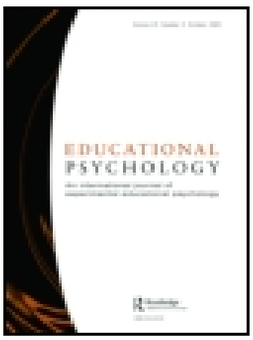

Routledge
Taylor & Francis Group

Educational Psychology

An International Journal of Experimental Educational Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cedp20

# Fewer students are benefiting from doing their homework: an eleven-year study

Arnold L. Glass & Mengxue Kang

**To cite this article:** Arnold L. Glass & Mengxue Kang (2020): Fewer students are benefiting from doing their homework: an eleven-year study, Educational Psychology, DOI: 10.1080/01443410.2020.1802645

**To link to this article:** https://doi.org/10.1080/01443410.2020.1802645

EDUCATIONAL PSYCHOLOGY
https://doi.org/10.1080/01443410.2020.1802645

Routledge
Taylor & Francis Group



# Fewer students are benefiting from doing their homework: an eleven-year study

Arnold L. Glass and Mengxue Kang

Rutgers University, Piscataway, NJ, USA

**ABSTRACT**

Performance on homework questions was compared with performance on related exam questions querying the same fact or principle, was used to assess the effect of answering online homework questions on subsequent exam performance. A distinctive pattern of performance was found for some students in which superior performance on online homework questions resulted in poorer exam performance. When assessed over an eleven-year period, for 2433 students in 12 different college lecture courses, the percent of students who did **not** benefit from correctly answering homework questions increased from 14% in 2008 to 55% in 2017. During the most recent two years of the study, when students were asked how they did their homework, students who benefitted from homework reported generating their own answers and students who reported copying the answers from another source did not benefit from homework.

**ARTICLE HISTORY**
Received 13 February 2019
Accepted 24 July 2020

**KEYWORDS**
Generation effect; testing effect; long-term memory

This study is the product of an ongoing project to use technology to monitor academic performance and to assess the effects of new instructional technologies on academic performance. The purpose of this study was to investigate the relationship between performance on homework assignments and exam performance. The project is motivated by a specific hypothesis about why and how cognition evolved, which provides the ultimate explanation for the relationship between homework strategy and exam performance. To motivate what follows, this hypothesis is presented briefly before describing details of the study.

Cognition begins with action. The entire purpose of cognition is to make action more effective. In humans, indeed in all mammals, two distinct cognitive systems work together to make action effective: the improvisational system generates ad hoc responses to novel targets, the habit system retrieves previously effective responses to familiar targets (Glass, 2016, 2019).

The improvisational system can generate an ad hoc response to a novel target in one of two ways. One way is to create the response by oneself. For example, to break open a particularly sturdy nut one tries striking it with a particularly sturdy stick

**CONTACT** Arnold L. Glass aglass@rutgers.edu Rutgers University, 152 Frelinghuysen Road, Piscataway, NJ 08854, USA

(Visalberghi et al., 1995). The other way is to imitate the response of another. For example, if you see a member of your community crack a nut by throwing it against a rock then you do the same. The brains of apes, including humans, are especially equipped for imitation by 'mirror neurons' that make it possible to accurately reproduce a seen action (Cook & Bird, 2013).

Whether a response has been created or observed has important consequences for whether it is remembered. This was established in a seminal study by Bobrow and Bower (1969). It was known that students were more likely to remember a pair of unrelated words, e.g. *car – tree*, if they were asked to create a sentence linking the two words, e.g. *the car crashed into the tree*, than if they were asked to remember the word pair (Bower & Clark, 1969). This result raised the issue of whether the sentence, *the car crashed into the tree*, was intrinsically more memorable than the pair of unrelated words, *car – tree*, or whether it was the act of creating the sentence that made it more memorable. To determine which explanation was the case, Bobrow and Bower performed a yoked control experiment in which one group of students, the generators, had to generate a linking sentence for each pair of words. Yoked to each generator was a control student who saw the linking sentence created by the generator. Bobrow and Bower (1969) found that generating one's own linking sentence resulted in retention of twice as many paired-associates (60%) than using a linking sentence generated by someone else (30%). Since then the generation effect has been well-established by the findings of 130 subsequent studies and remains an active area of research (e. g. McCurdy et al., 2017). In a variety of tasks, self – generated responses are remembered better than responses that are copied. One specific instance of the generation effect relevant to education is the testing effect. Answering a question about a just-read text produces better long-term retention of the text than extended study does (Pan & Rickard, 2018). In their recent review and meta-analysis of the experimental research on the testing effect, Pan and Rickard identified elaborative rehearsal as a causative factor for the testing effect. Elaborative rehearsal is the response an individual makes to new information connecting that information in a meaningful way with what the individual already knows.

The universal use of smartphones has produced both an opportunity and a challenge for academic instruction. This report will focus on the challenge. One challenge is that when students divide their attention in class between instruction and an unrelated activity on their cell phone (or other electronic device) elaborative rehearsal of the study material is suppressed because instead voluntary action is directed towards the unrelated activity. An important function of elaborative rehearsal by the improvisational system is to create a long-term representation of the response in the habit system as well (Glass, 2019). Consequently, when students divide their attention in class between instruction and an unrelated activity on their cell phone their exam performance is ultimately impaired (Carter et al., 2017; Glass & Kang, 2019; Sana et al., 2013).

The demonstrated effect of cell phone use on classroom learning raises the issue of whether cell phone use mediates the effect of homework on learning as well. Students are unconstrained in their cell phone use while doing homework. This report addresses this issue.

When answering homework questions, students have an opportunity that is not usually present in the classroom. The student can generate an answer to the question from memory, which would include guessing the answer. Alternatively, the student could use their cell phone to look up the correct answer. Within the framework that Pan and Rickard (2018) provide, these alternative actions have dramatically different effects on the long-term retention of the correct answer.

Within the framework that Pan and Rickard (2018) provide, answering homework questions ultimately increases exam performance because the retrieval of the answer has a mnemonic effect that increases long-term retention of the answer, hence the probability of retrieving it on the exam. Furthermore, answering a question provides the opportunity for elaborative retrieval, the association of the question with other relevant information. Pan and Rickard found that elaborative feedback increased transfer to *different* questions that had the same answer. Notice that since homework and exam questions are rarely identical, the elaborative rehearsal plays a crucial role in increasing long-term exam performance.

On the other hand, within Pan and Rickard (2018) framework, finding the answer to a question on the internet is equivalent to Bobrow and Bower's (1969) control condition, called the restudy control condition by Pan and Rickard. Copying an answer found on the internet does not involve the mental generation of an answer. In the context of academic study, the implication of these laboratory resuls is that a student who self-generated the answer to a homework question would be more likely to select the correct answer when subsequently asked a paraphrase of the question on an exam more than a week later than a student who found the answer on the internet.

Modern technology has made it possible to test this prediction of the testing effect during actual academic instruction. The availability of personal response systems (clickers) and online instructional platforms has made it possible to employ distributed questioning as an instructional methodology, providing a natural testbed for assessing the effects of answering earlier questions on performance for later ones (Glass & Sinha, 2013).

In the distributed questioning instructional methodology, for each fact-statement and principle tested on the exam, four questions are created such that the answers to all four questions are implied by the same fact-statement or principle. Each question from the set of four related questions is presented once during the semester. A Pre-lesson question is asked based on the homework reading assignment the day before the classroom lesson that contains the answer. The Post-lesson question is asked during the same classroom session that contains its answer shortly after the answer as been presented as part of the lesson. The Review question is asked at least a week after the lesson containing its answer. The Exam question appears on the unit exam (Glass & Sinha, 2013).

The purpose of the distributed questioning instructional methodology is to incrementally increase accurate performance through distributed repetition of the related questions so that ultimately the best performance is on the exam. Consistent with this purpose, in all studies performance was worst on the Pre-lesson questions and best on the Exam questions (Glass & Sinha, 2013). In some of the studies percent correct

on Review questions was significantly better than on Post-lesson questions (McDaniel et al., 2013; McDermott et al., 2014) but in the other studies percent correct on Review questions was no better or worse than on Post-lesson questions (Glass, 2009; McDaniel et al., 2011; Roediger et al., 2011), presumably because Post-lesson question performance benefitted from the fact that Post-lesson questions were asked in class shortly after the statement during that class providing the answer. In contrast, Review questions were asked at least a week or more later without the benefit of classroom review of the answer.

In most studies of distributed questioning (McDaniel et al., 2011, 2013; McDermott et al., 2014; Roediger et al., 2011), all four questions were presented in the classroom. However, the first author of this report has presented the Pre-lesson and Review questions as online homework questions in his college lecture course every year from 2008 through 2018. This provided an opportunity for examining the effect of homework on exam performance. The testing effect provided a framework for the analysis and provided an a priori prediction. Consider two possible strategies for completing an online homework assignment consisting of a set of multiple-choice questions. One strategy would be to retrieve the correct answers to the questions in advance of responding in order to maximise the score on the assignment. Recent technical advances have decreased the effort required by this strategy. The student might google a key phrase from the question, or use it to search an online version of the text, or text a fellow student a request for the answer. Notice that the first two of these actions are perfectly legitimate methods a dedicated student might use to do well on the homework assignment and to learn the material.

Another strategy would begin with the student generating an answer to each question without first checking the text, personal notes or an online source. The student might go on to check low-confidence responses using the same methods as the first strategy. However, the defining feature of this strategy is first generating an answer before it is checked. In the courses comprising this study, submitted quizzes were immediately graded. A student who submitted a quiz for grading without checking any of the answers received immediate feedback indicating the correct answer and the page in the textbook where the rationale for the answer could be found. Furthermore, by design, there was not a cost to the student for submitting a homework quiz with wrong answers in the calculation of the course grade.

Consider the implication of the testing effect for these two strategies. Within Pan and Rickard's (2018) framework, students who first mentally retrieve their answers should have better retention of their correct answers then students who copy their answers from an authoritative source. This should even be the case for students who generated wrong answers to the homework questions because they were given immediate feedback including the correct answer and when this was done long-term retention for the correct answer was improved even when the self-generated answer was wrong (Hays et al., 2013; Yan et al., 2014). In contrast, students who copied the answers from a source found on the internet were enacting the restudy control conditions of testing effect experiments, which did not produce long-term retention (Pan & Rickard, 2018). They should not subsequently remember the answers regardless of whether the answers were right or wrong. The prediction that found and recorded

information is poorly retained has been confirmed in non-academic contexts. There was poor retention for information retrieved by Google (Sparrow et al., 2011), poor retention of pictures photographed at a museum (Henkel, 2014), and poor memory for routes and locations found through a navigation system (Burnett & Lee, 2005).

Consequently, students who mentally retrieve answers to homework questions should have better performance on the exam one to three weeks later than students who find the answers on the internet. However, the students looking up the answers should perform better on the online homework assignment precisely because they have looked the answers up in an authoritative source. Hence, if, in fact, better versus worse performance on online homework assignments reflects the effect of a strategy of look-up versus a strategy of mental retrieval then the testing effect predicts that those who perform better on homework assignments will perform more poorly on the subsequent exams.

Notice that the testing effect implies a possible pattern of quiz and exam performance that has never been observed in any study of academic performance using the distributed questioning methodology: better performance on the Pre-lesson questions than on the exam questions. This extreme result would only occur when two conditions were both satisfied. First, the look-up of answers to online homework questions both insured correct answers to those questions and reduced the long-term retention of those answers. Second, whatever additional study was done by the student was insufficient to compensate for the poor retention of the online homework questions and answers. Consequently, a student who looked up the correct answers to the online homework questions would often fail to remember the answers to the questions on the exam despite additional study. In fact, Glass et al. (2008) found that distributed questioning caused long-term retention only when an answer was generated on more than one occasion. Consequently, the failure to use the online homework quiz as an opportunity to generate answers would impair the effectiveness of subsequent study efforts when generation was possible.

Given that better performance on the Pre-lesson questions than on the exam questions is extremely deviant from the pattern of performance over successive tests that has always been observed, reversing the best and worst conditions, and would only occur if students with histories of academic success employed a strategy at the college level that precluded success, one might expect such a pattern of performance to be rare. Nevertheless, it was decided to examine an appropriate data base to determine how often such a pattern was found; that is, had the percent of students performing better on homework than exams increased in the past decade? An increase in the percent of students who performed better on homework assignments than on exams would indicate that technology was influencing how students were doing their homework, for the worse. Presumably, homework answers were becoming easier to retrieve online, which would improve homework performance but make the answers less likely to be remembered, which would reduce exam performance.

**Table 1.** Sizes of enrolments and question sets for courses included in study.

| Title | Year | Enrollment | Question sets |
|---|---|---|---|
| Memory | 2008 | 309 | 47 |
| | 2009 | 346 | 112 |
| | 2010 | 129 | 80 |
| | 2010 | 379 | 84 |
| | 2011 | 378 | 73 |
| Cognition | 2012 | 169 | 108 |
| | 2013 | 147 | 89 |
| | 2014 | 138 | 100 |
| | 2015 | 98 | 126 |
| | 2016 | 121 | 132 |
| | 2017 | 116 | 105 |
| | 2018 | 103 | 134 |
| Total | | 2433 | 232 |

All courses took place only in Fall except for 2010, when the course was taught in both Spring and Fall.

## Method

### *Sample*

The data base was student homework, quiz, and exam performance from eleven years of the first author's college two-section lecture course at a public university in the United States. From 2008 through 2018 the distributed questioning methodology was implemented in a two-section college lecture course at a flagship state university. From 2008 to 2011 this was a course on human learning and memory and from 2012 to 2018 the course was on human cognition. The enrolments in the courses and the number of question sets asked are shown in Table 1. Over the eleven-year period, each year a new collection of question sets was created for a course by replacing some question sets from the previous year with new ones to support changes in the curriculum and to test new hypotheses relating to malleable features of the distributed questioning methodology. Also, some question sets were eliminated because performance across all questions in the set was above 90%.The entire collection of 232 question sets from which the question sets used in the eleven courses comprising the study are available among the instructor's materials for Glass (2016), which is the textbook that was used for the three most recent courses included in the study.

The total student sample was 59% women and 41% men. The age distribution was 5% 17 – 18, 92% 19 – 24, and 3% 25 – 36. Neither gender nor age distributions differed significantly across the courses during the eleven years of the study.

### *Measures*

For each class, students were divided into two groups on the basis of whether percent correct for Pre-lesson questions was less than percent correct for the related Exam questions. Students who performed better on exams were presumed to be generating their own homework answers and will be called homework-generators. It was predicted that their performance on Post-lesson and Review questions would fall between that for Pre-lesson and Exam questions because both in the laboratory and in the classroom, distributed questioning has always been shown to incrementally improve

long-term, mean class performance. Students who performed better on Pre-lesson questions than Exam questions were presumed to be answering questions through look-up and will be called homework-copiers. It was predicted that they would perform better than the homework-generators on the online-homework Pre-lesson and Review questions, where they could look up the answers, but worse than the homework-generators on the classroom Exam questions, for which they were less likely to remember the answers.

To provide converging evidence for the Copy versus Generate hypothesis, at the conclusion of the 2017 course students were asked which of two strategies, the generate strategy or copy strategy, they had used predominantly during the semester.

Even if the difference between pre-lesson and exam performance sorted students into two groups that also performed differently on post-lesson and review questions, it would only demonstrate the effect of a predominate strategy, not a ubiquitous one. There is no reason to believe that a student either always or never used google to find a homework answer and there is no reason to believe that a google search found the correct answer every time. During the 2018 course, to obtain more specific information about the frequency with which they used the Copy versus the Generate strategy, students were asked after eight homework assignments which strategy they had used.

At the conclusion of the 2017 course, immediately after finishing the final exam, students answered the following question:

When you do an online homework assignment, as each question appears, which of the alternatives below best describes how you usually answer the question?

A. I either answer the question from memory or guess at the answer.
B. I look the answer up in some way including searching the textbook and course materials, using google, and checking with someone else who has already answered the question.

Alternative A was intended to be a statement of the generate strategy and statement B was intended to be a statement of the copy strategy.

During the 2018 course, immediately after finishing each of eight of the online homework assignments, students answered the following online question:

When you just did this online homework assignment, which of the alternatives below best describes how you answered each question?

A. I always either answered the question from memory or guessed at the answer.
B. I looked the answer up in some way including searching the textbook and course materials, using google, and checking with someone else who has already answered the question.
C. Sometimes I looked the answer up and sometimes I did not look it up.

### *Procedure*

All of the courses were taught by the same instructor using a rigidly controlled procedure in order to be able to measure the effect of distributed questioning within a course-embedded experimental design. Hence, there was excellent control over what occurred in the classroom.

Lectures were presented as a sequence of Power Point slides that contained all the content spoken by the lecturer as well as all of the questions that were asked. Students were permitted only a single, timed, attempt on each online homework assignment. So there was complete information on when the online homework assignments were completed and submitted.The course met every Tuesday and Thursday. Forty-five hours before every class, an online homework quiz was posted that could be answered up until the start of the class. When a quiz was submitted for scoring, there was immediate feedback that included the quiz, the student's answers, the correct answers, a brief explanation of the correct answer, and the page in the textbook where a complete explanation could be found. This feedback was available to the student for the rest of the semester. So even though a student had only a 45 hour window to complete a quiz of $4-8$ questions, the student had the entire rest of the semester to study the quiz, their own answer, and the correct answer.

The course consisted of 28 classes distributed over 15 weeks because of Thanksgiving break in the fall and spring break in the spring. Three of the class sessions were devoted to non-cumulative unit exams. So the course was divided into three 4 to 4 and 1/2 week units terminating in an exam.

## Results

Two kinds of data were analysed. The quiz and exam results for the 11 years covered by the study were the primary data. Then, in the last two years of the study the students were asked how what actions they took to complete their homework assignments and these reports were related to exam performance.

### *Does answering homework questions influence subsequent performance? Quiz and exam results*

Percent correct on the Pre-lesson, Post-lesson, Review, and Exam questions is shown separately for homework-copiers and homework-generators in Table 2 for each class from 2008 to 2017. Notice that Review questions were not asked in 2012 (to assess whether their absence would affect subsequent exam performance). So the means for the entire 11-year period for all columns at the bottom of the table do not include the results for 2012. A repeated measures analysis of variance was performed for each course in which the factors were Question type (Pre-lesson, Post-lesson, Review, Exam) and Student type (Copier, Generator). The rightmost column of Table 2 shows the F-value for the critical Question x Student interaction testing the prediction that homework-copiers would perform better on online Pre-lesson and Review Questions and worse on Exam questions. The interaction was always significant at the $p < .001$ level.

**Table 2.** Mean percent correct (95% confidence intervals) for successive questions in a college lecture course.

| Year | Student Type | Pre-lesson M (C.I.) | Post-lesson M (C.I.) | Review M (C.I.) | Exam M (C.I.) | Question x Student Interaction |
|---|---|---|---|---|---|---|
| 2008 | Copier | 78 (72, 83) | 69 (63, 75) | 76 (70, 82) | 68 (64, 71) | $F(3, 921) = 65$ |
| | Generator | 47 (45, 49) | 77 (75, 80) | 71 (69, 74) | 78 (76, 79) | |
| 2009 | Copier | 81 (75, 87) | 69 (64, 74) | 82 (76, 88) | 73 (69, 76) | $F(3, 1032) = 77$ |
| | Generator | 50 (48, 52) | 70 (69, 72) | 61 (59, 63) | 84 (83, 85) | |
| 2010 | Copier | 88 (80,96) | 68 (62, 74) | 74 (66, 81) | 75 (70, 80) | F(3,381) = 37 |
| | Generator | 59 (56, 62) | 74 (72, 76) | 64 (61, 67) | 86 (84, 88) | |
| 2010 | Copier | 83 (80, 86) | 77 (74, 79) | 83 (80, 87) | 75 (73, 77) | $F(3, 1131) = 106$ |
| | Generator | 63 (61, 65) | 75 (74, 76) | 72 (71, 74) | 84 (83, 86) | |
| 2011 | Copier | 88 (84, 91) | 70 (68, 73) | 83 (80, 86) | 78 (76, 80) | $F(3, 1128) = 112$ |
| | Generator | 64 (62, 66) | 70 (69, 72) | 72 (70, 74) | 88 (86, 89) | |
| 2012 | Copier | 77 (73, 82) | 68 (64, 72) | | 70 (66, 74) | $F(2, 334) = 68$ |
| | Generator | 61 (59, 64) | 73 (71, 76) | | 80 (78, 83) | |
| 2013 | Copier | 83 (74, 91) | 79 (74, 85) | 73 (66, 80) | 74 (82, 85) | $F(3, 435) = 25$ |
| | Generator | 57 (54, 60) | 74 (72, 76) | 71 (69, 73) | 83 (69, 80) | |
| 2014 | Copier | 88 (84, 92) | 71 (68, 75) | 77 (72, 82) | 78 (75, 81) | $F(3, 408) = 52$ |
| | Generator | 59 (56, 62) | 68 (66, 71) | 66 (62, 69) | 83 (81, 85) | |
| 2015 | Copier | 86 (81, 91) | 71 (67, 76) | 83 (78, 87) | 77 (74, 80) | $F(3, 288) = 58$ |
| | Generator | 59 (54, 63) | 67 (61, 71) | 70 (66, 74) | 85 (83, 88) | |
| 2016 | Copier | 88 (85, 92) | 70 (67, 73) | 86 (83, 90) | 77 (75, 79) | $F(3, 357) = 53$ |
| | Generator | 68 (65, 71) | 75 (72, 78) | 79 (76, 82) | 85 (83, 88) | |
| 2017 | Copier | 89 (86, 93) | 94 (92, 96) | 88 (85, 92) | 76 (74, 79) | $F(3, 342) = 77$ |
| | Generator | 62 (58, 65) | 90 (88, 93) | 73 (69, 76) | 85 (82, 88) | |
| 2018 | Copier | 88 (84, 93) | 80 (74, 86) | 86 (83, 90) | 75 (72, 79) | $F(3, 303) = 21$ |
| | Generator | 66 (54, 78) | 76 (60, 91) | 80 (70, 90) | 88 (79, 88) | |
| Total | Copier | 81 (80, 83) | 75 (74, 76) | 80 (78, 81) | 78 (77, 78) | $F(3, 6786) = 244$ |
| | Generator | 58 (55, 60) | 73 (71, 76) | 70 (67, 72) | 84 (82, 86) | |

$p < .001$ for all interactions. Total does not include 2012.

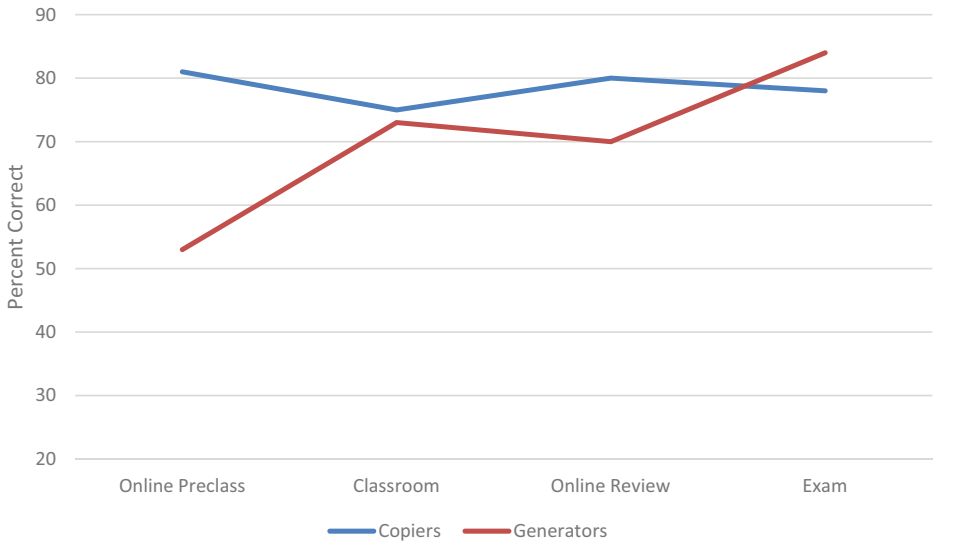


**Figure 1.** Interaction between copiers and generators and type of question on percent correct.

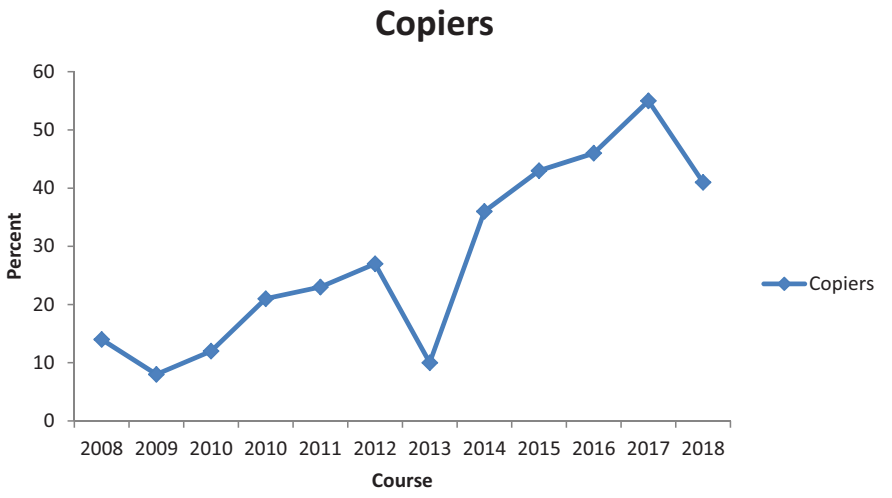


**Figure 2.** The percent of copiers from 2008 to 2017.

An analysis of variance was also performed for each course and on the cumulative results for all courses except in 2012 shown on the bottom two rows of Table 2. The factors were Question (Pre-lesson, Post-lesson, Review, Exam) and Student (Copier, Generator). Question, $F(3,6786) = 244$, Student, $F(1,2262) = 201$, and their interaction, $F(3,6786) = 462$, were all significant at the $p < .001$ level. In regards to the interaction, as shown in Figure 1, planned comparisons revealed that homework-copiers performed better than homework-generators on Pre-lesson Questions, $F(1,2262) = 795$, and Review Questions, $F(1,2262) = 154$, but homework-copiers performed worse than homework-generators on Exam Questions, $F(1,2262) = 144$, which were all significant at the $p < .001$ level.

### *Has the Percent of Students Benefiting from Answering Homework Questions Changed in the Past Decade?*

Figure 2 shows that with the exception of 2013 and 2018 the percent of homework-copiers has been increasing annually and is now around 50%.

### *Do homework strategies influence retention?*

The purpose of the study was to determine whether new information technologies that made the immediate retrieval of information online possible affected how students did their homework assignments. Having found that students could be divided into two groups on the basis of their online homework and subsequent exam performance, the next step was to ask students how they answered the homework questions to determine whether there was a correlation between reported strategy and performance. To provide a finer analysis of the data for the 2017 and 2018 classes that would make it possible to compute a correlation, the percent correct on Exams was subtracted from the percent correct on Pre-lesson questions to create a deviation score (D). Negative D scores (as the result of higher exam performance) were

presumed to indicate the positive effect of generation on exam performance and positive D scores were presumed to indicate the lack of effect of copying on exam performance.

In 2017, students were asked on the final exam to choose between alternatives A and B described in the Method section. Response A, describing generation, was assigned a score of 1 and response B, describing copying, was assigned a score of (–1) to create an Action score (A). Spearman's rho for A versus D, was $-0.208$, $p = .025$. The significant correlation was consistent with generating answers creating long-term retention of correct answers on a unit exam, hence $D < 0$, and copying answers from another source not being remembered on a unit exam, hence $D > 0$.

In 2018, to provide a more immediate measure of the homework strategy, at the end of each of eight online quizzes students were asked to choose among three alternatives mentioned in the Method. For each of the eight quizzes, the student received a score of 1 for generating, 0 for reporting both strategies, and $-1$ for reporting copying. The scores for the eight quizzes were averaged to compute an action score (A) ranging from 1 (always generated) to $-1$ (always copied). The Spearman's rho for the correlation between D and A was $-0.251$, $p = .013$.

## Discussion

The principal finding of this study is that students could be sorted into two groups on the basis of the relationship of their performance on homework questions to their performance on classroom quiz questions and exams. Questions were organised into sets containing a pre-lesson, post-lesson, review, and exam question whose answers could all be inferred from the same principle or fact statement in order to track the growth of knowledge of the course topic. The predominant pattern of student performance was a monotonic increase in the probability of answering each successive question in the sequence correctly, indicating that the online pre-lesson homework question, classroom post-lesson question, and online homework review question, all produced learning that ultimately increased the probability of answering the exam question correctly. However, a minority of the students performed much better on the homework assignments than the classroom assignments, indicating that reporting the correct answers for the homework assignments did not result in learning that prepared the students for the classroom exams. Furthermore, as shown in Figure 2, over the last eleven years there has been a steady increase in the number of students who exhibited this pattern of futility until it has recently approached nearly half the class.

As mentioned, there are radically different consequences from generating a response, such as an answer to a question, versus looking it up from another source (Bobrow & Bower, 1969; Pan & Rickard, 2018). Generated responses are remembered and may aid long-term learning even when they are wrong (Yan et al., 2014). Copied responses are not remembered so are not part of long-term learning. An obvious vector for increasing copying versus generating over the last eleven years is the growth of smart phone use, which provides google as an easy option for rapidly obtaining correct answers to homework questions. The presumptive use of cell phone look up would produce exactly the until-recently unusual, counter-intuitive, pattern observed:

high homework scores predicting lowered exam scores. Furthermore, consistent with this hypothesis, during the most recent two years students who obtained higher homework and lower exam scores were more likely to report higher frequencies of a copying strategy for homework. The cost of copying was from half to a full letter grade.

Homework-copiers do not perform worse than homework-generators on Post-lesson questions, which are short-term tests of material just presented in class. In 2017 a deliberate effort was made to increase performance on classroom Post-lesson questions by always presenting the question in class on the Power Point slide immediately after the slide containing the question. As can be seen from Table 2, the manipulation improved Post-lesson performance for both homework-copiers and homework-generators to levels above subsequent exam performance.

So homework-copiers can perform as well as homework-generators on tasks not requiring long-term retention and both groups are equally capable of inferring the answer to a question on a just presented slide. Unfortunately, homework-copiers apparently adopt a strategy that increases the homework quiz score at the cost of a lower exam score. There is no reason to believe that the students are aware that their homework strategy lowers their exam score. Rather, it is more likely that they make the commonsense inference that any study strategy that raises their homework quiz score raises their exam score as well.

On the other hand, the correlations between the self-reported strategies and actual patterns of performance, though significant, were not large. Other factors that may have influenced the decision to google the answer to a question, such as its apparent difficulty and whether the student was multi-tasking while doing the homework assignment or doing it under time pressure, were not investigated. Also, other factors besides the initial decision to look up the homework answer may have contributed to the ultimate decline in exam performance. Notice that the decision to copy the correct answers to online Pre-lesson and Review homework questions has a long-term effect on performance on exam questions that are answered one to three weeks later. Whatever studying homework-copiers are doing in addition to performing the online homework assignment is not sufficient to overcome the effect of the strategy for completing the homework assignment. One possible reason for this is that whatever other study strategies homework-copiers are using are similar to the strategy for completing the homework, hence also fail to induce long-term retention. A strategy of looking up the answer to any question whose answer the student was unsure of appears to be a plausible study strategy but absent a deliberate attempt to first generate the answer, it would not produce long-term retention of the answer. More research is needed to determine the causal factors producing the decline in exam performance.

An obvious possible causal factor for the increase in homework performance is the increasing ease of use and sophistication of social media. It is already difficult to remember that ten years ago college students would have had much less lifetime experience with cell phones and their cell phones would have only recently, if at all, included online search capabilities. Ten years ago, the least effortful strategy would have been to answer the question without looking up the answer, submit the quiz, and obtain the correct answers as feedback.

However, today's college students have a lifetime of experience using their cell phones as immediate sources of information for everything from the time and the weather to the answers to obscure trivia questions. They are skilful at rapidly obtaining whatever information they want from their phones and habitually do so. Given such a habit, it is plausible that for many the response to an online homework question is find the answer on their phone and then check that answer against the feedback provided when the quiz is submitted.

Unfortunately, there have been few experimental studies on the effects of cell phone technology, social media, and information retrieval technology, on long-term retention or on any aspect of academic performance (Tess, 2013). These results are consistent with the few other findings about the effects of the use of electronic devices for information retrieval and storage on the retention of that information by the user. There was poor retention for information retrieved by Google (Sparrow et al., 2011), poor retention of pictures photographed at a museum (Henkel, 2014), and poor memory for routes and locations found through a navigation system (Burnett & Lee, 2005). Students who took written notes had better exam performance than students who took notes on a laptop (Mueller & Oppenheimer, 2014). Students who read content on paper had better exam performance than student who read the same content on a laptop (Ackerman & Goldsmith, 2011; Mangen, Walgermo, & Bronnick, 2013). When Rosen et al. (2013) monitored students that were studying they found that they self-interrupted by switching to online social media and those who self-interrupted more had lower GPAs. Therefore, there may be a variety of consequences of doing online homework involving social media and information retrieval technology that collectively reduce the retention of the homework questions and answers.

The difference in the long-term retention between self-generated answers and answers retrieved from another source does not mean that a student must abstain from information retrieval technology when doing homework. What it means is that the student must first self-generate an answer before using information retrieval technology to check whether the self-generated answer is correct. In this case, the retrieved answer serves as self-generated feedback to the student response and feedback, especially delayed feedback, increases retention (Sinha & Glass, 2015).

Regardless of the explanation, the decline in the effectiveness of homework for improving exam scores should be of concern to any instructor whose purpose in assigning homework is to increase learning. It is surprising, if not shocking and dispiriting to see that after four weeks of diligent study, as measured by the number of completed assignments, some students are doing worse on Exam questions than they initially did on related Pre-lesson questions whose answers were implied by the same principles and fact statements. It is even more dispiriting to see the percent of students exhibiting this pattern increase annually.

These results strongly suggest that an increasing percent of students are adopting a strategy for completing homework quizzes that reduces the effect of quiz performance on exam performance. However, the results of only one course per year do not document the pervasiveness of the effect across course and disciplines nor do the results provide comparative evidence of how differences in instructional methods may contribute to or inhibit ineffective homework strategies. The phenomenon of high

homework and lower exam performance deserves immediate attention by researchers to determine its pervasiveness in academia, to identify its cause, and to identify a strategy for its amelioration. If not addressed immediately it may soon be so pervasive that its effect will be impossible to assess. At that point, homework may have become an ineffective ritual and cell phone use will be an existential threat to academic study.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

Arnold L. Glass http://orcid.org/0000-0002-3202-5953

## References

Ackerman, R., & Goldsmith, M. (2011). Metacognitive regulation of text learning: On screen versus on paper. *Journal of Experimental Psychology. Applied*, *17*(1), 18–32. https://doi.org/10.1037/a0022086

Bobrow, S. A., & Bower, G. H. (1969). Comprehension and recall of sentences. *Journal of Experimental Psychology*, *80*(3, Pt.1), 455–461. https://doi.org/10.1037/h0027461

Bower, G. H., & Clark, M. C. (1969). Narrative stories as mediators for serial learning. *Psychonomic Science*, *14*(4), 181–182. https://doi.org/10.3758/BF03332778

Burnett, G., & Lee, K. (2005). The effect of vehicle navigation systems on the formation of cognitive maps. *International Journal of Psychology*, *40*(1), 27–35. https://doi.org/10.1016/b978-008044379-9/50188-6

Carter, S. P., Greenberg, K., & Walker, M. (2017). The impact of computer usage on academic performance: Evidence from a randomized trial at the United States Military Academy. *Economics of Education Review*, *56*(1), 118–132. https://doi.org/10.1016/j.econedurev.2016.12.005

Cook, R., & Bird, G. (2013). Do mirror neurons really mirror and do they really code for action goals? *Cortex*, *49*(10), 2944–2945. https://doi.org/10.1016/j.cortex.2013.05.006

Glass, A. L. (2009). The effect of distributed questioning with varied examples on exam performance on inference questions. *Educational Psychology*, *29*(7), 831–848. https://doi.org/10.1080/01443410903310674

Glass, A. L. (2016). *Cognition: A Neuroscience Approach*. Cambridge University Press.

Glass, A. L. (2019). Within the framework of the dual-system model, voluntary action is central to cognition. *Attention, Perception & Psychophysics*, *81*(7), 2192–2216. https://doi.org/10.3758/s13414-019-01737-0

Glass, A. L., & Kang, M. (2019). Dividing attention in the classroom reduces exam performance. *Educational Psychology*, *39*(3), 395–408. https://doi.org/10.1080/01443410.2018.1489046

Glass, A. L., Brill, G., & Ingate, M. (2008). Combined online and in- class pre-testing improves exam performance in general psychology. *Educational Psychology*, *28*(5), 483–503. https://doi.org/10.1080/01443410701777280

Glass, A. L., & Sinha, N. (2013). Multiple-choice questioning is an efficient instructional methodology that may be widely implemented in academic courses to improve exam performance. *Current Directions in Psychological Science*, *22*(6), 471–477. https://doi.org/10.1177/0963721413495870

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

Henkel, L. A. (2014). Point-and-shoot memories: The influence of taking photos on memory for a museum tour. *Psychological Science*, *25*(2), 396–402. https://doi.org/10.1177/0956797613504438

Mangen, A., Walgermo, B. R., & Bronnick, K. (2013). Reading linear texts on paper versus computer screen: Effects on reading comprehension. *International Journal of Educational Research*, *58*(1), 61–68. https://doi.org/10.1016/j.ijer.2012.12.002

McCurdy, M. P., Leach, R. C., & Leshikar, E. D. (2017). The generation effect revisited: Fewer generation constraints enhances item and context memory. *Journal of Memory and Language*, *92*(2), 202–216. https://doi.org/10.1016/j.jml.2016.06.007

McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger, H. L. III (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, *103*(2), 399–414. https://doi.org/10.1037/a0021782

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. III (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology*, *27*(3), 360–372. https://doi.org/10.1002/acp.2914

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

Mueller, P. A., & Oppenheimer, D. M. (2014). The pen is mightier than the keyboard: The advantages of longhand over laptop note taking. *Psychological Science*, *25*(6), 1159–1166. https://doi.org/10.1177/0956797614524581

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin*, *144*(7), 710–756. https://doi.org/10.1037/bul0000151

Roediger, H. L., III, Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology. Applied*, *17*(4), 382–395. https://doi.org/10.1037/a0026252

Rosen, L. D., Carrier, M. L., & Cheever, N. A. (2013). Facebook and texting made me do it: Media-induced task-switching while studying. *Computers in Human Behavior*, *29*(3), 948–958. doi.: 10.1016/j.chb.2012.12.001 https://doi.org/10.1016/j.chb.2012.12.001

Sana, F., Weston, T., & Cepeda, N. (2013). Laptop multitasking hinders classroom learning for both users and nearby peers. *Computers & Education*, *62*(1), 24–31. https://doi.org/10.1016/j.compedu.2012.10.003

Sinha, N., & Glass, A. L. (2015). Delayed, but not immediate, feedback after multiple-choice questions increases performance on a subsequent short-answer, but not multiple-choice, exam: Evidence for the dual-process theory of memory. *The Journal of General Psychology*, *142*(2), 118–134. https://doi.org/10.1080/00221309.2015.1024600

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, *333*(6043), 776–778. https://doi.org/10.1126/science.1207745

Tess, P. A. (2013). The role of social media in higher education classes (real and virtual)—A literature review. *Computers in Human Behavior*, *29*(5), A60–A68. https://doi.org/10.1016/j.chb.2012.12.032

Visalberghi, E., Fragaszy, D. M., & Savage-Rumbaugh, S. (1995). Performance in a tool-using task by common chimpanzees (Pan troglodytes), bonobos (Pan paniscus), an orangutan (Pongo pygmaeus), and capuchin monkeys (Cebus apella). *Journal of Comparative Psychology*, *109*(1), 52–60. https://doi.org/10.1037/0735-7036.109.1.52

Yan, V. X., Yu, Y., Garcia, M. A., & Bjork, R. A. (2014). Why does guessing incorrectly enhance, rather than impair, retention? *Memory & Cognition*, *42*(8), 1373–1383. https://doi.org/10.3758/s13421-014-0454-6