B., & Metcalfe, J. (2001). Errors Committed With High Confidence Are Hypercorrected. *Journal of Experimental Psychology, 24*(7), 1491-1494. doi:10.1037//0278-7393.27.6.1491

Champagne, A., Klopfer, L., & Anderson, J. (1980). Factors influencing the learning of classical mechanics. *American Journal of Physics, 48*(12), 1074. doi:http://doi.org/10.1119/1.12290

Cohen, J., Asarnow, R., Sabb, F., Bilder, R., Bookheimer, S., Knowlton, B., & Poldrack, R. (2010). A unique adolescent response to reward prediction errors. *Nature Neuroscience, 13*(6), 669-671. doi:http://www.nature.com/doifinder/10.1038/nn.2558

Crouch, C., Fagen, A., Callan, J., & Mazur, E. (2004). Classroom demonstrations: Learning tools or entertainment? *American Journal of Physics*, *72*(6), 835-838. doi:10.1119/1.1707018

Davidson, M., Amso, D., Anderson, L., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*(11), 2037-2078. doi:10.1016/j.neuropsychologia.2006.02.006

Decker, J., Lourenco, F., Doll, B., & Hartley, C. (2015). Experiential reward learning outweighs instruction prior to adulthood. *Cognitive, Affective, & Behavioral Neuroscience, 15*(2), 310-320. doi:10.3758/s13415-014-0332-5

Delgado, M., & Dickerson, K. (2012). Reward-Related Learning via Multiple Memory Systems. *Biological Psychiatry, 72*, 134-141. doi:10.1016/j.biopsych.2012.01.023

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology, 64*(1), 135-168. doi:https://doi.org/10.1146/annurev-psych-113011-143750

Diamond, A., Barnett, W., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*, 1387-1388. doi:10.1126/science.1151148

Dickerson, K., Li, J., & Delgado, M. (2011). Parallel contributions of distinct human memory systems during probabilistic learning. *NeuroImage, 55*, 266-276. doi:10.1016/j.neuroimage.2010.10.080

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29*, 153-170. doi:10.1016/j.learninstruc.2012.05.003

Doll, B., Jacobs, W., Sanfey, A., & Frank, M. (2009). Instructional control of reinforcement learning: A behavioral and neurocomputational investigation. *Brain Research, 1299*, 74-94. doi:http://dx.doi.org/10.1016/j.brainres.2009.07.007

Doyon, J., Bellec, P., Amsel, R., Penhune, V., Monchi, O., Carrier, J., . . . Benali, H. (2009). Contributions of the basal ganglia and functionally related brain structures to motor learning. *Behavioural Brain Research, 199*, 61-75. doi:https://doi.org/10.1016/j.bbr.2008.11.012

Duit, R. (1991). On the role of analogies and metaphors in learning science. *Science Education, 75*, 649-672. doi:https://doi.org/10.1002/sce.3730750606

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

Ekman, P. (1992). Are there basic emotions? *Cognition & Emotion, 99*(3), 550-553. doi:http://doi.org/10.1080/02699939208411068

Fazio, L., & Marsh, E. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin & Review, 16*(1), 88-92. doi:10.3758/PBR.16.1.88

Filoteo, J., Maddox, W., & Davis, J. (2001). A possible role of the striatum in linear and nonlinear category learning: Evidence from patients with Huntington's disease. *Behavioral Neuroscience, 115*(4), 786-798. doi:https://psycnet.apa.org/doi/10.1037/0735-7044.115.4.786

Fiorella, R., & Mayer, R. (2016). Eight Ways to Promote Generative Learning. *Educational Psychology Review, 28*, 717-741. doi:10.1007/s10648-015-9348-9

Fox, C. J., Mueller, S. T., Gray, H. M., Raber, J., & Piper, B. J. (2013). Evaluation of a short-form of the Berg Card Sorting Test. *PLoS One, 8*, 6-9. doi:https://doi.org/10.1371/journal.pone.0063885

Freedberg, M., Toader, A., Wassermann, E., & Voss, J. (2020). Competitive and cooperative interactions between medial temporal and striatal learning systems. *Neuropsychologia, 136*, 1-13. doi:https://doi.org/10.1016/j.neuropsychologia.2019.107257

Funahashi, S. (2017). Working Memory in the Prefrontal Cortex. *Brain Sciences, 7*, 1-22. doi:10.3390/brainsci7050049

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

Gunstone, R. F., & White, R. T. (1981). Understanding of gravity. *Science Education, 65*(3), 291-299. doi:http://doi.org/10.1002/sce.3730650308

Gurlitt, J., & Renkel, A. (2008). Are high-coherent concept maps better for prior knowledge activation? Differential effects of concept mapping tasks on high school vs. university students. *Journal of Computer Assisted Learning, 24*(5), 407-419. doi:10.1111/j.1365-2729.2008.00277.x

Henson, R., & Gagnepain, P. (2010). Predictive, Interactive Multiple Memory Systems. *Hippocampus, 20*(11), 1315-1326. doi:10.1002/hipo.20857

Huelser, B., & Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory & Cognition, 40*(4), 514-527. doi:10.3758/s13421-011-0167-z

Inagaki, K., & Hatano, G. (1977). Amplification of Cognitive Motivation and Its Effects on Epistemic Observation. *American Educational Research Journal, 14*(4), 485-491.

Inspectie van het Onderwijs. (2020). *De Staat van het Onderwijs 2020.* Utrecht: Inspectie van het Onderwijs.

Justice, E. M., & Weaver-McDougall, R. G. (1989). Adults' knowledge about memory: Awareness and use of memory strategies across tasks. *Journal of Educational Psychology, 81*, 214-219. doi:https://doi.org/10.1037/0022-0663.81.2.214

Kang, M., Hsu, M., Krajbich, I., Loewenstein, G., McClure, S., Wang, J., & Camerer, C. (2009). The Wick in the Candle of Learning. Epistemic Curiosity Activates Reward Circuitry and Enhances Memory. *Psychological Science, 20*(8), 963-973.

Kirschner, P., Sweller, J., & Clark, R. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist, 41*(2), 75-86. doi:10.1207/s15326985ep4102_1

Klahr, D., & Nigam, M. (2004). The Equivalence of Learning Paths in Early Science Instruction. Effects of Direct Instruction and Discovery Learning. *Psychological Science, 15*(10), 661-667.

Klink, P., Jeurissen, D., Theeuwes, J., Denys, D., & Roelfsema, P. (2017). Working memory accuracy for multiple targets is driven by reward expectation and stimulus contrast with different time-courses. *Scientific Reports, 7*(9082), 1-13. doi:10.1038/s41598-017-08608-4

Knowlton, B., Mangels, J., & Squire, L. (1996). A neostriatal habit learning system in humans. *Science, 273*, 1399-1402. doi:https://doi.org/10.1126/science.273.5280.1399

Kolb, B., & Wishaw, I. (2015). Learning and Memory. In B. Kolb, & I. Wishaw, *Fundamentals of Human Neuropsychology* (pp. 480-514). New York: Worth Publishers.

Kolb, D. (1984). *Experiential Learning: Experience as the Source of Learning and Development.* Michigan: Prentice-Hall.

Kumaran, D., & Maguire, E. (2006). An unexpected sequence of events: mismatch detection in the human hippocampus. *PLoS Biology, 4*(12), 2372-2382. doi:10.1371/journal.pbio.0040424

Kumaran, D., & Maguire, E. (2007). Which computational mechanisms operate in the hippocampus during novelty detection? *Hippocampus, 17*, 735-748. doi:10.1002/hipo.20326

Laubach, M. (2005). Who's on first? What's on second? The time course of learning in corticostriatal systems. *Trends in Neurosciences, 28*(10), 509-511.

Lee, H., Lim, K. Y., & Grabowski, B. L. (2008). Generative learning: Principles and implications for making meaning. In J. Spector, D. Merrill, J. Van Marrienboer, & M. Driscoll, *Handbook of* *research and educational communications and technology* (pp. 111-124). New York, NY: Taylor & Francis Group.

Li, J., Delgado, M., & Phelps, E. (2011). How instructed knowledge modulates the neural systems of reward learning. *Proceedings of the National Academy of Sciences, 108,* 55-60.

Li, S., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. (2004). Transformations in the Couplings Among Intellectual Abilities and Constituent Cognitive Processes Across the Life Span. *Psychological Science, 15*(3), 155-163.

Liew, C., & Treagust, D. (1995). A Predict-Observe-Explain Teaching Sequence for Learning about Students' Understanding of Heat and Expansion Liquids. *Australian Science Teachers' Journal, 41*(1), 68-71.

Lisman, J., & Grace, A. (2005). The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron, 46*, 703-713.

Liston, C., Watts, R., Tottenham, N., Davidson, M., Niogi, S., Ulug, A., & Casey, B. (2006). Frontostriatal Microstructure Modulates Efficient Recruitment of Cognitive Control. *Cerebral Cortex, 16*(4), 553-560. doi:10.1093/cercor/bhj003

Marsh, E., Roediger, H., Bjork, R., & Bjork, E. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review, 14*(2), 194-199.

Mayer, R. (2004). Should There Be a Three-Strikes Rule Against Pure Discovery Learning? The Case for Guided Methods of Instruction. *American Psychologist, 59*(1), 4-9. doi:10.1037/0003-066X.59.1.14

McClure, S., Berns, G., & Montague, P. (2003). Temporal Prediction Errors in a Passive Learning Task Activate Human Striatum. *Neuron, 38,* 339-346.

McGaugh, J., & Roozendaal, B. (2009). Drug enhancement of memory consolidation: historical perspective and neurobiological implications. *Psychopharmacology, 202*(1), 3-14. doi:10.1007/s00213-008-1285-6

McIntosh, A., Reys, B., & Reys, R. (1992). A proposed framework for examining basic number sense. *For the Learning of Mathematics, 12*, 2-8.

Metcalfe, J., & Miele, D. (2014). Hypercorrection of high confidence errors: Prior testing both enhances delayed performance and blocks the return of the errors. *Journal of Applied Research in Memory and Cognition, 3*(3), 189-197. doi:https://doi.org/10.1016/j.jarmac.2014.04.001

Moreno, R. (2004). Decreasing Cognitive Load for Novice Students: Effects of Explanatory versus Corrective Feedback in Discovery-Based Multimedia. *Instructional Science, 32*, 99-113.

Morrison, A., & Chein, J. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review, 18*, 46-60. doi:10.3758/s13423-010-0034-0

Myers, C., Shohamy, D., Gluck, M., Grossman, S., Kluger, A., Ferris, S., . . . Schwarz, R. (2003). Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *Journal of Cognitive Neuroscience, 15*, 185-193. doi:https://doi.org/10.1162/089892903321208123

O'Doherty, J. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current Opinion in Neurobiology, 14*, 769-776. doi:10.1016/j.conb.2004.10.016

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. (2004). Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. *Science, 304*(5669), 452-454. doi:10.1126/science.1094285

Ordaz, S., Foran, W., Velanova, K., & Luna, B. (2013). Longitudinal Growth Curves of Brain Function Underlying Inhibitory Control through Adolescence. *The Journal of Neuroscience, 33*(46), 18109-18124. doi:10.1523/JNEUROSCI.1741-13.2013

Peirce, J. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods, 162*, 8-13. doi:https://doi.org/10.1016/j.jneumeth.2006.11.017

Piaget, J. (1926). *The language and thought of the child.* London: Rutledge and Keagan Paul.

Pike, C., & Forrester, M. (1997). The Influence of Number-sense on Children's Ability to Estimate Measures. *Educational Psychology, 17*(4), 483-500. doi:https://doi.org/10.1080/0144341970170408

Pinel, J., & Barnes, S. (2014). Learning, Memory and Amnesia. In J. Pinel, & S. Barnes, *Introduction to Biopsychology* (pp. 283-310). Essex: Pearson Education Limited.

Piper, B., Mueller, S., Geerken, A., Dixon, K., Kroliczak, G., Olsen, R., & Miller, J. (2015). Reliability and validity of neurobehavioral function on the Psychology Experimental Building Language test battery in young adults. *PeerJ, 3*(1460), 1-26. doi:10.7717/peerj.1460

Poldrack, R., Prabhakaran, V., Seger, C., & Gabrieli, J. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology, 13*, 564-574.

Potts, R., & Shanks, D. (2014). The Benefit of Generating Errors During Learning. *Journal of Experimental Psychology: General, 143*(2), 644-676. doi:10.1037/a0033194

Potts, R., Davies, G., & Shanks, D. (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. doi:http://dx.doi.org/10.1037/xlm0000637

Raichle, M., MacLeod, A., Snyder, A., Powers, W., Gusnard, D., & Shulman, G. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America, 98*(2), 676-682.

Ranganath, C., & D'Esposito, M. (2001). Medial temporal lobe activity associated with active maintenance of novel information. *Neuron, 31*, 865-873. doi:https://doi.org/10.1016/S0896-6273(01)00411-1

Rawley, J., & Constantinidis, C. (2009). Neural correlates of learning and working memory in the primate posterior parietal cortex. *Neurobiology of Learning and Memory, 91*(2), 129-138. doi:https://dx.doi.org/10.1016%2Fj.nlm.2008.12.006

Rawson, K., & Dunlosky, J. (2016). How Effective is Example Generation for Learning Declarative Concepts? *Educational Psychology Review, 28*, 649-672. doi:10.1007/s10648-016-9377-z

Reuter, T., Lew-Williams, C., & Borovsky, A. (2019). Predict and Redirect: Prediction Errors Support Children's Word Learning. *Developmental Psychology, 55*(8), 1656-1665. doi:http://dx.doi.org/10.1037/dev0000754

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15*, 243-257. doi:https://doi.org/10.1037/a0016496

Roediger, H., & Pyc, M. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition, 1*(4), 242-248. doi:http://dx.doi.org/10.1016/j.jarmac.2012.09.002

Rosner, Z., Elman, J., & Shimamura, A. (2013). The generation effect: Activating broad neural circuits during memory encoding. *Cortex, 49*(7), 1901-1909. doi:https://doi.org/10.1016/j.cortex.2012.09.009

Rutherford, F. (1964). The role of inquiry in science teaching. *Journal of Research in Science Teaching, 2*(2), 80-84.

Sadeh, T., Shohamy, D., Levy, D., Reggev, N., & Maril, A. (2011). Cooperation between the hippocampus and the striatum during episodic encoding. *Journal of Cognitive Neuroscience, 23*, 1597-1608.

Sağlam, Y., & Dost, Ş. (2016). A Qualitative Research on Example Generation Capabilities of University Students. *International Journal of Science and Mathematics Education, 14*, 979-996. doi:10.1007/s10763-015-9624-7

Schmidt, G., De Voider, M., De Grave, W., Moust, J., & Patel, V. (1989). Explanatory Models in the Processing of Science Text: The Role ofPrior Knowledge Activation Through Small-Group Discussion. *Journal of Educational Psychology, 81*(4), 610-619.

Schultz, F., Dayan, P., & Montague, P. (1997). A Neural Substrate of Prediction and Reward. *Science, 275*(5306), 1593-1599. doi:10.1126/science.275.5306.1593

Schultz, W., & Dickinson, A. (2000). Neuronal Coding of Prediction Errors. *Annual Review of Neuroscience, 231*(1), 473-500.

Shohamy, D., & Wagner, A. (2008). Integrating Memories in the Human Brain: Hippocampal-Midbrain Encoding of Overlapping Events. *Neuron, 60*, 378-389. doi:10.1016/j.neuron.2008.09.023

Slamecka, N., & Graf, P. (1978). The Generation Effect: Delineation of a Phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 592-604.

Squire, L., & Bayley, P. (2007). The neuroscience of remote memory. *Current Opinion in Neurobiology, 17*, 185-196. doi:http://dx.doi.org/10.1016/j.conb.2007.02.006

Stahl, A., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Cognitive Development*, *348*(6230), 91-94. doi:10.1126/science.aaa3799

Stahl, E., & Feigenson, L. (2017). Expectancy violations promote learning in young children. *Cognition, 163*, 1-14. doi:http://dx.doi.org/10.1016/j.cognition.2017.02.008

Steffe, L., & Gale, J. (1995). *Constructivism in Education.* New Jersey: Lawrence Erlbaum Associates, Inc.

Sweller, J. (2004). Instructional Design Consequences of an Analogy between Evolution by Natural Selection and Human Cognitive Architecture. *Instructional Science*, *32*(1), 9-31.

Tsui, C., & Treagust, D. (2003). Genetics Reasoning with Multiple External Representations. *Research in Science Education, 33*, 111-135. doi:http://doi.org/10.1023/A

Vallar, G. (2015). Short-Term Memory: Psychological and Neural Aspects. In J. Wright (Red.), *International Encyclopedia of the Social & Behavioral Sciences* (2nd ed., pp. 909-916). Amsterdam: Elsevier. doi:http://dx.doi.org/10.1016/B978-0-08-097086-8.51019-8

Van den Bos, W., Cohen, M., Kahnt, T., & Crone, E. (2012). Striatum--Medial Prefrontal Cortex Connectivity Predicts Developmental Changes in Reinforcement Learning. *Cerebral Cortex, 22*(6), 1247-1255. doi:10.1093/cercor/bhr198

Vaughn, K., & Rawson, K. (2012). When is guessing incorrectly better than studying for enhancing memory? *Psychonomic Bulletin & Review, 19,* 899-905. doi:10.3758/s13423-012-0276-0

Velanova, K., Wheeler, M., & Luna, B. (2008). Maturational changes in anterior cingulate and frontoparietal recruitment support the development of error processing and inhibitory control. *Cerebral Cortex, 18,* 2505-2522. doi:10.1093/cercor/bhn012

Voorn, P., Vanderschuren, L., Groenewegen, H., Robbins, T., & Pennartz, C. (2004). Putting a spin on the dorsal-ventral divide of the striatum. *Trends in Neurosciences, 27*(8), 468-474. doi:10.1016/j.tins.2004.06.006

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Wechsler, D. (2008). *Wechsler adult intelligent scale-fourth edition: Technical and interpretative manual.* San Antonio, TX: Pearson Assessment.

Wimmer, G., Braun, K., Daw, N., & Shohamy, D. (2014). Episodic Memory Encoding Interferes with Reward Learning and Decreases Striatal Prediction Errors. *The Journal of Neuroscience, 34*(45), 14901-14912. doi:10.1523/JNEUROSCI.0204-14.2014

Wittrock, M. (1974). Learning as a generative process. *Educational Psychologist, 11*(2), 87-95. doi:10.1080/00461520903433554

Wright, A., & Diamond, A. (2014). An effect of inhibitory load in children while keeping working memory load constant. *Frontiers in Psychology, 5*(213), 1-9. doi:https://doi.org/10.3389/fpsyg.2014.00213

Yin, H., & Knowlton, B. (2006). The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience, 7*, 464-476. doi:10.1038/nrn1919

Yin, H., Knowlton, B., & Balleine, B. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience, 19*(1), 181-189. doi:https://doi.org/10.1111/j.1460-9568.2004.03095.x

Yin, H., Ostlund, S., Knowlton, B., & Balleine, B. (2005). The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience, 22*(2), 213-223. doi:https://doi.org/10.1111/j.1460-9568.2005.04218.x