The comparative impact of practice-testing and learning-by-teaching on transfer among fifth and sixth graders

David Baars - 5348609

Master's Thesis

Utrecht University

Department of Social Sciences: Educational Sciences

Dr. Vincent Hoogerheide

Prof. Dr. Liesbeth Kester

June 17th, 2024

Word count: 7972

Abstract

Generative learning strategies are widely accepted and implemented by practitioners. Their effects hold high values in the educational environment but their effects on elementary school children and transfer remain limited. Moreover, teachers and students often do not know what learning strategy to implement when the aim is transfer. This study investigated whether learning-by-teaching (LBT) or practice testing (PT) was more favorable for transfer during a homework activity. 17 participants from three elementary schools in the Netherlands studied a text about the human digestive system and got placed in either the LBT, PT or restudy condition to perform a homework activity over the weekend. After the weekend, participants engaged in a post-test consisting of transfer questions. Results indicated that there was no significant difference between the conditions. Yet, mean differences were observed. However, since this study lacks substantial statistical power, no reliable inferences can be drawn from this sample size, failing to generalize to the population. Future research should consider the following aspects when replicating this study: 1) yield more participants, 2) better monitor the level of retrieval success in practice testing, and 3) increase internal consistency of the post-test.

Imagine that a teacher gives a geography homework assignment to a primary school student and lets the student decide which instructional strategy to use. The student likely chooses to use ineffective strategies such as re-study or underlining, because many students are not trained (or even told) about other effective strategies (Bjork et al., 2012). Yet one would hope that the student uses advanced strategies such as self-testing or explaining.

A wide range of generative learning strategies (i.e. GLSs) such as note-taking, teaching, concept mapping, drawing, and self-explaining all enhance learning (Klingenberg et al., 2020; Nesbit & Adesope, 2006; Rittle-Johnson, 2006; Schwamborn et al., 2010; Shrager & Mayer, 1989). GLSs apply to various types of learning outcomes such as retention, comprehension, and transfer, although, the exact effects on different learning outcomes can depend on the strategy used, age, and under conditions in which this strategy is used (Ritchie & Volkl, 2000). Practice testing (i.e. retrieval practice) is said to enhance retention due to repeated cognitive activity of the material which results in 'retention' of the material in long-term memory (Fiorella & Mayer, 2016). A meta-analysis by Adesope et al. (2017) states that practice testing is effective in all age groups. However, in their study, they plead for more research that examines testing effects with transfer (i.e., the ability to recall information and apply that in a new learning situation; Dori & Sasson, 2013) measures, indicating that this area is still rather unknown. The effects of GLSs on the learning outcome of transfer have been investigated by several researchers (e.g. Fiorella & Zhang, 2018; Klingenberg et al., 2020; Rittle-Johnson, 2006) and it could be argued that this is one of the main goals of education and learning (Hajian, 2019; Klausmeier, 1961; Rohrer et al., 2010). Therefore, investigating the effectiveness of GLSs for the learning outcome of transfer is of relevance for both learners and educators.

Because most research has focused on older students, it is an open question whether GLS helps children learn and which GLS children are best to use (Brod, 2021). While some research has contrasted the effects of different GLS on lower-order learning outcomes such as retention, it is particularly unclear which strategies foster transfer. Transfer is incredibly important because this helps prepare for future learning and is essential for lifelong learning and education (Dori & Sasson, 2013).

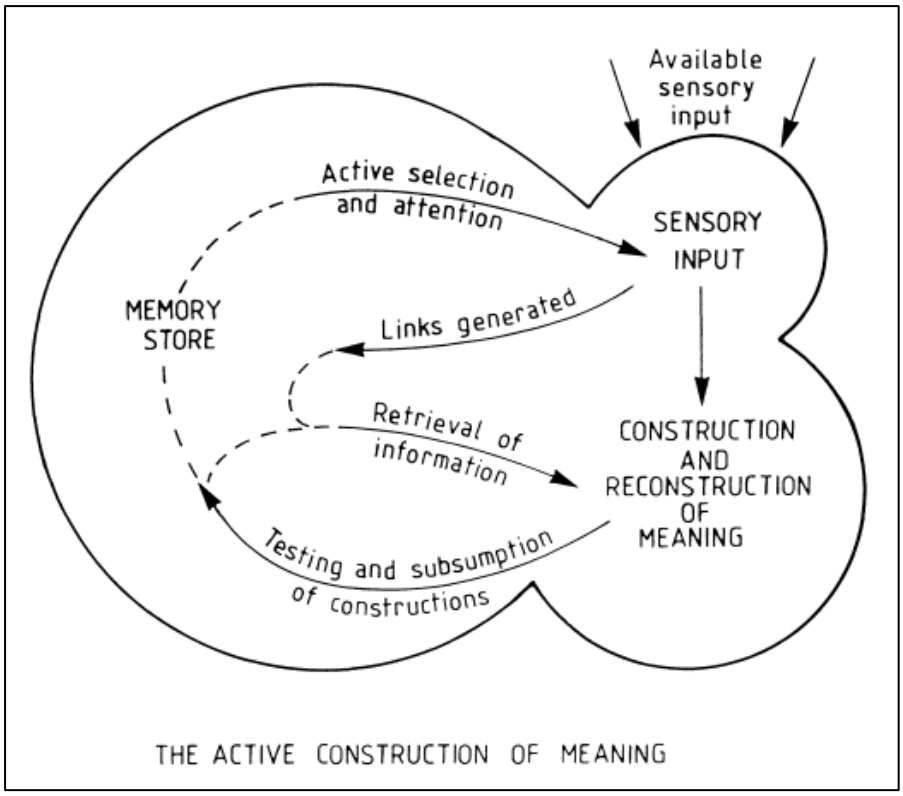
Moreover, Fiorella and Mayer (2016) also aim for more research on GLSs on outcomes such as transfer. Therefore, this study examines the effects of two popular and promising GLS on the learning outcome of transfer, namely: Practice Testing (i.e. PT) and Learning-by-Teaching (i.e. LBT).

Theoretical Framework

The generative learning theory was developed in the 20th century by Wittrock (1974a,b) and considered areas of cognitive development, human learning, human abilities, information processing, and aptitude treatment interactions (Lee et al., 2008). The generative learning theory is a cognitive theory that emphasizes the active and constructive nature of learning. It suggests that learners are active information seekers and generators of schemes and ideas based on existing knowledge and experiences, by organizing materials in a meaningful manner (Grabowski, 2013).

Figure 1

Benefits of Generative Learning Theory from a Cognitive Perspective



Osborne and Wittrock (1985) developed a schematic representation of their generative learning model. As seen in Figure 1; information that is being given attention to is selected in the sensory memory and influenced by prior knowledge. Then, links are generated between stimuli and other aspects of the core memory using retrieval. These links and information from the sensory memory are used to actively construct meaning. The constructed meaning may then be tested against other aspects and meanings of the memory store because of sensory input. If this new knowledge makes sense in terms of evaluation with other aspects of the memory store, it can then be incorporated into long-term memory.

This research includes the GLS LBT where students explain content to peers, stimulating cognitive processes (Lachner et al., 2022), and PT, which enhances understanding by recalling information from memory (Karpicke & Blunt, 2011). This study focuses on GLSs in a homework context; therefore, it will be worth mentioning how this may influence these GLSs.

Transfer

Fiorella and Mayer (2016) propose that increased cognitive engagement fosters meaningful learning outcomes like comprehension and transfer. Transfer is defined as the extent to which learning of a response in one task or situation influences the response in another task or situation (Adams, 1987). Dori and Sasson (2013) analyzed various forms of transfer and highlighted three main attributes of transfer that are similar in each form of transfer: a) task distance (i.e. level of similarity or difference from previous task), b) interdisciplinarity (i.e. context, domains, or disciplines), and c) skill set (i.e. accounts for thinking skills that the task requires). Blume et al (2009) mention in their article two different forms of transfer tasks, namely: near transfer and far transfer. Near transfer tasks refer to situations where the learning situation is highly similar to the learning task (i.e. high interdisciplinarity). Far transfer tasks refer to situations in which the learning situation is quite different from the learning task (i.e. low interdisciplinarity).

Kintsch (1994) mentions in his article that learning from a text implies the learner to use the information in other, novel ways, not just for reproduction. And that learning is defined in a way that

requires a deep understanding of the subject matter so that it can be used in novel environments (i.e. transferring knowledge). Research on the learning outcome of transfer often targets students and adults (e.g. Chmielewski & Dansereau, 1997; Chularut & DeBacker, 2004; Klingenberg et al., 2020; Leopold & Mayer, 2015) and does not always examine effects using texts.

Learning-by-Teaching and the Effects on Transfer

Learning-by-Teaching involves explaining study material to others, beneficial for the teacher (Bargh & Schul, 1980; Lachner et al., 2022). It can be interactive or non-interactive, like teaching an imaginary student or creating an instructional video. Interactive teaching involves preparation, presentation, and answering questions (Bargh & Schul, 1980; Fiorella & Mayer, 2016). Preparing entails selecting and organizing relevant information for clarity. Teaching to a peer enhances understanding for both parties (Duran & Topping, 2017). Research primarily focuses on students or adults (Fiorella & Mayer, 2013; Muis et al., 2016; Roscoe & Chi, 2007). Klingenberg et al. (2020) found significant improvements in transfer and retention for university students through teaching to others. Age-related differences in the effectiveness of this GLS might be explained by the maturation of the brain (Li et al., 2004). Li et al. argue that the brain of a child is still rather undifferentiated (i.e. components of the brain are still rather generic, like intelligence and ability) and becomes more differentiated (i.e. components develop into more specific skills or capacities, resulting in a more specialized structure of the brain) as we age. This will lead to a multifaceted ability structure. Thus, providing complete and accurate explanations might be difficult for young learners, meaning that LBT might not lead to the same beneficial results as would be the case for students and adults. However, a study by Hoogerheide et al. (2019) suggests that LBT can work for children. In their study, they examined the effects of generating an instructional video and summarizing comprehension through a homework activity. Since both conditions are comprised of effective GLS, the generating an instructional video condition still outperformed the restudying condition which can be explained by the social presence theory. This theory states that the effectiveness of LBT depends on the level of awareness of social presence (Hoogerheide et al., 2016). Presenting to an audience might be perceived

as stressful, meaning that LBT without an audience is effective for learning. However, Hoogerheide et al. examined comprehension rather than transfer.

Non-interactive teaching, despite being a recent area of research, yields effective results, including transfer (Lachner et al., 2022). Its effectiveness is attributed to retrieval practice, generative processing, and the social presence theory. However, studies on non-interactive teaching's impact on transfer in children are scarce.

Practice Testing and the Effects on Transfer

Practice testing involves assessing learners' knowledge through methods like practice questions or free recall to enhance retrieval from long-term memory to working memory (Fiorella & Mayer, 2016). Practice testing is effective since it mimics cognitive processes needed for the actual performance assessment (Adesope et al., 2017). Retrieval success relies on the number of cues associated with the information (Karpicke et al., 2014); stronger associated cues (e.g. Toast -> Bread) facilitate easier retrieval than weakly associated cues (e.g. Basket -> Bread) (Yang et al., 2019). For instance, recalling information in a specific location demonstrates cue association with location.

The practice testing effect is found to be effective for learning text passages (e.g., Roediger & Karpicke, 2006), rote learning (e.g. Carpenter & DeLosh, 2006), and nonverbal materials (e.g., Carpenter & Pashler, 2007). Moreover, Karpicke et al. (2014) found results that highlight the importance of guiding retrieval practices in elementary school. The effects of retrieval practice on elementary school children have been investigated by Karpicke et al. (2016) who found results that supported the benefits of retrieval practices on elementary school children (age 9-12). However, transfer was not included.

A study that did consider transfer was done by Butler (2010). He found that testing benefits transfer. The rationale for this was that retrieving information using numerous cues to access information should also be elicited during the transfer task, increasing the potential for transfer to occur. However, his study focused on university students, limiting generalization possibilities to lower

age groups (i.e. fifth and sixth-grade children). Rohrer et al. (2010) showed that testing enhances transfer, particularly in elementary school children. However, Rohrer et al. did not use free recall tests.

Free recall, common in retrieval practice, involves recalling information without prompted cues, yielding positive learning outcomes (Roediger & Karpicke, 2006b). Typically, free recall tests are comprised of wordlists that require rote learning rather than text learning (Moreira et al., 2019). However, a study by Wilson et al. (1985) found free recall to be more effective than illustrating for text learning among fourth graders, but, they assessed comprehension, not transfer. Younger individuals tend to excel at free recall compared to older adults (Rhodes et al., 2019), possibly due to challenges in event recall and declining attentional resources with age.

Comparing the Effects of Practice Testing and Learning-by-Teaching

Research shows that both GLSs are effective for several learning outcomes such as rote learning, comprehension, and transfer (Butler et al., 2010; Carpenter & DeLosh, 2006; Klingenberg et al., 2020; Rohrer et al., 2010). Learning by teaching promotes transfer (e.g. Klingenberg et al., 2020) and is suitable to promote comprehension for children (e.g. Hoogerheide et al.,2019). Practice testing enhances learning for all ages (Adesope et al., 2017) and according to Butler (2010), cues associated with the text should also be available during the transfer task. The effects of teaching to another and practicing retrieval were previously investigated by Lachner et al. (2020). In their study, it was investigated whether explaining in earlier phases of studying would be more beneficial for learning than explaining after the entire study phase. Lachner et al. also controlled for retrieval processes that may be involved in the explaining process. Results indicated that the hypothesis that explaining would be more beneficial than retrieval, was not supported. Also, transfer performance was similar across all conditions, showing low performance on transfer. Fortney (2016) examined the effects of self-explaining versus practice testing and rereading. Self-explaining refers to explaining the content to yourself, this usually happens silently and is according to Brod

(2021) similar to explaining to others since it involves similar cognitive processes. Fortney found mixed results. Indicating that the self-explanation condition outperformed the practice retrieval condition on verbatim test performance. However, no significant difference was found between these groups on inference questions. Fortney indicated that there is still a need for more research on this topic, stating that it is important to clarify conditions under which self-explaining and retrieval practices are most effective. Yet still little research has considered practice-testing and learning-by-teaching among fifth and sixth grades on transfer.

The Present Study

The purpose of the current study was to investigate to which degree two GLSs are suited for the learning outcome of transfer for fifth and sixth-grade graders. This research aims to seek results that might help teachers in their decision-making process for choosing appropriate learning strategies when the goal is transfer. Using GLSs wisely and with the right intention will not only foster the overall comprehension of materials but also help learners use learning strategies in future learning situations (e.g. college and future careers). Exploring this issue would benefit both teachers and learners, as GLSs are proven to be effective. With this research, the aim is to answer the following two research questions: 1) “What is the comparative impact of the generative learning strategies practice-testing and learning-by-teaching on transfer among fifth and sixth graders in the context of a homework activity?” and 2) “What is the effect of the generative learning strategies practice-testing and learning-by-teaching on transfer among fifth and sixth graders in the context of a homework activity compared to the control condition (restudy)?

I firstly hypothesize that learning by teaching will outperform the restudy condition. LBT is effective for children (Hoogerheide et al., 2019) and improves transfer (Klingenberg et al., 2020). Also, the effects of teaching to another involve many cognitive processes that do not apply to restudying. I secondly hypothesize that practice testing will outperform the restudy condition. Practice testing (also known as retrieval practice) requires learners to retrieve information from long-term

memory which fosters strong remembrance of material (Meyer & Logan, 2013). And, according to Butler (2010), retrieval practices are favorable over repeated studying (restudy), even for transfer. I thirdly hypothesize that the LBT condition will outperform the PT condition since teaching involves more processes than just explaining (Lachner et al., 2020). It also involves preparing the to-be-learned material to explain it, drawing inferences and elaborations, and mentally organizing the material. Thus, LBT involves more cognitive processes rather than just retrieving information from long-term memory as would be the case with PT (Bargh & Schul, 1980). Moreover, PT is said to be most effective when tests are delayed instead of immediately after learning (Rawson & Dunlosky, 2012). In the current study, participants will engage in practice testing only one hour after studying the material.

Method

Research Design and Participants

This study employed a between-subjects randomized experimental design to compare the

A desired sample size was calculated by Gpower using the following parameters: to-be-detected population effect size f = .25, significance level a = .05, and desired power level of at least 0.80. The desired sample size was calculated at 159 participants. Four (i.e. eight classes and 189 children) public primary schools in the Netherlands were contacted, one of which caters to both gifted and regular children. One school indicated not wanting to participate in the research. So, in total, three primary schools participated in the research (six classes and 142 children). Subsequently, both the fifth-grade and sixth-grade teachers at each participating school were contacted. Eventually, four teachers (four classes and 108 children) gave their consent to experiment in their classroom, the other two teachers did not respond to my request. Together with the teacher, an information letter was sent to all parents/caretakers in which they could give their consent. In total, 108 parents/caretakers were approached and only 17 gave their consent (15.74%), indicating that the consent rate is very low. The post-hoc power level was determined and was set at 0.13. This indicates a very poor power.

Participants were randomly distributed across groups to ensure that mean score and deviation in scores on all variables, both measured and unmeasured were comparable at the start (Morling et al.,

2022). The whole experiment will be held in Dutch. Four fifth-grade children and 13 sixth-grade children were included and randomly allocated to one of three conditions. Age varied from 10 to 12 (M = 11.2, SD = 0.75). Inclusion criteria comprised being a fifth or sixth-grade student registered at one of the aforementioned schools. Gifted children were already separated from non-gifted children within their school and were therefore easy to exclude. This study aims to make generalizable inferences about children in fifth and sixth grade and gifted children learn more rapidly than others, due to enhanced frontal cortical activation and faster neural processing speed (Geake, 2009). Therefore, gifted children were excluded. Participants with high prior knowledge, as indicated by a score of 13 on the pre-test, were excluded. Data showed that zero participants achieved a score of 13 or higher, therefore, no participants were excluded.

Materials

Learning Material

The learning material is comprised of information about the digestive system, different nutrients that your body needs, and the role of enzymes (605 words). The text is generated by ChatGPT-3.5 (Open AI, https://chat.openai.com). According to Kerndoelen (n.d.), fifth and sixth graders need to learn about the construction, shape, and function of parts of plants, animals, and humans. Therefore, study materials regarding the human digestive system are suitable for this age group. Kerndoelen describes all the learning objectives for all children in primary education, giving guidelines and the minimum level of knowledge and skills for primary education (Kerndoelen en Leerlijnen in het Onderwijs - Balans, 2023).

Pre-test

The prior knowledge test consists of 3 open questions (e.g. “Which five steps do we encounter in the process of digestion?”), designed by the researcher. Each question pertains to a different aspect of the study text. For instance, question 1 addresses digestion, question 2 concerns enzymes, and question 3 focuses on nutrients. Each correct answer will be granted 5 points (a total of 15 points). Points distribution is listed in the answer model. To allow for the calculation of interrater reliability

between the raters, a second rater was included to assess the pre-test, using the same answer model as the first rater. The intraclass correlation coefficient (ICC) was calculated to assess the reliability of the measurements using a two-way mixed effects model where people's effects are random and measures effects are fixed—the single measures ICC was set at 0.85, indicating a high level of agreement. Also, the Cronbach's alpha coefficient was used to determine the internal consistency of the pre-test. Results indicate that the pre-test has questionable internal consistency ($\alpha = .60$).

Homework Instructions

Instructions will be given on a Friday so that each participant can study the material (either, PT, LBT, or re-study) over the weekend. Each participant, independent of condition, was instructed to spend no longer than one hour on the entire homework assignment and to complete the whole assignment alone without any help from parents/friends/internet. After each participant was done with the homework assignment, they were asked to fill in a format in which they could indicate what they did during the homework assignment, including deviation from the instructions. (e.g. time spend studying). It was emphasized that it is important, to be honest and that it had no consequences for them. Also, each participant was required to bring their homework assignment with them on Monday.

Practice-Testing Condition

Karpicke et al. (2016) list several criteria for studying retrieval effects: 1) include a control condition, 2) exclude re-study moments, 3) ensure high initial retrieval success, and 4) prompt learners to recall prior episodic context (Karpicke & Zaromb, 2010; Lehman et al., 2014). This study will adhere to these criteria. Karpicke et al. (2014) found that retrieval practices must be guided to ensure high initial retrieval success. Therefore, participants are asked to study the text and then engage in question mapping. Question mapping helps the participants to form a conceptual network, guiding retrieval practices. A question map is a set of questions arranged in map format which participants will fill in.

In the practice-testing condition, participants receive the study material on Friday. During studying the material over the weekend participants will complete the question map while viewing the

text and then complete the question map without the text (criteria 3). On Sunday, participants are asked to perform a free recall test by themselves, and if possible, in the same location as where it was studied (criteria 4). Participants are prompted with the instruction that during the free recall test, they are not allowed to re-study the material and that this is crucial for the research. During the free recall test participants are allowed to write down everything they know about the text. After the free recall test, no re-study of the materials is allowed (criteria 2).

Learning-by-teaching Condition

Participants in this condition will be given the same text as the PT condition (on Friday) but will be instructed to generate a video instruction about the text using any video-recording device. Instructions comprised 1) explaining the study material to someone else (i.e. camera) as if that person has no prior knowledge on the topic, 2) explaining the content from your memory in your own words and not reading aloud. The generated video will not be taken into account when assessing the transfer test. Moreover, after the video was generated, participants were allowed to delete the video since this has no further relevance to this study.

Control Condition

Participants in the control condition (re-study) are instructed to study the material only by re-reading the text. Participants were allowed to determine when they were going to read the text throughout the weekend.

Post-test

The post-test (i.e. transfer-test) will be given to all participants on Monday and is different from the pre-test as this test requires participants to transfer their knowledge. The post-test consists of 3 open questions. The first and second questions are near transfer questions (Blume et al., 2009) since these tasks are similar to the context to be learned: 1) *"Brian is hungry and decides to eat a cheeseburger, fries, and a milkshake. Analyze this meal based on the nutrients that this meal offers. Give recommendations to Brian for an alternative meal that offers more healthy nutrients."* And 2) *"Kibo the chimpanzee lives in the zoo and receives a meal from his caretakers every day. Explain*

what happens to the food once Kibo puts it in his mouth. Describe the 6 steps that the food goes through, be as detailed as possible." The third question is a far transfer question since participants are required to apply characteristics of the enzyme collaboration process to another situation, meaning that the transfer task is quite different from the learning context: *"The cooperation of enzymes is similar to the teamwork in a sports team because…".*

This test will be given on paper. Participants are allowed to use a pencil and an eraser. Participants can achieve a score of 6 points on both questions 1 and 2 and three points on question 3. So, a total of 15 points can be achieved (See Appendix B for the answer model). The Cronbach's alpha coefficient was used to determine the internal consistency of the post-test. Results indicate that the pre-test has poor internal consistency ($\alpha = .21$). The ICC was calculated to assess the reliability of the measurements using a two-way mixed effects model where people's effects are random and measures' effects are fixed. ICC was set at 0.98, indicating a high level of agreement.

Procedure

Before the start of the experiment, informed consent was obtained from the parents/caretakers of the participants. Before the start of the experiment, each participant will either be placed in the PT condition, LBT condition, or control condition. At the start of the first session, the researcher will introduce himself and provide the participants with basic background information. Then, in the participants' classroom, participants with parental informed consent received information about the study and were given a form to indicate whether they would like to participate in the study or not. Afterward, the participants with consent received two envelopes consisting of 1) the pre-test and a questionnaire regarding age and gender and 2) home-work instructions and a short note for parents that states that it is important that their child sticks to the instructions and may not be given any help since it is part of a scientific study. Next, the researcher explained the information of all three conditions, indicating that participants will complete a transfer test on Monday which will not influence the participants' report card (this will also be emphasized in the note for parents/caretakers). At the end of the first session, participants can ask questions and will be instructed to write this down in their agenda before putting the envelope into their school bags.

At the start of the second session (Monday) the researcher distributes an envelope for each participant consisting of the transfer task regarding the digestive system, enzymes, and nutrients, and a blank paper. Participants received 20 minutes to finish the post-test.

Data-Analysis

Data were analyzed using a one-way ANOVA with condition as between-subject factor (conditions A, B, and C) to examine the effect of two GLSs on the transferability of knowledge. The purpose of this analysis was to determine whether there were significant differences in transfer among the different GLSs. Both the conventional analysis and Bayes testing methods were conducted to check whether there were significant differences and to examine which hypothesis gained the most support based on data. Statistical analyses were conducted using SPSS and Rstudio (Bayes) with a significance level set at $\alpha = 0.05$. When assessing the post-test, the total score of all questions combined was used (the same goes for the pre-test). For instance, when a participant scores two points on question 1, three points on question 2, and 1 point on question 3, the participant has a total score of 6 points. Using this score, the one-way ANOVA was conducted.

First, preliminary data screening was conducted to ensure that the assumptions of a one-way ANOVA were met. The assumption of a randomized controlled trial is met due to the design of this research. The dependent variable (transfer) is of continuous measurement level since a score is being linked to the post-test. There are independent observations as each participant is only being included in one condition. Participants are independent of each other across conditions; therefore, the assumption of independent groups is assured. To examine whether there are any outliers in the data, a scatterplot was employed to visually inspect the distribution of data and to detect any outliers. Potential outliers were inspected and subsequently removed from the dataset. The normal distribution of the data was examined by creating a visual representation of the distribution of the data as well as conducting a Shapiro-Wilk normality test. Finally, scores must have homogeneity in all conditions. This was examined using Levene's test.

Results

Before testing the hypotheses, conditions to perform an ANOVA were checked. All the assumptions were met. The Levene's test to check whether the scores have homogeneity showed a significance level of $p = .434$. Therefore, the null hypothesis of homogeneity in the scores may not be rejected. Also, data were normally distributed. The assumption of normality was met because the Shapiro-Wilk test showed a significance level of $p = .161$. Therefore, the null hypothesis of normally distributed data may not be rejected. Moreover, a histogram with standardized residuals was used to account for any outliers.

Table 1 shows the descriptive statistics of the scores for each condition on the pre-test and post-test. On the post-test, a maximum score of 15 points could be achieved. Interpretation of these descriptive statistics shows that for the control condition, the mean score was highest and that the learning-by-teaching condition achieved both the lowest average and minimum score. To determine whether the prior knowledge of each participant might have contributed to the results, it was investigated whether the groups were equally distributed at the start of the experiment. A one-way ANOVA was employed on the pre-test. Results indicated that there was no significant difference between conditions $F(2, 14) = 2.88$, $p = .09$, $\eta^2 = .29$, meaning that it can be assumed that the groups were equally distributed at the start of the experiment.

A one-way ANOVA was performed to compare the effect of three study strategies on transfer. The test shows that there was no significant effect of condition on mean post-test scores for transfer, $F(2, 14) = 2.37$, $p = 0.13$, $< \eta^2 = .25$. These results are not in line with the hypotheses which suggested that there was a significant difference between conditions. As seen in the descriptive data there is a difference in mean score between all three groups with the LBT condition (M = 5.00) and control condition (M = 8.40) having the biggest difference in mean scores. The effect size of $\eta^2 = .25$ indicates a large effect, suggesting that 25% of the variance on the post-test can be attributed to the condition.

Table 1

Descriptive Statistics of the Scores for Each Condition on Pre-test and Post-test (points range 0-15)

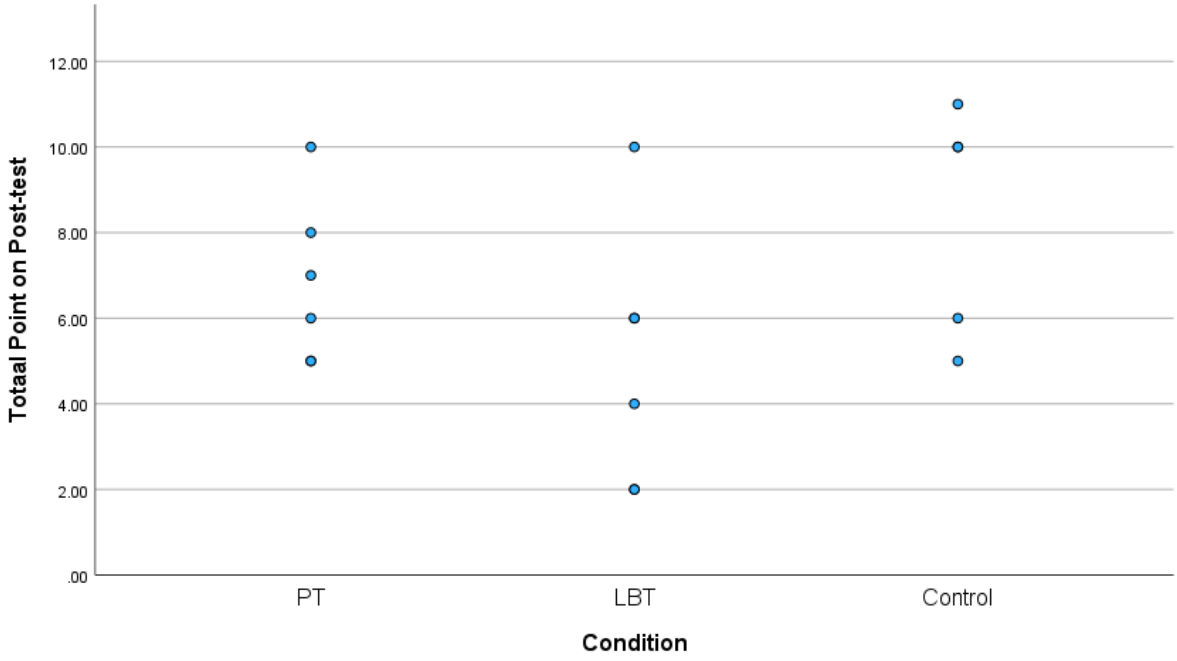
								95% Confidence Interval for Mean							
Condition	N	Mean		SD		Std. Error		Lower Bound		Upper Bound		Minimum		Maximum	
		Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test
PT	6	4.42	6.83	1.86	1.94	.76	.79	2.47	4.80	6.36	8.87	2.00	5.00	7.00	10.00
LBT	6	1.33	5.00	2.34	3.03	.95	1.24	-1.12	1.81	3.79	8.18	.00	2.00	6.00	10.00
Control	5	2.80	8.40	2.49	2.70	1.11	1.21	-.29	5.05	5.89	11.75	1.00	5.00	7.00	11.00
Total	17	2.85	6.65	2.47	2.80	.60	.68	1.58	5.20	4.12	8.09	0.00	2.00	7.00	11.00

The current statistical analysis holds poor value due to the small sample size. A post-hoc power analysis showed a power of 0.13. This indicates a very poor power, meaning that there is a high chance that no existing effect will be detected, even if it exists. In other words, the risk of a Type II error is high. Because of this reason, other steps are taken to check whether there are alternative ways lend themselves to explore any differences between groups. First, raw data will be very carefully interpreted, looking for possible clues that might explain group differences. Second, a Bayesian test was conducted. Bayesian testing solely relies on the observed data and enables an assessment of the evidence supporting the null hypothesis (Wei et al., 2022) In other words, the Bayes testing method was used to assess the strength of the evidence (Wagenmakers et al., 2017). Moreover, Bayesian testing is less sensitive for (too) small sample sizes (Van de Schoot et al., 2015). Bayesian testing compares evidence that would support the null hypothesis (H_0: $\mu_{pt} = \mu_{LBT} = \mu_{rs}$) relative to the alternative hypothesis (H_0: $\mu_{pt} \neq \mu_{LBT} \neq \mu_{rs}$) that would indicate that at least one of the conditions would have differing results. This testing method uses data as an extraction point, relative to conventional analyses, that use the null hypothesis as an extraction point. The Bayesian testing

method assesses to what extent a certain hypothesis fits the data the most and is called the Bayes factor (BF). This indicates which hypothesis gains the most support based on the data provided. For instance, a $BF_{01} = 5.00$ means that there is 5 times more support for the null hypothesis than for the alternative hypothesis. As mentioned, raw data will be carefully interpreted. In Figure 2, the scores of all the participants are visualized, and distributed by condition. The control condition has the highest mean score but did not consistently outperform the other conditions since considerable overlap is observed.

Figure 2

Score Distribution of Post-test (N = 17)

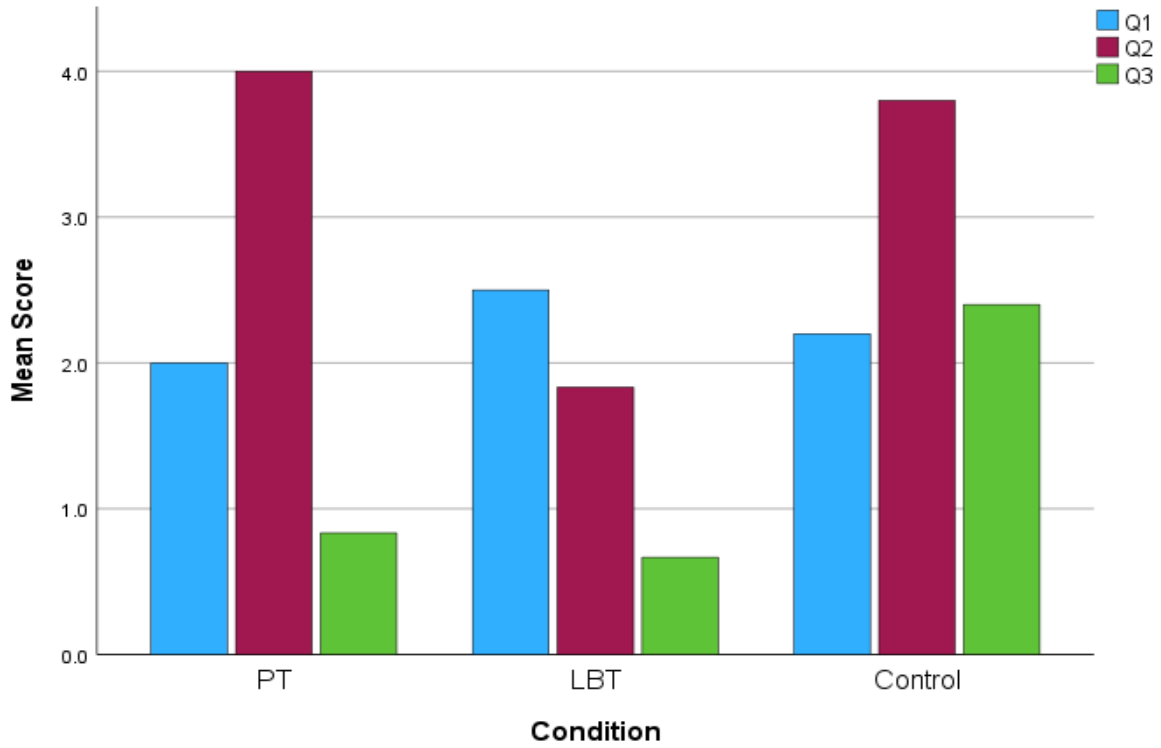


Note. PT - *n* = 6, LBT - *n* = 6, Control - *n* = 5. Total - *n* = 17

Interpretation of the raw data showed that participants in the Control condition did best on question 3 (far transfer), and question 2 (near transfer). At the same time, the LBT condition outperformed the other conditions on question 1 (near transfer). Figure 2 shows the distribution of scores on each question on the post-test. These results look promising, but no reliable inferences can be drawn from this. This means that despite the promising results, it is still not clear to what extent one learning strategy is favorable over the other.

Figure 3

Distribution of Scores per Question on Post-test



Note. PT - *n* = 6, LBT - *n* = 6, Control - *n* = 5. Total - *n* = 17

Bayesian Testing

In this experiment, the cut-off significance level of a = .05 using the Null Hypothesis Significance Testing (NHST) does not allow to show any significant difference. Moreover, the likelihood that the hypothesis under study is correct or that the data were the result of pure chance is not measured by p-values (Wei et al., 2022). For this reason, Bayesian testing was conducted on the data that would allow room to analyze whether hypotheses were likely true or not. The Bayes testing was performed using RStudio. To perform the Bayesian testing method, assumptions were checked. Since the same assumptions of the NHST apply to the Bayes test (Hoijtink, 2019) no further inspection regarding the assumptions was needed. Also, the Bayes factor and posterior model probabilities were determined (PMPs). A PMP is another way of expressing the BF and allowing to report type 1 and type 2 errors.

Interpretation of the results shows a Bayes factor of $BF_{10} = .45$. Meaning that there is 1 / .45 = 2.23 times more support for the null hypothesis than for the alternative hypothesis (Dienes, 2021). Also, a PMP of .69 for the null hypothesis and a PMP of .31 for the alternative hypothesis was found. Meaning that the chance for a type 1 error (rejecting the null hypothesis while in reality, there is no difference between conditions) is $\alpha = .69$, and the chance for a type 2 (not rejecting the null hypothesis while in reality, there is a difference between groups) error is $\beta = .31$. Moreover, the chance that H_0 is true is .69 while the chance that H_1 is true is .31.

Both the conventional analysis and the Bayes test show little support for the probability that reliable inferences can be drawn from this experiment. The conventional analysis does not show significant results and the Bayes testing method confirms that the null hypothesis gets more support to fit the data than the alternative hypothesis does. However, mean differences are found in the descriptive statistics, indicating that the control condition scores higher on the post-test than the PT and LBT conditions. But, since both testing methods confirm little effect, differences are more likely to be the result of random variability rather than a true underlying effect. Regarding the hypotheses after the literature review, it is not reliable to conclude the effects of these generative learning strategies on the transfer among fifth and sixth graders.

Discussion

The goal of the current study was to examine the effectiveness of two generative learning strategies among fifth and sixth-grade children on transfer. While earlier studies have explored the effects of both LBT and PT on children's learning outcomes (e.g., comprehension and retention), the effects of these GLSs on the learning outcome of transfer among this age group are quite limited. Therefore, this study aims to fill this gap using two GLSs, namely: practice testing (i.e. practicing retrieval practice) and learning by teaching (making a video in which participants explain content to the camera). A third condition (as a control group) was added in which participants were instructed only to restudy the text. In the experiment the participants were asked to study a text by either practicing with the material and engaging in a free recall activity (PT condition), explaining the content to a fictitious peer while making a video of oneself (LBT condition), or just reading the text

(restudy condition). The experiment entailed a homework assignment that would only last for about one hour. The day after the weekend, participants engaged in a post-test to assess their ability to transfer their knowledge on both near and far transfer questions.

To give concluding comments on the formulated hypotheses, it is necessary to have reliable evidence. Currently, both ways of testing do not show a substantial amount of evidence, meaning that no reliable inferences can be drawn and that it can be assumed that the difference between conditions is a result of random variability. This means that there are, perhaps, other causes for the difference in means between groups. The low statistical power of this study might be a possible explanation for why the findings are not in line with the hypotheses. Due to a poor sample size, data was not sufficient to show any statistical difference between groups. A larger sample size would result in a higher statistical power, meaning that reliable inferences can be drawn to generalize to the population (Uttley, 2019). Since this study is not representative to be making inferences, I will still attempt to interpret results (based on mean scores) to shed light on the hypotheses.

Results of the conventional statistical one-way ANOVA and the Bayesian test showed that there is no evidence for any meaningful difference between conditions on transfer. These results are not in line with the hypotheses. Moreover, looking at the means of each condition, the restudy condition tends to outperform both the PT and LBT conditions. These findings are not congruent with the findings of other studies. In the study by Hoogerheide et al. (2019), results were found that suggest that explaining on video was more beneficial for test performance than restudying. Hoogerheide et al. state that the key to the effectiveness of teaching over restudying seems to lie in the social presence theory. However, mean differences in this study do not show similar results. A reason for this contradictory finding might be that in the study by Hoogerheide et al., participants were aware of potential audience, promoting psychological arousal. In the current study, participants were told that they could delete the video afterward and that it was not being assessed, perhaps missing the positive effects of psychological arousal (Cui et al., 2013). Also, Hoogerheide et al. examined conceptual knowledge whereas the current study examined transfer. Therefore, the result might differ.

The LBT condition has the lowest mean score (M = 5.00) and the highest standard deviation (SD = 3.03) on the post-test, contradicting hypotheses that it would outperform the restudy and PT conditions. If these non-significant results are valid and not due to low statistical power, a section of the cognitive load theory (Sweller, 2011) may offer an explanation. This theory posits that an individual's working memory has a limited capacity, and exceeding this capacity leads to cognitive overload and poor recall. This suggests that children, usually in possession of lower cognitive capacity (Esposito, 2022), might struggle more with teaching, as it occupies more working memory compared to adults (Feldon, 2007), who can allocate cognitive resources more efficiently.

Brod (2020) supports this, finding that older students benefit more from generating explanations. However, Muis et al. (2016) indicate that studying as if to teach improves learning outcomes for young children, implying that the teaching effect increases with age. The poor LBT post-test scores might also be due to insufficient explanation quality. Fiorella and Mayer (2015) stress that meaningful learning requires high-quality explanations. Participants might not have studied the text well enough or relied on reading aloud from the text during their explanations, which Lachner et al. (2020) found less effective than retrieval practice.

In this study, the PT condition did not outperform the restudy condition (in terms of means), also contradicting the hypothesis. Numerous studies have described the positive effects of retrieval practice on different kinds of learning outcomes, under which transfer (Butler, 2010). However, recalling information might not always be evenly successful (Kornell, 2009) resulting in reduced effectiveness of the GLS (Butler, 2010). Moreover, recalling wrong information and being awarded with feedback after a delay, produce reduced retention in comparison to being awarded with feedback immediately afterward (Hays, 2013). Moreover, Kang et al. (2007) state that practice testing is most effective when corrective feedback is provided. In the current experiment, the participants were not awarded with feedback on their free recall activity due to the homework assignment. Therefore, the benefits of feedback did not apply in this GLS resulting in the probability that participants recalled the wrong information. Also, in the current experiment, it was not clear to what extent the participants were successful in their retrieval practice. Another possibility for the results contradicting the

hypotheses is the immediate post-test. In a study by Roediger and Karpicke (2006), participants engaged in a free-recall test (study-test condition) or merely reread the study material (study-study condition). Their results indicate that the study-study condition outperformed the study-test condition in an immediate post-test. However, when the post-test was administered after a delay (2 days), the study-test condition outperformed the study-study condition. The findings of Roediger and Karpicke might explain the findings of this study, indicating that the beneficial retrieval effects are larger after a delay (e.g. two weeks). The findings in this study also contradict the findings of Butler (2010). Butler found that repeated testing enhances transfer. However, Butler incorporated immediate feedback after retrieval attempts and engaged in more retrieval moments than in this study. Meaning that, when retrieval is successful, information would be easier to retrieve from long-term memory.

Limitations

A major limitation of this study was the poor sample size (n = 17) compared to a population size of approximately N = 301.852. With the current sample size, the achieved power was set at 0.13 indicating a low power. This limitation caused research to be of low value to the educational field. Having a low power, and therefore, a low probability that any variations between groups will be the result of a learning strategy is detrimental to this research.

Aside from the low statistical power, another limitation in this research is that there was no sight of how well the participants in the PT condition performed on their free recall activity. Also, participants were given zero feedback on the free recall activity. According to Hays (2013) and Kang et al. (2007), corrective feedback on retrieval practices is crucial to achieve successful retrieval. When no feedback is given on the retrieval process, it could be the case that individuals remember information incorrectly. In the current experiment, the participants in the PT condition engaged in a post-test after the weekend, meaning that there was little delay between the retrieval activity and the post-test. In the experiment, participants were asked to fill in a form in which they could indicate what they did during the homework activity. Of the six participants in the PT condition, only one participant indicated to have performed the free-recall activity on the day of the post-test. The other 5 participants did not give such information. This shows the probability that participants engaged in the

homework activity on Sunday, meaning that there is no delay between the retrieval activity and the post-test which results in a less effective remembrance of information.

Another limitation to this research is the poor internal consistency of the post-test ($\alpha = .21$). It could be argued that the reason for the results to be not in line with the hypotheses, is due to the low internal consistency. This means that the questions on the post-test may not be reliably measuring the same underlying construct (transfer). Aside from the fact that there is a low statistical power, the low internal consistency also limits the generalizability of the results. This limitation raises concerns about the reliability of this study. Since the measurement tool may not be reliable, the findings based on this tool may not be generalized to different contexts.

In this study, the participants in the LBT condition were not required to hand in their generated video. This decision might be a limitation of this study since results might be better explained by the generated videos. These videos could give insight into what the participants did, what the quality of their explanation was, and why these strategies did not work with this sample.

Implications for Practice

Because both testing methods (Conventional one-way ANOVA and Bayes) do not show significant results nor give support for any differences between groups, indicating insufficient evidence to favor either PT or LBT on transfer for fifth and sixth graders. However, other studies did succeed in obtaining reliable evidence to make inferences about the effectiveness of PT and LBT, such as Butler (2010) and Hoogerheide et al. (2019). Butler implied that practice testing promotes transfer among adults and Hoogerheide et al. implied that generating an explanatory video has positive effects on learning outcomes. And even for elementary school children, testing promotes transfer (Rohrer et al., 2010). Yet it remains unclear which GLS is more favorable for transfer on fifth and sixth graders. A comparative study between these two GLSs was done by Lachner et al. (2020), implying that explaining was not more beneficial for learning than retrieval practice, however, this study measured conceptual knowledge, not transfer. If this study had sufficient statistical power and if the results were significant, there would be several benefits for practice. First, teachers will be better

able to instruct their students in picking the most fitting learning strategy when the aim is transfer. Secondly, students would also benefit from this, solving the problem stated by Bjork et al. (2012), indicating that students often pick the wrong study strategy for their learning purposes and that they are not instructed on how to use a certain strategy.

Future Research

Both the conventional statistical analysis and the Bayesian testing conducted in this study indicated non-significant results and offered limited support for the alternative hypothesis. Aside from repeating this study with a larger sample size to obtain adequate statistical power, future research should also focus on obtaining a larger internal consistency of the post-test. The current post-test is not reliable and does not yield beneficial results (Sürücü & Maslakçi, (2020).

Future research could incorporate a broader range of variables since this might provide a more comprehensive understanding of the research question. This might involve demographic variables, psychosocial factors, or contextual influences that were not accounted for in the current study. By adopting a more holistic approach, researchers can control for potential confounders and better isolate the effect of the primary variables of interest (Westfall & Yarkoni, 2016).

As stated in the discussion section, cognitive abilities tend to change over time. Meaning that as you get older, cognitive capacities tend to increase. Moreover, Brod (2020) state that there is mixed evidence as to how much elementary school children benefit from explaining content. This could mean that teaching to others might have different results that change over time. Doing this in a within-subjects design, the effect of learning by teaching can be better isolated and results can subsequently be accounted to individual differences.

Conclusion

In this paper, experimental research was conducted to answer the following research question: “What is the comparative impact of the generative learning strategies practice-testing and learning-by-teaching on the learning outcome of transfer among fifth and sixth grades in the context of a homework activity?”

In this study, 17 participants engaged in the homework activity. Results indicated that LBT and PT do not improve transfer relative to restudying in the context of a homework activity. Given these results, no meaningful inferences for the educational practice can be drawn. Rather, future research should complement this research since the gap in research has not been filled yet by this study. This research did not fulfill its mission to answer the research question. Future research on this topic should 1) yield more participants to gain larger statistical power, 2) improve the internal consistency of the post-test, 3) incorporate more variables, 4) better monitor the level of success in retrieval practices to subsequently give feedback, 5) analyze generated videos to gain more insight into results, and 6) isolate the effects of LBT to examine age-related differences.

Acknowledgments

In this study, Generative AI (Open AI, https://chat.openai.com) was used in order to generate the study text. After the text was generated and a few adjustments were prompted, the final text was evaluated on correctness and suitability for the age group. Not everything that Chat-GPT has generated was used. For instance, a figure was added, and difficult words were replaced with easier, suitable words. A few examples of my exact prompts:

- "design an experiment in which fifth and sixth grade students are required to learn on a topic. A few days later these students are required to transfer their knowledge and apply that to solve problems on near transfer tasks" – February 27th, 16:53.
- "give me another topic" – February 27th, 16:57
- "give me a topic related to the human body" February 27th, – 16:59
- "design the homework text that these students need to learn. be very elaborate and specific. Use at least 1000 words" February 27th, 17:04
- "translate to dutch" February 27th, 17:05

As you can tell by the first prompt, an experimental design was requested. This was done only for inspiration purposes. The actual experiment in this study was not obtained via Chat-GPT. Only the text that Chat-GPT had generated was used as a basis for the study text. No other outputs were used.

References

Adams, J. A. (1987). Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. *Psychological Bulletin, 101*(1), 41-74. https://doi.org/10.1037/0033-2909.101.1.41

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: a meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659–701. https://doi.org/10.3102/0034654316689306

Bargh, J. A., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology, 72*(5), 593–604. doi:10.1037/0022-0663.72.5.593

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

Brod, G. (2021). Generative learning: Which strategies for what age?. *Educational Psychology Review, 33*(4), 1295-1318. DOI:10.1007/s10648-020-09571-9

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

Carpenter, S. K. & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent attention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34*(2), 268–276. DOI: 10.3758/bf03193405

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

Chmielewski, T. C., & Dansereau, D. F. (1998). Enhancing the recall of text: Knowledge mapping training promotes implicit transfer. *Journal of Educational Psychology, 90*(3), 407. https://doi.org/10.1037/0022-0663.90.3.407

Chularut, P., & DeBacker, T. K. (2004). The influence of concept mapping on achievement, self-regulation, and self-efficacy in students of English as a second language. *Contemporary Educational Psychology, 29*, 248–263. doi:10.1016/j.cedpsych.2003.09.001.

Cui, G., Lockee, B., & Meng, C. (2013). Building modern online social presence: A review of social presence theory and its instructional design implications for future trends. *Education and information technologies, 18*, 661-685. DOI:10.1007/s10639-012-9192-1

Dienes, Z. (2021). How to use and report Bayesian hypothesis tests. *Psychology of Consciousness: Theory, Research, and Practice, 8*(1), 9. https://doi.org/10.1037/cns0000258

Dori, Y. J., & Sasson, I. (2013). A three-attribute transfer skills framework – part I: establishing the model and its relation to chemical education. *Chem. Educ. Res. Pract., 14*(4), 363–375. doi:10.1039/c3rp20093k

Duran, D., & Topping, K. (2017). Learning by teaching: Evidence-based strategies to enhance learning in the classroom. Taylor & Francis. https://doi.org/10.4324/9781315649047

Esposito, A. G., & Bauer, P. J. (2022). Determinants of elementary-school academic achievement: Component cognitive abilities and memory integration. *Child Development, 93*(6), 1777-1792. https://doi.org/10.1111/cdev.13819

Feldon, D. F. (2007). Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational psychologist, 42*(3), 123-137. https://doi.org/10.1080/00461520701416173

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology, 38*(4), 281–288. doi:10.1016/j.cedpsych.2013.06.00

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

Fiorella, L., & Zhang, Q. (2018). Drawing boundary conditions for learning by drawing. *Educational Psychology Review, 30*(3), 1115–1137. https://doi.org/10.1007/s10648-018-9444-8.

Fortney, C. (2016). How Best to Study for a Test: A Comparison of Practice Retrieval and Self-Explanation.

Geake, J. G. (2009). Neuropsychological Characteristics of Academic and Creative Giftedness. International Handbook on Giftedness, 261–273. doi:10.1007/978-1-4020-6162-2_11

Grabowski, B. L. (2013). Generative learning contributions to the design of instruction and learning. In Handbook of research on educational communications and technology (pp. 713-737). Routledge.

Hajian, S. (2019). Transfer of learning and teaching: A review of transfer theories and effective instructional practices. *IAFOR Journal of education, 7*(1), 93-111. DOI:10.22492/ije.7.1.06

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(1), 290–296. doi:10.1037/a0028468

Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological methods, 24*(5), 539. https://doi.org/10.1037/met0000201

Hoogerheide, V., Deijkers, L., Loyens, S. M. M., Heijltjes, A., & van Gog, T. (2016). Gaining from explaining: Learning improves from explaining to fictitious others on video, not from writing to them. *Contemporary Educational Psychology,* 44-45, 95–106. doi:10.1016/j.cedpsych.2016.02.005

Hoogerheide, V., Visee, J., Lachner, A., & van Gog, T. (2019). Generating an instructional video as homework activity is both effective and enjoyable. *Learning and Instruction, 64*, 101226. https://doi.org/10.1016/j.learninstruc.2019.101226

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

Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval mode distinguishes the testing effect from the generation effect. *Journal of Memory and Language, 62*(3), 227-239. https://doi.org/10.1016/j.jml.2009.11.010

Karpicke, J. D., Blunt, J. R., Smith, M. A., & Karpicke, S. S. (2014). Retrieval-based learning: The need for guided retrieval in elementary school children. *Journal of Applied Research in Memory and Cognition, 3*(3), 198–206. https://doi.org/10.1037/h0101802

Karpicke, J. D., Blunt, J. R., & Smith, M. A. (2016). Retrieval-based learning: Positive effects of retrieval practice in elementary school children. *Frontiers in Psychology, 7*, 350. https://doi.org/10.3389/fpsyg.2016.00350

Kerndoelen. (z.d.). SLO. https://www.slo.nl/sectoren/po/kerndoelen/

Kerndoelen en leerlijnen in het onderwijs - Balans. (2023, 18 oktober). Balans. https://balansdigitaal.nl/kennisbank/onderwijs/neem-de-stille-regie/kerndoelen-en-leerlijnen-in-het-onderwijs#:~:text=kerndoelen%20zijn%20samen%20met%20de,Het%20zijn%20streefdoelen.

Kintsch, W. (1986). Learning from text. *Cognition and instruction, 3*(2), 87-108. https://doi.org/10.1037/0003-066X.49.4.294

Klingenberg, S., Jørgensen, M. L., Dandanell, G., Skriver, K., Mottelson, A., & Makransky, G. (2020). Investigating the effect of teaching as a generative learning strategy when learning through desktop and immersive VR: A media and methods experiment. British *Journal of Educational Technology, 51*(6), 2115-2138. doi:10.1111/bjet.13029

Klausmeier, H. J. (1961). Educational psychology. Learning and human abilities. New York: Harper.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(4), 989. DOI: 10.1037/a0015729

Lachner, A., Backfisch, I., Hoogerheide, V., van Gog, T., & Renkl, A. (2020). Timing matters! Explaining between study phases enhances students' learning. *Journal of Educational Psychology, 112*(4), 841–853. https://doi.org/10.1037/edu0000396

Lachner, A., Hoogerheide, V., van Gog, T., & Renkl, A. (2022). Learning-by-teaching without audience presence or interaction: When and why does it work?. Educational *Psychology Review,* 1-33. https://doi.org/10.1007/s10648-021-09643-4

Lee, H. W., Lim, K. Y., & Grabowski, B. L. (2008). Generative learning: Principles and implications for making meaning. In Handbook of research on educational communications and technology (pp. 111-124). Routledge.

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

Leopold, C., & Mayer, R. E. (2015). An imagination effect in learning from scientific text. *Journal of Education and Psychology, 107*, 47–63. doi:10.1037/a0037142.

Li, S.-C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the Couplings Among Intellectual Abilities and Constituent Cognitive Processes Across the Life Span. *Psychological Science, 15*(3), 155–163. doi:10.1111/j.0956-7976.2004.01503003.x

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

Moreira, B. F. T., Pinto, T. S. S., Starling, D. S. V., & Jaeger, A. (2019, February). Retrieval practice in classroom settings: A review of applied research. *In Frontiers in Education, 4(*5). Frontiers Media SA. https://doi.org/10.3389/feduc.2019.00005

Morling, B. & Carr, D. et al. (2022). Introduction to and Application of Research Methods and Statistics. Universiteit Utrecht. Norton. ISBN 978-1-324-02429-3

Muis, K. R., Psaradellis, C., Chevrier, M., Di Leo, I., & Lajoie, S. P. (2016). Learning by preparing to teach: Fostering self-regulatory processes and achievement during complex

mathematics problem solving. *Journal of Educational Psychology, 108*(4), 474–492. doi:10.1037/edu0000071

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of educational research, 76*(3), 413-448. https://doi.org/10.3102/00346543076003413

Osborne, R., & Wittrock, M. (1985). The generative learning model and its implications for science education.

Rawson, K. A., & Dunlosky, J. (2012). When is practice testing most effective for improving the durability and efficiency of student learning?. *Educational Psychology Review, 24(*3), 419-435. DOI 10.1007/s10648-012-9203-1

Ritchie, D., & Volkl, C. (2000). Effectiveness of two generative learning strategies in the science classroom. *School Science and Mathematics, 100*(2), 83-89. https://doi.org/10.1111/j.1949-8594.2000.tb17240.x

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child development, 77*(1), 1-15. https://doi.org/10.1111/j.1467-8624.2006.00852.x

Rhodes, S., Greene, N. R., & Naveh-Benjamin, M. (2019). Age-related differences in recall and recognition: A meta-analysis. *Psychonomic Bulletin & Review, 26*, 1529-1547. https://doi.org/10.3758/s13423-019-01649-y

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

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 36*(1), 233. https://doi.org/10.1037/a0017678

Roscoe, R. D., & Chi, M. T. H. (2007). Tutor learning: the role of explaining and responding to questions. *Instructional Science, 36*(4), 321–350. doi:10.1007/s11251-007-9034-5

Schwamborn, A., Mayer, R. E., Thillmann, H., Leopold, C., & Leutner, D. (2010). Drawing as a generative activity and drawing as a prognostic activity. *Journal of Education and Psychology, 102*, 872–879. doi:10.1037/a0019640

Shrager, L., & Mayer, R. E. (1989). Note-taking fosters generative learning strategies in novices. *Journal of Educational Psychology, 81*(2), 263.

Sürücü, L., & Maslakçi, A. (2020). Validity and reliability in quantitative research. *Business & Management Studies: An International Journal, 8*(3), 2694-2726.

Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press. DOI 10.1007/978-1-4419-8126-4

Uttley, J. (2019). Power Analysis, Sample Size, and Assessment of Statistical Assumptions—Improving the Evidential Value of Lighting Research. LEUKOS, 1–20. doi:10.1080/15502724.2018.1533851

Van de Schoot, R., Broere, J. J., Perryck, K. H., Zondervan-Zwijnenburg, M., & van Loey, N. E. (2015). Analyzing small data sets using Bayesian estimation: the case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology, 6*(1), 25216. doi:10.3402/ejpt.v6.25216

Wagenmakers, E. J., Verhagen, J., Ly, A., Matzke, D., Steingroever, H., Rouder, J. N., & Morey, R. D. (2017). The need for Bayesian hypothesis testing in psychological science. *Psychological science under scrutiny: Recent challenges and proposed solutions,* 123-138. https://doi.org/10.1002/9781119095910.ch8

Wei, Z., Yang, A., Rocha, L., Miranda, M. F., & Nathoo, F. S. (2022). A review of Bayesian hypothesis testing and its practical implementations. *Entropy, 24*(2), 161. https://Doi.org/10.3390/e24020161

Westfall, J., & Yarkoni, T. (2016). Statistically Controlling for Confounding Constructs Is Harder than You Think. *PLOS ONE, 11*(3), https://doi.org/10.1371/journal.pone.0152719

Wilson, R. M., Gambrell, L. B., & Pfeiffer, W. R. (1985). The effects of retelling upon reading comprehension and recall of text information. *The Journal of Educational Research, 78*(4), 216-220.

Yang, B. W., Razo, J., & Persky, A. M. (2019). Using testing as a learning tool. *American journal of pharmaceutical education, 83*(9), 7324. https://doi.org/10.5688/ajpe7324

Appendices

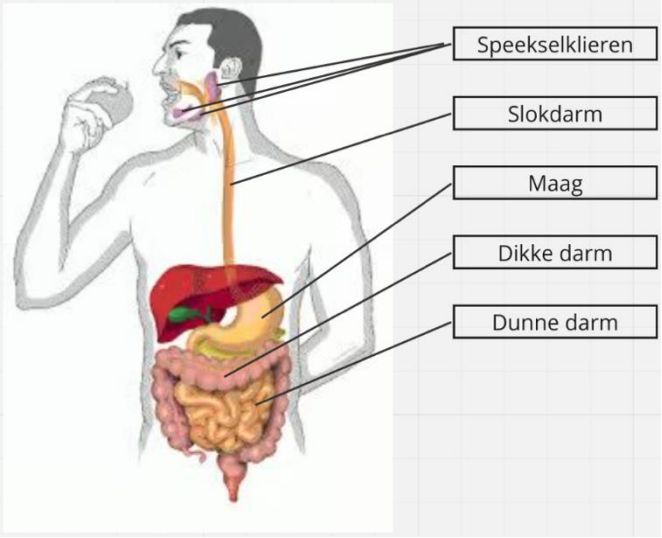
Appendix A

Study Text

Begrip van het Spijsverteringsproces:

Laten we eens kijken naar hoe ons lichaam voedsel verteert. Het spijsverteringsstelsel is als een team van organen dat samenwerkt om voedsel klein te maken, zodat ons lichaam het kan gebruiken. Het begint allemaal in de mond, waar het kauwen en speeksel het voedsel zacht maakt om door te slikken.

Terwijl de brij door de slokdarm naar beneden gaat, duwt de slokdarm het in de maag, waar maagsappen met zoutzuur en enzymen de afbraak van eiwitten, vetten en koolhydraten voortzetten. Doordat de maag als het ware kneedt wordt de inhoud verder gemengd, wat leidt tot een soort vloeibaar geheel.



Vervolgens komt dit geheel in de dunne darm, waar het grootste deel van de voedingsstofopname plaatsvindt. Hier helpen enzymen uit de alvleesklier en gal uit de lever bij de spijsvertering van koolhydraten, eiwitten en vetten. Alles wat overgebleven is gaat naar de dikke darm, waar water en elektrolyten worden opgenomen, en afvalproducten worden gevormd tot poep.

Verkenning van de Rol van Enzymen:

Centraal in het spijsverteringsproces staan enzymen. Dit zijn speciale eiwitten en zijn als het ware kleine helpers in je lichaam. Ze breken eten af in stukjes, zodat je lichaam het kan gebruiken. Elk soort enzym richt zich op specifieke voedingsstoffen van het eten om te zorgen dat het eten goed verteerd kan worden.

In je mond en buik zit er een enzym dat grote stukken eten, zoals brood en pasta veranderd in kleine stukjes die je lichaam gemakkelijk kan gebruiken. Het verandert die grote stukken eten in glucose, dit is een soort suiker.

Dan zijn er enzymen die werken hard om eiwitten, die je krijgt van dingen zoals vlees en kaas, in kleine stukjes te hakken die je lichaam kan gebruiken om sterk en gezond te blijven. Ze veranderen die grote stukken eiwit in kleine stukjes die 'aminozuren' worden genoemd.

Voeding: De Sleutel tot Gezondheid en Vitaliteit:

Naast het begrijpen van de spijsvertering, is het handig om te weten wat voeding nou doet met je gezondheid. Een gebalanceerd dieet levert alle voedingsstoffen (koolhydraten, eiwitten, vetten, vitaminen, mineralen en water) die je lichaam nodig heeft om optimaal te functioneren.

Koolhydraten zijn je belangrijkste bron van energie, waardoor je lichaam brandstof krijgt om te bewegen. Gezonde bronnen van koolhydraten zijn onder andere granen, fruit, groenten en peulvruchten. Ongezonde bronnen van koolhydraten zijn onder andere bewerkte koolhydraten zoals wit brood, frietjes, witte rijst, suikerrijke vruchtensappen en frisdranken.

Eiwitten zijn de bouwstenen van weefsels, spieren en organen. Dit is essentieel om te groeien en om je lichaam zichzelf te laten herstellen. Gezonde voedingsmiddelen rijk aan eiwitten zijn onder andere vlees, vis, eieren, zuivelproducten, noten en zaden. Ongezonde bronnen rijk aan eiwitten zijn bijvoorbeeld bewerkt vlees zoals worstjes, bacon en hamburgers met veel verzadigd vet.

Vetten spelen een belangrijke rol bij het aanmaken van hormonen en het helpen bij de opname van voedingsstoffen. Gezonde bronnen van vetten zijn onder andere avocado's, noten, zaden, olijfolie en vette vis. Ongezonde bronnen van vetten zijn verzadigde vetten uit bijvoorbeeld koekjes en gebak. Ook bevat gefrituurd eten en fastfood veel ongezonde vetten.

Vitaminen en mineralen zijn onmisbaar voor je lichaam en zorgen ervoor dat je gezond blijft en geen chronische ziektes krijgt. Zout is een onderdeel van een mineraal maar het eten van te veel zout kan leiden tot hoge bloeddruk en het risico op hart- en vaatziekten. Het eten van producten met toegevoegde suikers en verzadigde vetten zorgt ervoor dat je lichaam minder goed vitaminen en mineralen kan opnemen.

Water is belangrijk voor hydratatie, voedingsstoffentransport, en afvalverwijdering. Het drinken van een voldoende hoeveelheid water per dag is belangrijk voor het behouden van gezondheid en het voorkomen van uitdroging.

Appendix B

Post-Test Answer Model

Vraag 1

Brian heeft honger en besluit om een cheeseburger en frietjes te eten.

Welke ongezonde <u>bronnen</u> van voedingsstoffen zitten er in deze maaltijd?

Verzadigde vetten **(1pt):** Het vlees en de kaas in de cheeseburger bevatten doorgaans aanzienlijke hoeveelheden verzadigde vetten, die geassocieerd zijn met een verhoogd risico op hart- en vaatziekten en andere gezondheidsproblemen wanneer ze in overmaat worden geconsumeerd.

Bewerkte koolhydraten **(1pt):** De broodjes van de cheeseburger en de aardappelen van de frietjes bevatten vaak bewerkte koolhydraten, die snel worden afgebroken tot glucose in het lichaam en kunnen leiden tot schommelingen in de bloedsuikerspiegel en insulineresistentie bij regelmatige consumptie.

Toegevoegde suikers **(1pt):** Sauzen en dressings die in de cheeseburger worden gebruikt, evenals mogelijk toegevoegde suikers in de broodjes, kunnen de totale hoeveelheid toegevoegde suikers verhogen, wat kan bijdragen aan obesitas, diabetes en andere metabole aandoeningen.

Zout **(1pt):** In een cheeseburger met friet zit vaak een groot gehalte aan zout wat kan leiden tot klachten in de nieren.

Brian kiest de volgende dag een kipfilet sandwich op volkorenbrood. Leg uit waarom dit gezonder is (tip: denk aan de voedingsstoffen):

Eiwitten **(1pt)**: Kipfilet is een uitstekende bron van eiwitten, die essentieel zijn voor de opbouw en reparatie van weefsels in ons lichaam. Eiwitten helpen ook om ons langer verzadigd te houden, waardoor ze een goede keuze zijn voor een vullende maaltijd.

Gezonde koolhydraten **(1pt)**: Volkorenbrood bevat complexe koolhydraten, vezels, vitamines en mineralen die ontbreken in wit brood. De vezels helpen de spijsvertering te reguleren, verminderen het risico op hart- en vaatziekten en helpen om een gezond cholesterolgehalte te behouden.

Vraag 2

Kibo de chimpansee woont in de dierentuin en krijgt elke dag een maaltijd van zijn verzorgers. Leg uit wat er met het eten gebeurt zodra Kibo het eten in zijn mond stopt. Benoem de 6 stappen die het voedsel aflegt, wees zo uitgebreid mogelijk.

1. Kauwen **(1pt)**: Het eten wordt in kleine stukjes gehakt waardoor het eten makkelijker kan worden doorgeslikt.
2. Speeksel **(1pt)**: Tijdens het kauwen wordt speeksel geproduceerd, dat enzymen bevat die helpen bij het afbreken van voedsel.
3. Slikken **(1pt)**: de zachte brij wordt doorgeslikt en gaat via de slokdarm naar de maag.
4. Maag **(1pt)**: In de maag breken enzymen en maagsappen de voedingsstoffen af en wordt er als het ware gekneed.
5. Dunne darm **(1pt)**: Vanuit de maag komt de brij in de dunne darm waar allerlei voedingsstoffen worden opgenomen als gevolg van de spijsvertering.

6. Dikke darm **(1pt)**: na de dunne darm komt de brij in de dikke darm waar water en elektrolyten worden opgenomen. Alles wat overblijft wordt poep.

Vraag 3

Denk terug aan hoe enzymen samenwerken om voedsel in kleine stukjes te maken. Denk daarna aan een teamsport dat samenwerkt om te winnen.

Vul in:

Je lichaam heeft enzymen die Eiwitten/kleine stukken eten **(1pt) afbreken en je lichaam heeft enzymen die** Koolhydraten/grote stukken eten **(1pt) afbreken.**

Denk nu aan een teamsport (bijvoorbeeld voetbal, hockey, honkbal enz.) en benoem twee verschillende posities die belangrijk zijn bij het halen van een overwinning. Vul in:

De eerste positie die belangrijk is bij de teamsport [ANTWOORD] **is [**ANTWOORD],

want [ANTWOORD].

De tweede positie die belangrijk is, is [ANTWOORD] **want [**ANTWOORD].

Het samenwerken van enzymen is vergelijkbaar met het samenwerken van een teamsport want:

Elk enzym heeft een specifieke rol bij het afbreken van voedsel, vergelijkbaar met verschillende spelers in een sportteam die elk een specifieke taak hebben. Net zoals verschillende spelers in een team samenwerken om een doel te bereiken, werken enzymen samen om voedsel af te breken tot bruikbare componenten die het lichaam kan gebruiken **(1pt)**