Answering questions after initial study guides attention during restudy

THIS ARTICLE IS PUBLISHED IN INSTRUCTIONAL SCIENCE UNDER

DOI 10.1007/s11251-014-9330-9

RUNNING HEAD: THE TESTING-EFFECT AND RESTUDY BEHAVIOR

Answering Questions after Initial Study Guides Attention during Restudy

Kim. J. H. Dirkx[a], Gun-Brit Thoma[b], Liesbeth Kester[a], and Paul. A. Kirschner[a]

[a] CELSTEC, Open University in the Netherlands, Heerlen, the Netherlands

[b] Leibniz Institute for Science and Mathematics Education, Kiel, Germany

Author Note

Correspondence concerning this article should be addressed to Kim J. H. Dirkx, Open University in the Netherlands, CELSTEC, NL - P.O. Box 2960, 6401 DL, Heerlen, the Netherlands. E-mail: kim.dirkx@ou.nl.

Abstract

Though the testing-effect can be boosted by including a restudy phase after answering test questions, we do not know precisely why it does so. One possible explanation is being tested here. The present study measured attention allocation during initial reading and rereading with a remote eye tracker to gain information on the cognitive processes during restudy, with and without prior testing. The results show that at the final study moment, students in the study-test-condition attended longer to information pertaining to the initial test questions as compared to students in the study-only condition (i.e., who did not take the test). No differences in attention allocation were found for information only questioned on a posttest one week later. In addition, it was found that performance on the initial test questions heavily affected which information students restudy; students in the study-test-condition paid namely more attention to the answers of questions they answered incorrectly during the initial test than to the answers of the questions they answered correctly on the initial test.

Keywords: testing-effect, restudy behavior, eye tracking.

Answering Questions after Study Guides Attention during Restudy

Traditionally, tests and examinations are used as summative assessment tools that measure what a student has learned (cf. Harlen, 2006). However, memory researchers have found that retrieving information during learning by means of testing enhances retention significantly more than spending a comparable amount of time rereading the information. This effect of taking tests to enhance memory is well-known as the testing-effect and has a long research history (Gates, 1917; see Karpicke & Grimaldi, 2012; see Roediger & Karpicke, 2006; see Roediger & Butler, 2011; see Roediger, Putnam, & Smith, 2011; Spitzer, 1939).

Whereas many testing-effect studies used research designs in which a single test phase was compared with a single restudy phase, quite a number of testing-effect studies also included a restudy phase *after* initial testing to boost the testing-effect (e.g., Gates, 1917). However, the cognitive processes which cause this boosting have not been well investigated. The present study tries to shed some light on the question why restudy boosts the effect of retrieval thereby investigating if attention allocation plays a role by making use of eye tracking.

The Testing-effect with Restudy

One of the earliest studies on the testing-effect has been conducted by Gates (1917). He gave students biographies and syllables and asked them to study this material. Some of the students were then tested, others were asked to keep on studying. Of the students who were tested, some of them were given the opportunity to look back at the materials after they were being tested. Gates found that restudy after initial testing plays an important role in boosting the beneficial effect of testing and, consequently, quite a number of studies exploring the testing-effect in real educational settings have included a restudy phase (as Gates) after initial testing in their designs (e.g., Karpicke & Roediger, 2007).

To explain why restudy after testing boosts the testing-effect, some researchers have suggested that giving students test questions prior to or during learning guides future study

behavior (e.g., Andre, 1979; Hamilton, 1985; Karpicke, 2009; Karpicke & Roediger, 2007; Kornell, Hays, & Bjork, 2009; Pyc & Rawson, 2010; Rothkopf, 1966). That is, questions are posited to have a potentiating effect (see Roediger et al., 2011) and to help students establish a "search set" (Raaijmakers & Shiffrin, 1981) aiding them in allocating their limited attentional resources during study. Rothkopf called this effect of questions on restudy behavior "*mathemagenic behavior"* (literally 'behavior that gives birth to learning'). It implies that test questions given before or during (re)study are beneficial for learning because they cue students' attention towards a specific type of information and/or the answers on the initial test questions (i.e., targeted information), and, they help (i.e., cue) students to ignore information irrelevant for the intended purpose (i.e., non-targeted information). Thus, answering questions is not only seen as a strategy having a direct effect on learning (i.e., the retrieval process changes memory), but also an indirect effect. It guides learners to examine specific instructional material better which results in improved learning of that material, but not of non-questioned material (cf. Roediger & Karpicke, 2006; Roediger et al., 2011).

A related explanation comes from the 'Productive Failure' literature (Kapur & Toh, 2013; Kapur, 2014). Productive failure refers to the finding that trying to solve problems before instruction, leads to better problem solving and transfer as compared to direct instruction where the information is presented before trying to solve the problem. An important prerequisite is however that instruction is given after attempted problem solving. One of the explanations for productive failure is that it helps them better to prepare for subsequent instruction. Another explanation is that students better process the instruction because they are less confident about their knowledge when first trying to solve problems as compared to first studying the material.

Processes during Restudy

Although the research on mathemagenic behavior gives some information on the potentiating effect of testing for learning, it has not investigated the online study processes

during restudy. Instead it looked at the effect of answering a specific type of question on subsequent performance and draws inferences from those results about the processes that might explain such effects. Research on online reading-behavior shows, however, that answering questions *during* reading (i.e., when text and questions are presented simultaneously) indeed affects the online reading process (Britt & Angliskas, 2002; Cerdán, Vidal-Abarca, Martínez, Gilabert, & Gill, 2009; Goldman & Rakestraw, 2000; Roll, Holmes, Day, & Bonn, 2012). When students are given questions during study, their learning goal is to answer the given questions and this goal steers subsequent attention allocation. Prior research in which students answer factual knowledge questions *with the text present* shows, for example, that students search only a small number of *paragraphs* per question to find the answers to those questions (e.g., Roll, et al.,; Vidal-Abarca, Gilabert, & Rouet, 1998). Hyönä, Niemi, and Underwood (1989), on the contrary, found that rereading a text without answering questions prior to or during that reading leads to quicker reading behavior with less detailed attention to specific information (i.e., scanning the whole text instead of detailed rereading).

Performance on Initial Tests as a Mediator for Restudy Behavior

Although it thus seems that posing questions *during reading* has an effect on restudy behavior, Carrier and Pashler (1992) suggested that restudy is particularly effective when initial retrieval is rather low and/or inaccurate, indicating that not only giving test questions to students or not giving questions, but also initial performance on these questions, affects attention allocation. This is confirmed by research of, for example, Gurthie (1972) which shows that feedback (i.e., restudy) has almost no effect on initially correct answers, but can increase subsequent performance up to 400% for initially incorrect answers. In addition, research on study-time allocation - where students have to deliberately select the material they want to restudy - shows that, when students are asked to select material for restudy, they generally choose material that they believe they have not very well mastered and pay less attention (i.e., do not select) to material they believe they have already mastered (e.g., Bisanz,

Vesonder, & Voss, 1978; Karpicke, 2009). In other studies, however, students have selected material for restudy which they believed to be relatively easy (i.e., information that was retrievable), probably because they want to rehearse this material (e.g., Metcalfe & Kornell, 2003; Thiede & Dunlosky, 1999; Son & Metcalfe, 2000) and this seemed much more beneficial than dropping material from further study (Karpicke & Roediger, 2007).

Research Questions

Restudy plays an important role in boosting the beneficial effects of retrieval practice (e.g., Gates, 1917). In addition, in educational practice, most students will choose to restudy a text after initial testing (Karpicke, et al., 2009). Thus, investigating what actually happens during such a restudy phase could yield interesting and important information when aiming to implement the testing-effect in schools. Although there is some research which shows that answering questions *during reading* leads to specific search behavior and increases attention to goal related material, no such research has been conducted for the testing-effect, where the most interest has been in the retrieval function of testing and questions are presented *prior* to subsequent restudy. The present study is therefore an attempt to investigate if mathemagenic behavior can also explain the boosting effect of restudy for the testing-effect. More specifically, the study explores if answering questions prior to restudy – as compared to not answering such questions – affects attention allocation during subsequent study and if and how *performance* on those questions is of influence. For this purpose eye tracking is used. As compared to prior research (i.e., Rothkopf & Bisbicos, 1967), the present study investigated *online* study behavior *after* answering initial test questions. The following hypotheses are tested:

H1) During initial study, there are no differences in attention allocation between conditions since no questions were given prior to initial study (i.e., manipulation check).

H2a) During restudy, students in the study-test-condition will focus longer on the

information asked for in the initial test questions (i.e., targeted information) as compared to students who are not tested (i.e., study-only condition).

H2b) No attention allocation differences are expected to occur for information not covered by the initial test questions (i.e., non-targeted information). In other words, students in the study- test-condition will pay the same amount of attention to non-targeted information than students in the study-only-condition.

<u>For the testing-condition, the following additional hypotheses are tested:</u>

H3a) Students in the study-test-condition use the restudy phase to focus on incorrectly recalled information

H3b) Students in the study-test-condition use the restudy phase to rehearse correctly recalled information.

H4) If students in the study-test-condition focus longer on the information they initially answered incorrectly, then this will lead to increased performance on those questions (i.e., a high percentage of 'learned items'; as has been suggested by for example Gurthie, 1971) whereas when students use the restudy phase to rehearse correctly retrieved information, this will lead to a high percentage of remembered items.

Method

Participants

For this experiment, 56 students in 7th or 8th grade of a secondary academic track school (Gymnasium) in Germany (50% males) who volunteered to participate (*Mean Age* = 13.16 years*; SD* = .69 years) were included when they did not wear contact lenses or glasses and calibration was sufficient (calibration scores > .08). Students were paid 20 Euros for participation at the end of the experiment.

Apparatus and Materials

Eye-tracker. The complete experiment ran on a video based remote eye tracking device (RED) from Sensomotoric Instruments (SMI) set at 120 Hz. This infrared camera was

fixated under the screen of the stimulus PC, located in a recording room at the university and enabled the investigation of reading behavior and attention allocation on a very detailed level (Van Gog & Scheiter, 2010).

Expository text. An expository text of 766 words on a – for the students – unfamiliar mathematics topic (i.e., probability calculations) was used as stimulus in this experiment (see appendix I). The text was used in other experiments (XXX, et al., 2013) and translated into German. It was written in such a way that all students should be able to comprehend the text. The text contained five short paragraphs (*Mean length* = 180 words).

Factual knowledge test. The factual knowledge test was comprised of five questions (see appendix I for example questions). These questions have been used in a prior experiment (XXX, et al., 2013) and were translated into German. For each paragraph one factual knowledge question was constructed. The answers to the factual knowledge questions could be found literally in the text and comprised maximally one sentence. A pilot study showed that, on average, students were able to answer about 50% of the questions correctly after having studied the text (i.e., on a posttest), indicating that the questions were of moderate difficulty.

Posttest. The posttest was comprised of the five factual knowledge questions in the initial test (i.e., repeated questions) expanded with one new factual knowledge question per paragraph (i.e., 5 in total). These new questions were not related to the initial test question (see appendix I for example questions).

Scoring

Factual knowledge test and posttest questions. A full point was awarded to each correctly answered factual knowledge question on the initial test or the posttest. When the questions consisted of two partial answers, 0.5 points were awarded for each partial correct answer. For example, the answer to the question – *"There are different situations for which you can calculate probabilities. For each situation, there is a different procedure that you*

need to follow. What was the first situation explained in the text?" was awarded 0.5 points for *"Without putting the marbles back in the vase"* and 0.5 points for *"taking the marbles one by one"*. Zero points were awarded for missing or wrong answers. The total score was divided by the total number of points that students could attain for that test (i.e., 5 for the initial test, 5 for repeated posttest questions, and 5 for new posttest questions). A total of 29 (51.78%) posttests were scored by two raters with very high inter-rater agreement (*ICC* = .99 for the repeated questions and 1.0 for the new questions).

To yield information on the effect of restudy on posttest performance, the scores on the initial test questions were related to posttest performance scores (per question). When students in the study-test-condition answered a question correctly on both the initial test and the posttest (i.e., correct-correct or c-c) o, the question was scored as remembered. When students answered a question correctly on the initial test but incorrectly (i.e., i = incorrect) on the posttest, the question was scored as forgotten (c-i). Questions answered incorrectly on both tests were scored as not learned (i-i). Finally, questions answered incorrectly on the initial test, but correctly on the final test, were scored as learned (i-c).

Eye tracking analyses

For each answer on the initial and final test, an AOI was defined in the expository text. The AIOs were between one word and one sentence long and comprised no more than 5% of the expository text. For example, the answer to questions 1 *("What is the name of the inventor of the probability definition?")* in the initial test was Pierre so the name Pierre was defined as AOI in the expository text. In line with the choices made for the test questions, the AIOs were evenly distributed over the 5 text sections.

Figure 1 presents an example of the data generated by the eye tracker. The red circles represent fixations, thus were a student looks at for a period of time. With the eye tracker, not only the location, but also the duration (in milliseconds) of the gaze can be recorded. The longer a student looks at a certain area, the larger the fixation circles. The lines between the

circles represent the saccades; the eye movements between fixations. During saccades no information is assumed to be processed. Taking the durations of all fixations and saccades that hit the areas of interest (i.e., AIOs) together leads to the dwell and provides a measure of the duration (in milliseconds) that students look at certain AOIs (i.e., in this case the answers to the initial test questions and posttest questions).

** insert Figure 1 about here**

There were two types of AOIs, namely answers to initial test questions (i.e., targeted AOIs) and answers to new posttest questions (i.e., non-targeted AOIs). *T*-tests were calculated for dwell time on AOIs between the conditions. For the study-test-condition, dwell time on questions that were initially answered correctly and incorrectly were compared with a paired sample *t*-test.

Design and Procedure

The experiment took place at the university where a recording room was specially prepared to collect eye tracking data (i.e., the windows were covered to prevent sunlight falling into the room and a room divider separated the participant and experimenter). The recording room was located in a quiet area of the building to prevent disruptions and noise. For the study, students were randomly assigned to one of the two conditions (see Table 1).

** Insert Table 1 about here **

For the design, previous testing-effect research was used as an example. In the control condition, students studied the text three times (i.e., study-only-condition; SSS) while in the experimental condition students studied the text, took a test (i.e., were asked to retrieve the information needed to answer the question from memory), and then restudied the text (i.e., study-test-condition; STS).

Before the experiment began, students were briefly instructed about the experiment and signed an informed consent form. This was followed by the calibration of the eye tracking apparatus (see Van Gog, Paas, and Van Merriënboer, 2005). The experimenter asked the

participant to assume a comfortable position and keep this position during the experiment. After successful calibration (calibration values below .5) the experiment began with the following experimenter instruction: *"You will read an expository text consisting of five paragraphs which will take approximately 8 minutes. The computer will automatically go to a subsequent paragraph. After reading the text, you will either be given a test or will be asked to reread the text. The computer will decide randomly which it will be. Either way, you will be tested on the text next week. This test will contain both factual knowledge questions as well as application questions"*. This instruction was the same for all students independent of condition as prior research has shown that test expectation can affect students' performance (Kausler, Laughlin, & Trapp, 1963; Kausler & Trapp, 1962; McDaniel, Blischak, & Challis, 1994).

Students in both conditions studied each of the five paragraphs of the text (range = 131 - 266 words) during 90 seconds. The total reading time was thus 8 minutes (480,000 milliseconds) per study phase. Reading fluency scores (Hasbrouck & Tindal, 2006) showed that this should be sufficient time for the students to carefully read each paragraph (i.e., 133 words per minute in grade 8). During the initial reading phases, the computer automatically went to the subsequent paragraph, whereas during the final reading phase (the final phase in both conditions), students were presented the complete text and could scroll through it as they liked. After the initial study phase, students in the study-only-condition restudied the text in two consecutive study phases of each 8 minutes. Each study phase in this condition was separated by a 2-minute distracter task (i.e., a visual search task) after which the students were recalibrated. Students in the study-test-condition on the other hand were given the factual knowledge test after initial study and then restudied the text a final time (for 8 minutes). Here too, each learning phase was followed by the distracter task and recalibration. The students were instructed to keep on reading during the complete 8 minutes

and no explicit feedback was given to the students after testing. One week after the learning phase, all students completed the posttest (13 minutes).

Results[1]

Analyses between Conditions. An independent-sample *t*- test[2] revealed no significant differences in dwell time between the two conditions during the first reading for both the targeted and non-targeted AOIs ($t(47) = 1.70$, $p = .10$, $r = .24$) and $t(47) = .20$, $p = .84$, $r = .23$). Students in the study-test-condition focused equally long on targeted AIOs and non-targeted AOIs as compared to students in the study-only-condition (see Table 2 for descriptives). However, on the final reading, students in the study-test-condition looked significantly longer at the targeted AOIs as compared to students in the study-only-condition ($t(46) = 2.06$, $p = .05$, $r = .29$) indicating that test questions guide students' attention towards the information covered by the those question. As expected, no dwell time differences between the two conditions were found for non-targeted AOIs ($t(46) = .07$, $p = .95$, $r = .01$). Students in the study-test-condition looked, as expected, equally long at the non-targeted AOIs as students in the study-only-condition.

** Insert Table 2 about here **

On the posttest, the two conditions significantly differed on the repeated questions ($t(54) = 2.27$, $p = .027$, $d = .60$). Students in the study-test-condition performed significantly better on the repeated questions of the posttest than students in the study-only-condition (see Table 3 for descriptive statistics). As expected, no significant differences were found between the two conditions on the new factual knowledge questions of the posttest ($t(54) = 1.70$,

[1] Due to imprecise eye tracking data the results of 8 participants during the first reading and 6 participants during the final reading could not be analyzed.

[2] Independent sample *t*-tests were used because there was no significant correlation between dwell time on targeted and non-targeted AOIs and between performance on identical and new posttest questions.

p = .096, r = .23). Students in the study-test-condition did not score significantly higher or lower on the new factual knowledge questions than students in the study-only-condition.

** Insert Table 3 about here **

Analyses Study-Test-Condition. Students in the study-test-condition answered on average 14% (*SD* = 15) of the answers correctly on the initial test. Comparing dwell time between areas containing the answer to the question that were initially answered correctly and those initially answered incorrectly with a paired-sample *t*-test shows that students looked longer at the answers to the questions they initially answered incorrectly than at answers to the questions they initially answered correctly ($t(23) = 5.96$, $p < .001$, $r = .78$). Students looked, on average, 1.29% (*SD* = 1.85) of the time at the answers to the questions they initially answered correctly and 6.35% (*SD* = 2.99) of the time at the answers to the questions they initially answered incorrectly (see also Table 3 for descriptive statistics).

Comparing performance on the initial test and the posttest showed that students in the study-test-condition mainly improved on the items they initially answered incorrectly (61.48% of the learned information or i-c). However, they forgot more than 20% of the answers which they had correctly given on the initial factual knowledge test (i.e., c-i) and also remembered (c-c) relatively few items (see Table 4).

** Insert Table 4 about here **

General Discussion

In the present study, the boosting effect of restudy after initial testing was studied by investigating attention allocation. For this purpose, students either studied a text, made a test on that text, and then restudied the text (i.e., study-test-condition) or they read the text three times (study-only-condition) without testing. The hypotheses are revisited here one by one.

H1) During initial study, there are no differences in attention allocation between conditions since no questions were given prior to initial study (i.e., manipulation check).

This hypothesis has been confirmed in the present study. Students in both conditions looked about equally long at the AIOs which marked the answers to the initial and final test questions which were not yet presented during initial reading.

H2a) During restudy, students in the study-test-condition will focus longer on the information asked for in the initial test (i.e., targeted information) as compared to students who are not tested (i.e., study-only condition).

This hypothesis has been confirmed in the present study. The students in the study-test-condition looked longer at the information that was covered by the test questions as compared to students who were not initially tested. It thus seems that initial testing does guide students' subsequent reading behavior. These findings are in line with prior research showing that questions affect the reading process (e.g., Cerdán et al., 2009) even if the questions are provided prior to the restudy phase.

H2b) No attention allocation differences are expected to occur for information not covered by the initial test questions (i.e., non-targeted information). In other words, students in the study- test-condition will pay the same amount of attention to non-targeted information than students in the study-only-condition.

This hypothesis has been confirmed in the present study and provides confirmation for hypothesis 2a.

H3a) Students in the study-test-condition use the restudy phase to focus on incorrectly recalled information.

This hypothesis has been confirmed. Students seem to use the restudy phase especially to focus on the information they did not know during prior testing. These findings are in line with research by, for example, Gurthie (1971) who found that feedback (in this case restudy) is especially beneficial for initially incorrectly recalled information and has little effect on initially correctly recalled information.

H3b) Students in the study-test-condition use the restudy phase to rehearse correctly recalled information.

This hypothesis has been only partially confirmed. Although students revisit the information they answered correctly on an initial test, they have more attention for information they answered initially incorrect.

H4) If students in the study-test-condition focus longer on the information they initially answered incorrectly, then this will lead to increased performance on those questions (i.e., a high percentage of 'learned items'; as has been suggested by for example Gurthie, 1971) whereas when students use the restudy phase to rehearse correctly retrieved information, this will lead to a high percentage of remembered items on the final test.

In line with hypothesis 3a, students in the study-test-condition had a very high percentage of learned items which supports the conclusion that students use the restudy phase to correct initial errors.

In sum, the results of the present study show that there is an attentional effect of giving test questions prior to restudy. In addition, performance on the initial test questions seems to influence attention allocation during restudy. In that sense, test questions appear to have a strong diagnostic and formative function. By being tested, students get an idea of how well they learned the information (i.e., testing can thus be seen as a diagnostic tool) and they use this information to steer their subsequent reading. In addition, testing with restudy leads to higher performance on a posttest as compared to reading-only.

An implication from these findings for classroom practice is that problem solving before instruction is helpful for learning and that instruction after attempted problem solving is a necessity, especially when initial problem solving fails (see also Kapur & Toh, 2013; Kapur, 2014). We would recommend to consider and use problem solving as a formative testing opportunity in educational design which fosters learning instead of an activity that assesses how much students have learned.

Although the present research sheds some light on how test questions actually affect restudy behavior, dwell time on answers to initial test question was only a small fraction of total reading time. However, considering that the AOIs (i.e., the answers to the initial test questions) made only about 5% of the total text, one might conclude that attention effects have taken place. Another limitation of the present study is that the study reported here cannot give causal explanations about the contribution of retrieval practice versus attention allocation to performance because students in the study-test-condition both retrieved information *and* attended (longer) to the tested facts. For future research it might be informative to compare a condition in which students are given only targeted information in the form of statements - to control attention allocation during restudy - whereas another condition practices overt recitation to control retrieval practice and compare the effects achieved in both conditions. A final limitation is that the study presented here only looked at recall of facts whereas, for educational practice, it would also be interesting to look at the effects of attention in

application and transfer of information and to do this in a context where the testing is more included in an authentic situation because of possible motivational aspects.

Nevertheless, the present study is to our knowledge the first study which uses eye tracking to gain more insight into the cognitive processes that might explain the testing-effect and it shows that testing affects attention allocation during restudy and that restudy especially plays a role when initial test performance is incorrect.

References

Andre, T. (1979). Does answering high-level questions while reading facilitate productive learning? *Review of Educational Research, 49*(2), 280-318. doi: 10.3102/00346543049002280

Bisanz, G. L., Vesonder, G. T., & Voss, J. F. (1978). Knowledge of one's own responding and the relation of such knowledge to learning. *Journal of Experimental Child Psychology, 25,* 116-128. doi: 10.1016/0022-0965(78)90042-5

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20,* 633-642. doi: 10.3758/BF03202713

Cerdán, R., Vidal- Abarca, E., Martínez, T., Gilabert, R., & Gill, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning & Instruction, 19,* 13-27. doi: 10.1016/j.learninstruc.2007.12.003

XXX, XXX & XXX (2013). XXXXXXXXX, *XXXXXXX*, XX-XX

XXX, XXX, & XXX (2014). *XXXXXXXXXXX*, XXXXX

Gates, A. I. (1917). Recitation as a factor in memorizing. *Archives of Psychology, 6,* (40).

Goldman S., & Rakestraw, J. (2000). Structural aspects of constructing meaning from text. In M. Kamil , P. Mosenthal, P. D. Pearson, R. Barr (Eds.), *Handbook of Reading Research,* Vol III (311-335). New York, United States: Lawrence Erlbaum Associates.

Guthrie, J. T. (1971). Feedback and sentence learning. *Journal of Verbal Learning and Verbal Behavior, 10,* 23-28. doi: 10.1016/S0022-5371(71)80088-9

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research, 56*, 212-242.

Hamilton, R. J. (1985). A framework for the evaluation of the effectiveness of adjunct questions and objectives. *Review of Educational Research, 55,* 47-85. doi: 10.3102/00346543055001047

Harlen, W. (2006). On the relationship between assessment for summative and formative purposes. In: J. Gardner (Ed.). *Assessment and learning* (pp. 103-118). London, England: Sage Publications ltd.

Hasbrouck, J., & Tindal, G. A. (2006). Oral Reading Fluency Norms: A Valuable Assessment Tool for Reading Teachers. *The Reading Teacher, 59,* 636–644. doi: 10.1598/RT.59.7.3

Hyönä, J., Niemi, P., & Underwood, G. (1989). Reading long words embedded in sentences: Informativeness of word halves affects eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 142-152. doi: 10.1037/0096-1523.15.1.142

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science, 38*, 1008-1022. doi: 10.1111/cogs.12107

Kapur, M., & Toh, P. L. L. (2013). Productive failure: From an experimental effect to a learning design. In T. Plomp, & N. Nieveen (Eds.), *Educational design research – Part B: Illustrative cases* (pp. 341-355). Enschede, the Netherlands: SLO.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138,* 469-486. doi: 10.1037/a0017341

Karpicke, J. D., Butler, A. C., & Roediger, H. L., III (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory, 17,* 471-479. doi: 10.1080/09658210802647009

Karpicke, J. D., & Grimaldi, P. J. (2012). Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review, 24,* 401-418. doi: 10.1007/s10648-012-9202-2

Karpicke, J. D., & Roediger, H. L., III (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57,* 151-162. doi: 10.1016/j.jml.2006.09.004

Kausler, D. H., Laughlin, P. R., & Trapp, E. P. (1963). Effects of incentive-set on relevant and irrelevant (incidental) learning in children. *Child Development, 34*(1), 195-199. Retrieved from: http://www.jstor.org/stable/1126839.

Kausler, D. H., & Trapp, E. P. (1962). Effects of incentive-set and task variables on relevant and irrelevant learning in serial verbal learning. *Psychological Report, 10,* 451-457. Retrieved from: http://www.amsciepub.com/doi/pdf/10.2466/pr0.1962.10.2.451

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 989-998. doi: 10.1037/a0015729.

McDaniel, M. A., Blischak, D. M., & Challis, B. (1994). The effects of test expectancy on processing and memory of prose. *Contemporary Educational Psychology, 19*(2), 230-248. doi:10.1006/ceps.1994.1019

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy. Effective and portable. *Psychological Science, 20,* 516-520. doi: 10.1111/j.1467-9280.2009.02325.x

Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General, 132,* 530–542.

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science,* 333-335. doi: 10.1126/science.1191465

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93-134. doi: 10.1037/0033-295X.88.2.93

Roediger, H. L., III & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Science, 15,* 20-27. doi: 10.1016/j.tics.2010.09.003

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

Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre & B. Ross (Eds.), *Psychology of learning and motivation: Cognition in education* (pp. 1-36). Oxford: Elsevier. doi 10.1016/B978-0-12-387691-1.000

Roll, I., Holmes, N. G., Day, J., & Bonn, D. (2012). Evaluating metacognitive scaffolding in guided invention activities. *Instructional Science, 40,* 691-710. doi:10.1007/s11251-012-9208-7

Rothkopf, E. Z. (1966). Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. *American Educational Research Journal, 3,* 241-249. Retrieved from http://www.jstor.org/stable/1162033.

Rothkopf, E. Z., & Bisbicos, E. E. (1967). Selective facilitative effects of interspersed questions on learning from written materials. *Journal of Educational Psychology, 58*(1), 56-61. doi: 10.1037/h0024117

Sagerman, N., & Mayer, R. E. (1987). Forward transfer of different reading strategies evoked by adjunct questions in science texts. *Journal of Educational Psychology, 79,* 189-191. doi: 10.1037/0022-0663.79.2.189

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 204-221. doi: 10.1037/0278-7393.26.1.204

Spitzer, H. F. (1939). Studies in retention. *The Journal of Educational Psychology, 30,* 641-656. doi:10.1037/h0063404

Thiede, K. T., & Dunlosky, J. (1999). Toward a General Model of Self-regulated Study: An Analysis of Selection of Items for Study and Self-paced Study Time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1024-1037.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology, 19,* 205-221. doi: 10.1002/acp.1112

Van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*(2), 95-99. doi: 10.1016/j.learninstruc.2009.02.009

Vidal-Abarca, E., Gilabert, R., & Rouet, J. F. (July, 1998). *The role of question type on learning from scientific text.* Paper presented at the Meeting on Comprehension and Production of Scientific Texts, Aveiro, Portugal.

Appendix I

The probability definition contains different symbols. The symbol *P* for example stands for probability and *G* stands for the situation for which you want to calculate the probability, such as the situation of throwing two (sum is two) with two dice. How do you calculate this probability? When you throw two dice, there are 6X6 = 36 possible outcomes (sum scores), because every dice has six sides, each with a different number of dots. For the situation *(G)* 'sum is two' there is only one possibility, namely that you throw one with the first dice and one with the second. Thus, there is only one possible outcome for 'sum is two' and 36 possible outcomes of throwing two dice. According to the formula *P*(sum is two) = 1/36.

Initial test question	**Posttest question**
1) What does the symbol G in the formula stand for?	1) In the probability definition the symbol *P* is included. Wherefore stands *P*?