

Fall 9-30-2019

## Hacking the Non-Technical Brain: Maximizing Retention in a Core Introductory IT Course

Eric Sturzinger

*United States Military Academy*, [eric.sturzinger@westpoint.edu](mailto:eric.sturzinger@westpoint.edu)

Daniel Hawthorne

*United States Military Academy*

Thomas Babbitt

*United States Military Academy*

Follow this and additional works at: [https://digitalcommons.usmalibrary.org/usma\\_research\\_papers](https://digitalcommons.usmalibrary.org/usma_research_papers)



Part of the [Engineering Education Commons](#)

---

### Recommended Citation

Eric M. Sturzinger, Daniel S. Hawthorne, and Thomas A. Babbitt. "Hacking the Non-Technical Brain: Maximizing Retention in a Core Introductory IT Course." Proceedings of the 20th Annual SIG Conference on Information Technology Education. ACM, 2019.

This Conference Proceeding is brought to you for free and open access by USMA Digital Commons. It has been accepted for inclusion in West Point Research Papers by an authorized administrator of USMA Digital Commons. For more information, please contact [nicholas.olijnyk@westpoint.edu](mailto:nicholas.olijnyk@westpoint.edu).

# Hacking the Non-Technical Brain: Maximizing Retention in a Core Introductory IT Course

A Quantitative Study on the Performance Impacts of High-Frequency, No-Risk Quizzing

Eric M. Sturzinger  
United States Military Academy  
West Point, New York  
eric.sturzinger@westpoint.edu

Daniel S. Hawthorne  
United States Military Academy  
West Point, New York  
daniel.hawthorne@westpoint.edu

Thomas A. Babbitt  
United States Military Academy  
West Point, New York  
thomas.babbitt@westpoint.edu

## ABSTRACT

Maximizing student retention of, and ability to apply, technical material in introductory information technology courses is a complex task, especially with respect to the general student population. This population struggles with the application of programming concepts in the time-constrained testing environment. Our study considers the implementation of daily quizzes in a core-curriculum information technology and programming course as a means to improve student concept retention and application. Between the first and second exams, the instructors implemented a series of high-frequency, no-risk quizzes. Of the four sections of the course that each instructor taught, two sections each were provided with the quizzes as the experimental group and two remained with the standard curriculum as the control. The results demonstrate the benefits of frequent, effortful recall on student performance in a core-curriculum information technology and programming course.

## CCS CONCEPTS

• **Social and professional topics** → **Information technology education**; **Student assessment**; *Computer science education*; • **Applied computing** → *Collaborative learning*;

## KEYWORDS

high-frequency; no-risk; effortful recall; fluid intelligence; information technology core education

## ACM Reference Format:

Eric M. Sturzinger, Daniel S. Hawthorne, and Thomas A. Babbitt. 2019. Hacking the Non-Technical Brain: Maximizing Retention in a Core Introductory IT Course: A Quantitative Study on the Performance Impacts of High-Frequency, No-Risk Quizzing. In *The 20th Annual Conference on Information Technology Education (SIGITE '19)*, October 3–5, 2019, Tacoma, WA, USA. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3349266.3351411>

## 1 INTRODUCTION

Recent literature demonstrates the necessity for the introduction of information technology (IT) curriculum earlier in the formative educational years of a student's development. Our institution mandates all graduates pass an introduction to computing and information technology course, defined as a Computer Science 0 (CS0) course

in [1]. In addition, students not enrolled in IT-related programs are required to take a Data Science 0 (DS0) course; the previous version is described in [11]. These requirements ensure all graduates understand basic IT and DS theory and applications. The CS0 course is taken in the first year and the DS0 course in the third year of a four year program. The DS0 course focuses on data science core principles while the CS0 course covers a more traditional set of IT topics, such as: programming, networking, cybersecurity, and cyber ethics [19].

In 2018, the Institute for Electrical and Electronics Engineers (IEEE) Spectrum designated Python the top programming language [5]. Our CS0 course utilized Jython Environment for Students (JES), a simplified version of Python, as its programming language [10]. Many of our students are exposed to programming for the first time in CS0 and have difficulty designing, implementing, and testing a programming solution on midterm and final exams. This necessitates an additional concept reinforcement technique to increase student performance on significant graded events and maximize long-term retention and skill.

This paper proposes and assesses the impact of high-frequency, no-risk quizzes on graded event performance in our CS0 course. The course consists of 40 lessons, approximately half of which are focused on introductory programming concepts and their application. As the course progressed through the semester, these concepts increased in complexity. In previous semesters of instruction, we observed a significant decrease in student performance from the first programming exam to the second programming exam. While this is not unusual, we noticed students struggled to apply concepts learned in the first block when combined with more advanced programming topics from the second block. The decrease in overall exam score prompted us to explore means to improve retention and integration of programming topics from the first block into subsequent concepts taught later in the course. The goal was to increase student retention and ultimate performance on the second and third exams.

Because of the nature of our CS0 course, we were required to stay within the bounds of the schedule and graded event template. We therefore assigned a short, ungraded quiz at the beginning of each lesson. Following each quiz, students were required to justify and explain their answers to a partner. This allowed each student to receive immediate, peer-led feedback. We conducted multiple statistical tests to assess whether our approach reduced the historical drop in performance between the first and second exams. We provided these additional ungraded quizzes to half of the observed sections to establish a control and an experimental

This paper is authored by an employee(s) of the United States Government and is in the public domain. Non-exclusive copying or redistribution is allowed, provided that the article citation is given and the authors and agency are clearly identified as its source.

SIGITE '19, October 3–5, 2019, Tacoma, WA, USA

2019. ACM ISBN 978-1-4503-6921-3/19/10.

<https://doi.org/10.1145/3349266.3351411>

subset of students. While the results do not provide a clear effect of the quizzes across the entire experimental group, there is a clear effect for certain subsets of students. The number of students who performed much worse (defined later in Table 3) on the second exam relative to the first exam decreased by just over 50% for those belonging to the experimental group. On the other hand, students who performed much better increased by 100%. We can therefore claim that we found “a” method to improve the fluid intelligence of students in a core, introductory information technology course.

The remainder of this paper is organized as follows: Section 2 discusses key definitions and related work, Section 3 describes our approach, Section 4 and Section 5 documents how we evaluated and assessed our results, and Section 6 concludes the study and provides a direction for future work.

## 2 DEFINITIONS AND RELATED WORK

Much of the research and experiments related to retrieval-based learning and effortful recall is naturally in the context of Psychology courses. While we found a few loosely related experiments in technical subjects, the vast majority came from educators in Psychology related topics. They are still relevant to our work as the mechanics and objectives are largely identical.

### 2.1 Definitions

*Effortful recall* is used interchangeably with retrieval-based learning in much of the literature. The popular pedagogical book *Make It Stick* notes effortful recall requires “that you ‘reload’ or reconstruct the components of the skill or material anew from long-term memory rather than mindlessly repeating them from short-term memory [4]. Chapter two of the same text defines examples when “no-risk” quizzes were used. The primary benefit was that student stress was slightly reduced, potentially allowing for more mental energy to be devoted to recall. Effortful recall is the foundation on which most of the related work in the next section is based.

*Fluid intelligence* is one of two types of general intelligence first defined by Raymond Cattell in the 1970s, the other being crystallized intelligence [6]. Fluid intelligence is the ability to use logical thinking to solve new and complex problems, regardless of domain or previously acquired knowledge. Conversely, crystallized intelligence is developed from experience in specific tasks leading to domain knowledge. We sought to avoid relying on the crystallized intelligence of students by interleaving question topics from different lessons so as to not promote short-term memorization of each lesson’s concepts.

*Basic* concepts include the most fundamental structures and skills required to perform virtually any programming task, such as: storing a value or a reference to an object in a variable, using print statements, iteration (single for loop), selection (single if statement), and defining a function, among others.

*Advanced* concepts include a more complex version of a basic concept or an entirely new concept. Examples are: nested for loops, nested if statements, while loops, and file input/output. The delineation between the two groups aided us in quantifying how many different concepts were covered in each quiz question.

### 2.2 Related Work

Many experimental studies related to retrieval-based learning have been conducted on undergraduate courses. Jeffrey D. Karpicke and Henry L. Roediger III are two of the most prolific researchers in this field and have both contributed a large body of work. The pair described historical pedagogical views on testing along with a comprehensive review on contemporary laboratory experiments and classroom-focused studies related to retrieval-based learning [23]. They found that free-recall tests as the primary review method promoted increased long-term retention while rereading material was more ideal for short-term retention [14, 22]. In another pair of experiments, Karpicke and Roediger showed that interleaving restudying with effortful recall-focused tests produced the greatest performance results over testing only [15]. Additionally, repeated recall produced the maximum retention results compared to studying previous content alone or dropping previously tested content from future tests.

Karpicke compared long-term retention rates between three different learning methods: studying alone, recall from a single test, and repeated recall from multiples tests [12]. Repeated retrieval produced the largest proportion of recalled ideas. Roediger noted the importance of two of the main components of our study: 1) mixed retrieval practice (covering a wide spectrum of subject matter) and 2) using low-stakes quizzes (worth very little or zero points) [21]. In a Psychology course experiment, Blunt and Karpicke found that both paragraph and content-mapping recall produced roughly equivalent performance increases [3]. This refuted the common notion that the act of writing is the primary reason for successful retention, rather it is the act of retrieval in the various ways it can occur.

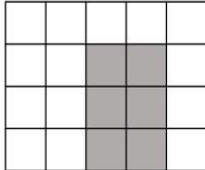
Nevid et al. examined the effect of effortful recall for both open and closed-book quizzes [17]. They demonstrated that open-book quizzes resulted in short-term retention but failed in long-term retention, the solution to which is repeated effortful recall through intelligently-spaced quizzes. Glenn presented several anecdotal stories related to effortful recall [9]. His main concern was that quizzes would steer students toward studying narrow, specific topics. However, he explained how it was a benefit as students had to recall other concepts related to the quizzed content, which strengthened the understanding of the targeted concept itself. Pyc and Rawson found that a longer interval between episodes of effortful recall increased exam performance [20]. McDaniel et al. examined three different quiz scenarios: quiz questions were identical to exam questions, similar to exam questions, and the use of multiple choice versus short answer questions [16]. Naturally, providing identical quiz questions ensured the best results while providing quiz questions similar in subject to final exam questions produced “nominally better” exam performance than rereading material.

The following two papers provided a more theoretical, psychological perspective of retrieval-based learning. Karpicke and Grimaldi offered a thorough description of both direct and indirect learning benefits of retrieval-based learning, the latter of which assists in identifying gaps or weaknesses in knowledge or understanding, focusing students toward their weak areas [13]. Rowland also provides a theoretical analysis of the mental mechanics in how retrieval-based learning is an effective learning method in [24].

### Lesson 14 Quiz

```

1. width = getWidth(picture)
   height = getHeight(picture)
   for pixel in getPixels(picture):
       x = getX(pixel)
       y = getY(pixel)
       if 2*width/5 <= x < 4*width/5 :
           if height/4 <= y < height :
               setRed(pixel, getRed(pixel)/2)
    
```



Fill in the missing lines above such that the red component is decreased by 50% in the shaded region of the picture on the right.

```

2. x = 7
   y = 10
   if (x < y):
       print "a"
       if (x + 2 < y):
           print "b"
       if (x < y/2):
           print "c"
       if (x < 2 * y):
           print "d"
       print "e"
       print "f"
       print "g"
    
```

Which letters are printed? (write all on line below)  
  a b f g  

3. Write two nested if statements which evaluate whether the distance between the color *orange* and the color of the pixel object *pixel* is less than 165 and the distance between the color *red* and the color consisting of red = 100, green = 35, and blue = 180 is greater than 200.

```

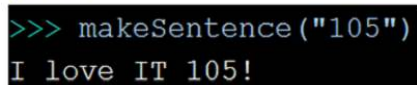
if distance(orange, getColor(pixel)) < 165:
    if distance(red, makeColor(100, 35, 180)) > 200:
    
```

(a) Nested if statements quiz

### Lesson 20 Quiz

```

1. def makeSentence(lastWord):
    word = "I "
    word += "love "
    word += "IT "
    word += lastWord + "!"
    print word
    
```



Fill in the missing lines in the function definition above so that the proper text is printed with the respective function call.

2. An HTML file contains the content & structure while a CSS file defines the style of the webpage.  
 (presentation, format)

3. An HTML file can point to multiple CSS files. True / False (Circle one)  
 Multiple HTML files can point to a single CSS file. True / False (Circle one)

```

4. superHeroes = ["Iron Man", "Black Panther", "Wonder Woman"]
   heroFile = open(getMediaPath("heroText.txt"), "wt")
   for hero in superHeroes:
       heroFile.write( hero )
   heroFile.close()
    
```

Fill in the lines above so that each super hero is written (each on a new line) to the document "heroText.txt" – Assume setMediaPath() has already been executed in the command area.

(b) HTML part 2 w/ file I/O quiz

Figure 1: Two of the seven total quizzes. Both contain a mix of syntactical and conceptual questions.

Although most studies on effortful recall primarily use Psychology and related courses, we found several related to introductory CS and other STEM-related courses.

Dobson et al. studied effortful retrieval in the context of a physical anatomy course and confirmed that distributed, retrieval-based strategies outperformed massed, restudy-focused strategies [8]. Norris employed a pre- and post-quizzing scheme where a pre-quiz was designed to prepare students for class discussion while post-quizzes were given after class discussion [18]. She observed an 8% increase in pass rate in her CS1 course and a 10% increase in her CS2 course. Another study crowd-sourced retrieval-based learning tests where more than 100 teachers compared a blocked practice approach (all questions on the same topic) to an interleaved approach (questions on different topics) [25]. The authors found that the interleaved approach produced a range of 5-30% better performance than the blocked approach. Cicirello instituted a pop quiz experiment in a CS1 course over 3 years and 9 semesters where some sections were randomly given pop quizzes and others were not [7]. Juniors were the best performing class (of those in the quizzed sections), scoring higher on exams as well as programming assignments. The primary benefit to students in this scenario was the more frequently received indirect feedback on their understanding of course material.

Zingaro et al. used the peer instruction technique to encourage students to explain and defend their individual answers to

peers, reach a consensus, and produce a group-wide answer for broader discussion (in a remedial CS1 course) [26]. The class used the "C" programming language; the authors found completing reading quizzes correlated positively with course performance. Bangert-Drowns et al. performed a large experiment through a mix of high school and college STEM courses where 83% of classes showed a positive effect of frequent testing on final exam performance [2]. The number of quizzes correlated with grades and students gave higher feedback marks in the classes when queried.

Although most quiz experiments were conducted within Psychology and related courses due to its nature, we were unable to find any effortful recall-focused quiz experiments for the general student population in an introductory IT course. This lack of literature is likely due to the rather unique curriculum requirements imposed upon students at this institution.

### 3 METHODOLOGY

As mentioned in Section 1, the overall goal was to minimize the degradation in student performance from the first to the second midterm exam by maximizing fluid intelligence, the ability to apply general logic and common sense to solve problems. We achieved this goal through effortful recall utilizing daily no-risk quizzes. We sought to accomplish two general objectives within the bounds of the course structure:

- (1) Strengthen and reinforce student understanding and application of multiple “basic” programming concepts learned prior to the first midterm exam.
- (2) Repetitively integrate recently learned “advanced” programming topics to strengthen long-term recall.

We designed quiz questions such that students had to perform effortful recall. Students did not receive prompts or hints such as those given in multiple choice questions where they simply must choose the correct answer. The following subsections document our rationale behind decisions pertaining to question content and type. The strategy in selecting question content was the primary driver in improving fluid intelligence while choosing the appropriate question type had the most impact on how we ensured effortful recall by the students.

### 3.1 Question Content

We sought to avoid only quizzing students on the exact concept taught in the previous lesson to minimize the exclusive reinforcement of crystallized intelligence. We instead integrated basic and advanced concepts learned in multiple prior lessons. Thus, quizzes not only covered material from the previous lesson, but also included concepts from the cumulative body of content up to that point in the course.

Table 1 shows lesson numbers and the primary subject covered on the quiz. For example, the Lesson 14 quiz, shown in Figure 1a, primarily covered nested if statements - taught in Lesson 13 - along with potentially all previous material thus far presented in the course. The subject areas listed are a combination of common introductory programming tasks, language-specific themes of the course (copying pictures into a canvas), and Hypertext Markup Language (HTML) - the language by which webpage content is presented.

**Table 1: Quiz Subject by Lesson**

Lesson	Subject
14	nested if statements
15	if statements w/ elif, else
16	nested for loops
17	copying pictures into a canvas
18	while loops
19	HTML Part 1
20	HTML part 2, file I/O

Each quiz consisted of three or four questions depending on difficulty. Students were given five minutes to complete the entire quiz. Every question integrated either a basic concept, advanced concept, or a combination of both into a single question. For example, question one in Figure 1a tests a student’s ability to recognize variable types for  $x$ ,  $y$ , *width*, and *height* and what they represent (a basic concept). In addition, it demonstrates a student’s ability to apply variables properly in an advanced concept like nested if statements.

We decided not to quiz students within the JES Integrated Development Environment (IDE), where the students performed their daily homework assignments. The purpose of this decision was to

introduce variance in recall methods to avoid similarity between homework and quiz questions. Finally, once complete, students exchanged quizzes and briefly discussed correct answers with a partner. The instructors led students in peer feedback, where confident students explained their answers to the rest of the section. The instructor usually only discussed a question in depth if there was a challenging question that much of the class answered incorrectly.

### 3.2 Question Type

The two primary question types we used requiring effortful recall were: short answer (conceptual) and code completion (syntax). Learning both programming concepts and respective language syntax in parallel is critical to mastering the fundamental outcomes of our CS0 course. Thus, we designed quiz questions that would test student retention of the cumulative body of conceptual topics as well as their application of syntax.

Short answer questions require the student to truly understand how a concept is applied in an example scenario or the nature of its relationship to other concepts to solve a problem. Question two in Figure 1b is an example of a conceptual question based on content from a previous lesson. As previously stated, these questions come from any prior lesson for the purpose of enhancing a student’s ability to apply learned concepts on demand.

Code completion concepts consist of completing a single line or multiple lines of code with proper syntax. These questions contain various scenarios that require the student to initially decipher learned concepts and understand the objective of the code snippet. The student can then judge what the most appropriate statement or expression is needed to accomplish the objective of the code snippet. Question one in Figure 1a is an example of a code completion problem where the student must provide the missing code to make the code block functional.

Both question types ensured that students were not able to rely on prompts to refresh their memory. Instead, given the general context of the question, they had to recall specific conceptual and syntactical information from the cumulative body of course content.

## 4 EXPERIMENT

The two instructors employed an experimental design to assess the effects of the high-frequency, no-impact quizzes on student performance. The previous observation, a decline in scores between the first and second exams, made the second programming block the most appropriate for the experiment. The experimental design includes the selection of groups and minimization of non-experimental factors in addition to the selection of the appropriate statistical tests to identify and measure any present effects.

To systematically assess the impact of the experiment, we chose the following research question: *What effect does group membership, experimental versus control, have on the difference in student performance on the first exam versus the second exam.* This research question led to the following null hypothesis:  $H_0$  *Group membership has no effect on student performance.* The alternate hypothesis is given based on the purpose of the research, focusing on improving performance:  $H_A$  *Experimental group membership has a statistically significant, positive impact on student performance.*

### 4.1 Design

The instructors, each teaching four sections of 16 to 18 students, first identified instructor variance as a potential non-experimental factor. To minimize the potential variance, the instructors each selected two sections as experimental (EXP) and two sections as control (CTL) as seen in Table 2. To mitigate the possibility of class days and sectioning differences, the instructors also varied which sections they chose as control versus experimental.

**Table 2: Instructor Sections**

Section	A	B	C	D
Instructor 1	EXP	EXP	CTL	CTL
Instructor 2	CTL	CTL	EXP	EXP

The second identified factor was GPA. To limit experimental complexity, the approach mitigated the effects of student GPA rather than including it as a factor. We recognized the inexorable fact that students perform differently as a potential source of noise in our results. To limit the influence of individual performance variance, the study focused on the difference in performance between the first and second exams rather than the cardinal scores.

Lastly, the instructors had access to the population results. This access allowed a comparison between the population results and both categories of the sample results to ensure the sample did not have performance anomalies that would render the results inconclusive or invalid.

### 4.2 Tests

The selection of tests focused on answering the research question, making determinations on the hypotheses, and implementing the aforementioned mitigations. We selected a one-tailed, two-sample t-test for the first test to address both hypotheses. Although the t-test is highly applicable, we envisioned a situation where highly variant individual performance of a few members had disproportionate effects on the means, possibly rendering the results invalid. The second test, logistic regression, provided the means to categorically assign performance, reducing the impact of those potential anomalies.

The chosen categories, depicted in Table 3, relate to the relatively normal distribution of the exam results. The results are in terms of the difference (DIF) in performance between the first and second exams. The neutral category (=) contains the results within one-half of a standard deviation (SD) of the average (AVG). The slight improvement category (+) contains those results with improvements between one-half standard deviation and one standard deviation, whereas much better (+ +) contains all improvements one standard deviation above the average. The worse (-) and much worse (- -) categories are the inverse.

## 5 RESULTS

The eight sections, four control and four experiment, represent a significant portion of the 30-section population. The sample contained 142 students of the 521 enrolled in the course. Since the instructors had access to the population results, the comparison of total population and sample population occurred first. The raw

**Table 3: Logistic Regression Categories**

Category	Description
++	$DIF \geq AVG + SD$
+	$AVG + SD > DIF > AVG + SD/2$
=	$AVG + SD/2 \geq DIF \geq AVG - SD/2$
-	$AVG - SD/2 > DIF > AVG - SD$
--	$AVG - SD \geq DIF$

data is presented next, followed by the results of, and observations garnered from, the planned statistical tests.

### 5.1 General

The aggregate sample results, including both experiment and control groups for both instructors, closely resembled the population results. Although the aggregate of the sample groups had slightly better performance, the averages on both exams, the average differences between exam performance, and the standard deviations were all within one percent of those of the population. The previously observed decline in performance, although only a couple percent, was present in both the population and sample. A final observation from the comparison of the sample and population performances was that both were generally normal distributions.

**Table 4: Control Results**

	Exam 1	Exam 2	DIF
AVG	86.34%	83.10%	-3.24%
SD	11.73%	14.51%	10.42%

The control and experimental results, depicted in Tables 4 and 5, elude to the effects of each group. Specifically, the reduced average (AVG) difference (DIF) suggests that membership in the experimental group had a positive impact on performance.

**Table 5: Experiment Results**

	Exam 1	Exam 2	DIF
AVG	84.06%	83.88%	-0.18%
SD	13.06%	11.87%	10.76%

### 5.2 T-Test

The one-tailed, two-sample t-test confirmed the significance of the performance improvement between the control and experimental groups seen in Tables 4 and 5. It resulted in a p-value of 0.04356, below the common threshold for significance of 0.05. This result led to the rejection of the null hypothesis; the difference is statistically significant. Furthermore, the difference being an improvement led to the acceptance of the alternate hypothesis.

### 5.3 Logistic Regression

Although logistic regression did not yield statistically significant results, Table 6 led to two interesting observations regarding the likelihood of group membership. First, it is twice as likely that a

randomly selected result found in the much better category (+ +) belongs to the experimental group. Second, it is about half as likely that a randomly selected result found in the much worse category (– –) belongs to the experimental group.

**Table 6: Logistic Regression Results**

	--	-	=	+	++
CTL	13	5	36	10	6
EXP	7	8	36	9	12
Total	20	13	72	19	18

Given a larger sample, the logistic regression results may have indicated statistical significance for the much better and much worse categories, however, the middle categories of better, about equal, and worse were similar to the degree that it is unlikely they would provide any significant difference. Regardless, a decrease in the number of students that perform much worse on the second exam and an increase in those that perform much better is a positive result. To provide more concrete results in the future, similar work should include as many sections as possible of a similarly-sized course.

## 6 CONCLUSION

This paper demonstrated the application of the established pedagogical practices of varied, frequent, and no-risk quizzes as a means to elicit effortful recall, a catalyst to improved learning and retention in the information technology core curriculum environment. This also fostered the development of the students’ fluid intelligence, their ability to identify patterns, and their use of fundamental logic to solve complex and novel problems.

Our results aligned with the best practices and advice found throughout the literature. The varied, effortful recall presented by the no-risk quizzes had a positive, statistically significant impact on student performance in the introductory IT course. Additionally, we showed that there was an approximately 50% reduction in likelihood that a student in the experimental group performed much worse on the second exam and a 100% increase in likelihood that a student performed much better.

As part of the planned transition from 40 lessons of 55-minutes to 30 lessons of 75-minutes, the CS0 course has been rewritten. The course now includes daily quizzes, which focus on the subject to be discussed in each respective lesson. The quizzes usually contain at least one question randomly chosen from any previous lesson. A direction for future work would include defining two separate question pools for the daily quizzes (assigned to two separate groups of students), one of which contained more cumulative questions and one focused around the current lesson. This approach would better quantify the effect on performance based on the distribution of subjects from which quiz questions were generated, something we did not explicitly and measure in this study.

## REFERENCES

- [1] T. Babbitt, C. Schooler, and K. King. Punch Cards to Python: A Case Study of a CS0 core course. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education, SIGCSE '19*, pages 811–817, New York, NY, USA, 2019. ACM.
- [2] R. L. Bangert-Drowns, J. A. Kulik, and C.-L. C. Kulik. Effects of frequent classroom testing. *The Journal of Educational Research*, 85(2):89–99, 1991.
- [3] J. R. Blunt and J. D. Karpicke. Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, 106(3):849, 2014.
- [4] P. C. Brown, H. L. Roediger, and M. A. McDaniel. *Make it stick*. Harvard University Press, Cambridge, Massachusetts, 2014.
- [5] S. Cass and P. Bulusu. Interactive: The top programming languages 2018, July 2018. [Online; posted 31-July-2018].
- [6] R. B. Cattell. Abilities: their structure, growth, and action. 1971.
- [7] V. A. Cicirello. On the role and effectiveness of pop quizzes in cs1. In *ACM SIGCSE Bulletin*, volume 41, pages 286–290. ACM, 2009.
- [8] J. L. Dobson, J. Perez, and T. Linderholm. Distributed retrieval practice promotes superior recall of anatomy information. *Anatomical sciences education*, 10(4):339–347, 2017.
- [9] D. Glenn. You Will Be Tested on This. *Chronicle of Higher Education*, 53(40):A15–A17, 2007.
- [10] M. Guzdial and B. Ericson. *Introduction to computing and programming in python*. Pearson, 2016.
- [11] J. Hussey and J. Shaha. Educational Approach to Cyber Foundations in an Undergraduate Core Program. In *Proceedings of the 18th Annual Conference on Information Technology Education*, pages 21–26. ACM, 2017.
- [12] J. D. Karpicke. Retrieval-based learning: Active retrieval promotes meaningful learning. *Current Directions in Psychological Science*, 21(3):157–163, 2012.
- [13] J. D. Karpicke and P. J. Grimaldi. Retrieval-based learning: A perspective for enhancing meaningful learning. *Educational Psychology Review*, 24(3):401–418, 2012.
- [14] J. D. Karpicke and H. L. Roediger. The critical importance of retrieval for learning. *science*, 319(5865):966–968, 2008.
- [15] J. D. Karpicke and H. L. Roediger III. Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57(2):151–162, 2007.
- [16] M. A. McDaniel, K. M. Wildman, and J. L. Anderson. Using Quizzes to Enhance Summative-assessment Performance in a Web-based Class: An Experimental Study. *Journal of Applied Research in Memory and Cognition*, 1(1):18–26, 2012.
- [17] J. S. Nevid, Y. S. Pyun, and B. Cheney. Retention of text material under cued and uncued recall and open and closed book conditions. *International Journal for the Scholarship of Teaching and Learning*, 10(2):10, 2016.
- [18] C. Norris. An examination of layers of quizzing in two computer systems courses. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, pages 48–53. ACM, 2016.
- [19] T. G. on Information Technology Curricula. *Information Technology Curricula 2017: Curriculum Guidelines for Baccalaureate Degree Programs in Information Technology*. ACM, New York, NY, USA, 2017.
- [20] M. A. Pyc and K. A. Rawson. Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, 60(4):437–447, 2009.
- [21] H. Roediger. How tests make us smarter. *The New York Times*, page SR12, 2014.
- [22] H. L. Roediger III and J. D. Karpicke. The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3):181–210, 2006.
- [23] H. L. Roediger III and J. D. Karpicke. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*, 17(3):249–255, 2006.
- [24] C. A. Rowland. The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6):1432, 2014.
- [25] O. Yasar, P. Veronesi, J. Maliekal, L. Little, and J. Tillotson. Computational Pedagogical Strategies to Help Learners Retain STEM Content Knowledge. In *Society for Information Technology & Teacher Education International Conference*, pages 351–360. Association for the Advancement of Computing in Education (AACE), 2019.
- [26] D. Zingaro, C. Bailey Lee, and L. Porter. Peer instruction in computing: the role of reading quizzes. In *Proceeding of the 44th ACM technical symposium on Computer science education*, pages 47–52. ACM, 2013.