

THE TESTING EFFECT

An evaluation of Achieve Read & Practice and the investigation of whether retrieval practice helps close the achievement gap in Psychology courses



Kara McWilliams PhD, Vice President Impact Research, Macmillan Learning

Alexandra Stone, University of Connecticut

Ethics and peer review. This research complied with APA ethical standards for research. It was approved by a third-party Institutional Review Board (IRB) prior to recruitment, and then approved by individual institutional IRBs at each participating institution where required. This paper and the results herein, have been peer reviewed, revised, and approved for publication as a Macmillan Technical Report by the Impact Research Advisory Council, a panel of experts in applied research, measurement, educational technology, and the learning sciences.



Contents

- 3 Foreword**
- 4 Acknowledgements**
- 5 Introduction**
- 7 Literature Review**
 - Retrieval Practice
 - Reading Quizzes
 - Differential effectiveness
 - Summary
- 10 Achieve Read & Practice**
 - Study Procedures
 - Research Questions
 - Data Collection
- 15 Sample Descriptions**
- 17 Results**
 - Research Question 1
 - Research Question 2
 - Research Question 3
 - Variables
 - Model Building
 - Research question 4
- 34 Discussion**
- 36 Implications for Instruction**
- 37 Limitations and Future Research**
- 38 Note on Data Privacy**



Foreword

At Macmillan Learning we are committed to providing our instructors and students with practical, actionable, and timely insights derived from studies that meet standards for educational and psychological testing. Our goal is to improve teaching and learning by enabling evidence-based decision making and to contribute to the methods and outcome research on digital learning tools. To that end, we take a comprehensive approach to measuring the effectiveness and efficacy of the digital learning tools we produce. Beginning in development, and continuing through use at scale, we partner with instructors and students to conduct studies that are appropriate for the tool's stage in the development lifecycle. Each study contributes unique and increasingly rigorous evidence to the validity and efficacy argument of that tool. Studies also produce insights into usage and engagement patterns among educational contexts that instructors might consider implementing in their own courses. This report represents one study that makes up the larger body of Achieve efficacy research. We are confident in this approach but acknowledge that measuring efficacy is complex, and we are always learning. The authors of this report, and the impact research team as a whole, welcome any comments or feedback on this report or our approach to measuring efficacy.

Kara McWilliams PhD, Vice President Impact Research, Macmillan Learning



Acknowledgements

The work in this report reflects the efforts of many people beyond the authors. We appreciate the commitment that Macmillan Learning has made to ensuring that the tools we are developing are meeting the needs of instructors and students, and improving higher education. We thank our Chief Executive Officer, Ken Michaels for his continued support on this journey. This study would not have been a success without the partnership of Gareth Hancock and Tim Flem's Product Team, Chelsea Valentine's Technology Team, Alicia Nachman's User Experience team, and Susan Winslow's Editorial group. Our partner instructors and their students were fortunate to have the support of Marcie Smith's Customer Experience team throughout the study. To each of the leaders at Macmillan and their talented teams, our gratitude.

In an effort to offer timely, peer reviewed insights to instructors, we are grateful to the Impact Research Advisory Council for their peer review of this report. Their guidance and critique since we began developing our approach to efficacy, ongoing insight throughout each study, and honest reviews of findings have been invaluable. Chris Dede, Michael Feldstien, Sara Finney, Suzanne Lane, Thanos Patelis, and Elana Zieda we are indebted to you.

We also thank our colleagues on the Learning Research and Learning Analytics teams and our Chief Learning Officer, Adam Black for their reviews and helpful feedback on early drafts of this report.

Most importantly, we want to thank the instructors and students who partnered with us on this study and those who partner with us on every efficacy study we conduct. Your close partnership, keen insight, and honest feedback enable the continual optimization of the tools we develop. For the work presented in this report we particularly thank the administrators, instructors, and students at the following institutions:

Borough of Manhattan Community College, Duquesne University, El Paso Community College, Georgia Military College - Augusta Campus, Middlesex Community College, Montgomery College Rockville, Morrisville State College, Onondaga Community College Polk State College, Prince George's Community College, Rogers State University, University of Cincinnati, University of Mary Washington.



Introduction

Retrieval practice, generally facilitated by including practice tests in one's learning regime, contributes to comprehension and memory (Endres, 2017; Carpenter, et al., 2016; Agarwal 2019). The learning sciences refer to this phenomenon as the “testing effect”. An emerging body of literature compares outcomes between students who studied and engaged in practice tests and those who were exposed to extra studying but did not engage in practice tests. The findings have suggested that summative test performance was significantly better among students who had engaged in practice tests (Roediger & Karpicke, 2006a; Toppino & Cohen, 2009). Researchers conclude that because the testing effect improves retention and recall, students then perform better on subsequent final assessments. Research has also shown that repeated practice testing positively influences summative assessment scores (e.g., Karpicke & Roediger, 2007b, 2008; Roediger & Karpicke, 2006a).

Another body of research suggests that assigning activities before class that are related to the content that will be covered during class time positively influences summative exam scores. Researchers postulate that because pre-class formative assessment gives students a basic understanding of vocabulary and core concepts of a topic, they come to class better prepared to engage. And, that this basic understanding enables instructors to reduce the time they spend on direct instruction and focus more on higher-order thinking and active learning (Brown, Roedinger, & McDaniel, 2014).



Achieve Read & Practice is a new adaptive gamified reading and quizzing system that was built based on the learning science of retrieval practice. Read & Practice was available for instructors to use at scale beginning in the Fall 2018 semester. In advance of release, the authors conducted a set of focused implementation studies with a targeted set of instructors in the tool's beta form, but the effectiveness of the fully developed tool had never been examined with a representative set of instructors. And, in general, while there is a strong body of research around reading quizzes and pre-class assessments at four-year institutions, there is little evidence of effectiveness among the important populations at two-year institutions. Consequently, this study investigated how instructors from various educational contexts (two-year and four-year institutions) and backgrounds (length of time teaching, title, comfort with technology) chose to use Read & Practice, whether use was related to learning outcomes, and whether there were any differences in the relationships based on the instructor's selected implementation model or students' level of motivation to succeed in the course and/or their academic preparedness.

Although existing research has demonstrated that the testing effect positively influences summative performance, it is important to systematically evaluate the tools that are delivering the reading and assessment. Since Read & Practice is a new tool that is being widely used, the study expands the literature by contributing the first evidence of effectiveness of this new tool. The research also expands the literature on pedagogical strategies in higher education by systematically examining whether different implementation patterns are more effective. And, to contribute to the literature on the testing effect, we investigated whether voluntarily re-taking quizzes for practice positively influenced exam scores when student prior academic performance, baseline level of motivation, and instructor were controlled.

If results suggested that use of Achieve Read & Practice was related to stronger exam scores, and if certain implementation models were more influential, it could help instructors make

adoption and usage decisions in their course. Also, if findings suggested that re-taking quizzes for practice positively influenced exam scores, instructors might consider influencing their students to retake quizzes as part of their preparation for in-class assessments. Possibly most important, if the results demonstrated that use of Read & Practice could help bridge the gap between students more academically prepared to succeed and their less prepared peers at two- and four-year institutions it could provide insight into closing the higher education achievement gap at scale. If a positive relationship between the use of the tool and exam scores did not emerge, the development team could use the findings to optimize Achieve Read & Practice; and a replication study would be conducted to evaluate whether improvements to the product changed the outcomes.

In the Spring 2019 semester, thirteen instructors from thirteen institutions elected to use Read & Practice as the primary curricular material in their Introductory Psychology course. Of the 847 students enrolled across the courses, 670 (79%) consented to participate in this study. Instructor implementation patterns were not mandated as part of the study and instructors used the tool in different ways. Consequently, students engaged with the tool in different ways. For example, some instructors chose to assign Read & Practice prior to the course in which the content would be covered (the "flipped model") and some instructors chose to assign it after the class in which the content would be covered. In all courses Read & Practice was assigned for credit. None of the instructors required that students retake quizzes for practice or offered extra credit for doing so, but some students elected to retake quizzes for practice.

This paper provides the theoretical background for the research, offers a complete description of the study procedures, provides the technical details of the analyses, outlines the results, and discusses the implications for instructors and how the results contribute to the broader educational research.



Literature Review

To provide context and a theoretical grounding for this study the authors first reviewed the literature on retrieval practice (testing effect) and reading quizzes.

RETRIEVAL PRACTICE

An extensive body of research supports claims of improved academic performance among students in various disciplines in higher education courses when they used retrieval practice (Endres, 2017; Carpenter, et al., 2016; Agarwal 2019). Research from the past five years emphasizes study of retrieval practice in classroom settings to examine how varying approaches influence students' course-specific success (Carpenter, et al. 2016). These methods enable analyses of factors like individual student performance, complexity of learning tasks, and types of retrieval tasks most effective for course-specific content, classroom setting, and individual student performance (Endres, 2017; Carpenter, et al., 2016; Agarwal 2019).

Building on the research base that retrieval supports comprehension, Endres, et al. (2016) investigated whether there were differences in learner outcomes based on retrieval types. The authors hypothesized “that the expected effect on comprehension is mediated by the elaboration strategies employed during recall (Mediation Hypothesis)” (Endres, et al., 2016). Fifty-six undergraduate students were randomly assigned to a free recall condition or a prompted recall condition. Students in the prompted recall condition were provided an elaborative prompt with the assumption that this would influence elaboration strategies among students. Results showed that elaborative prompts enhanced the learning effects of retrieval practice, however it was unclear whether all types of prompts would yield the same outcomes. Thus the researchers suggested that more research should be conducted on whether various recall conditions supported comprehension.



In addition to differences in recall conditions, researchers are interested in understanding whether there are individual differences in learning from recall. Carpenter, et al (2016) investigated the differences in effects of retrieval practice among cohorts of students in an introductory Biology course. Research questions examined differences in outcomes between students classified as “middle-and-low performing students” and “high performers” based on their scores on four mandatory exams they had previously completed for the course. The relationship between use of retrieval practice and quiz scores was stronger for students classified as high-performing as was the ability to predict future performance on quizzes based on exposure to retrieval practice. Middle-performing and lower-performing students benefited more from copying correct answers than being asked to retrieve course-specific information and apply it through more dynamic exercises. While the study supports research about the effectiveness of retrieval practice, the authors encourage further examination of the relationship of individual student achievement in relationship to effectiveness of retrieval practice, noting that while “high performers appear to get better at aligning their predictions with performance as a result of practice, middle and low performers may be in need of additional metacognitive training to improve calibration” (Carpenter et al., 2016).

Agarwal (2019) examined the role of retrieval practice in the development of higher order thinking skills. Across three experiments (two 4x2 within-subject designs with two groups of college students, and a 3x2 within-subjects design with a group of sixth grade students), Agarwal observed retrieval with consideration for varying conditions and delayed test types (fact test, higher-order test). Students exposed to retrieval practice earned higher scores on final tests than those who simply copied course material, but only when test questions matched the retrieval practice type (fact learning questions vs higher-order learning questions). Agarwal concluded that students were more successful in developing higher-level thinking skills when they engage in higher-level retrieval practice as opposed to “building a foundation of knowledge via fact-based retrieval practice”, challenging popular assumptions about the need for building foundational knowledge before practicing higher order skills as defined in Bloom’s taxonomy (Agarwal 2019).

READING QUIZZES

There is a substantial body of research that supports the claims that online reading quizzes promote students completing assigned reading and positively influence student academic performance on subsequent assessments. Most of the available literature from the past five years investigates reading quizzes that were assigned for credit prior to the student attending the class in which the content would be lectured on. To our knowledge, there is no published research that compares various implementations of assignment of reading quizzes including a combination pre- and post-implementation.



To extend the research on active learning and the flipped classroom, Hodges et al. (2015) explored whether there were differences between the effectiveness of paper based or online reading quizzes that were assigned before class. The authors presented case studies of various implementations of reading quizzes that were designed to focus on student-centered learning and developing questions associated with the readings that were the appropriate level. The authors concluded that a highly structured course that implemented reading quizzes “pre-class” positively influenced student learning, which reiterated to the findings of a study by Freeman, et al (2011) that investigated whether a highly structured course that implemented pre-class reading quizzes and/or active learning in the classroom could reduce failure rates in Introductory Biology courses. After controlling for students’ prior academic performance, results showed that the failure rate was dramatically lower among students in highly structured courses that implemented both pre-class reading quizzes and active learning. The authors suggested that reading quizzes may help bridge the gap between less academically prepared students and their more academically prepared peers.

Pape-Lindstrom, et al. (2018) investigated the effect of pre-class reading quizzes among community college students by assigning students to one of three control groups and one of three experimental groups. Students in the experimental groups realized an average 4.9% gain in exam scores as compared to the students in the treatment groups. The authors suggested that any or all of the following hypotheses gleaned from previous research on the topic could explain the benefits of reading quizzes: pre-class preparation through reading quizzes may allow students to gain better mastery during in-class time, that regular practice of taking reading quizzes contributes to a testing effect and results in higher subsequent exam scores, and/or the structure by which the assessments in the course were scored and weighted might benefit test-anxious students (Gross et al., 2016, Cassidy & Johnson, 2002, Roediger & Butler, 2011).

A substantial body of research also investigates the efficacy of gamified reading quizzes as compared to static reading quizzes. In a five-year longitudinal study, Barata et al. (2013) compared the first three years of non-gamified class assignments and two successive years of experimental gamified quizzing. Results showed significantly higher grades among students participating in the gamified assignments. Additionally, those students showed improvements in attention to reference materials, online participation, and proactivity (Barata et al., 2013).

SUMMARY

A combination of a highly structured (flipped) classroom model and the implementation of pre-class, gamified quizzes/activities may be the most fitting combination, placing higher expectations on students to commit sufficient hours of learning outside of class while enabling the testing effect to support comprehension and future recall of information.

Achieve Read & Practice was built on the foundation of the learning science that underpins the constructs discussed in the literature review. With a solid grounding in the literature and the learning science of retrieval practice and gamified reading quizzes, the next section describes the digital learning tool studied and the study design.



Achieve Read & Practice

Achieve Read & Practice (Read & Practice) is one tool in the broader suite of next generation digital learning tools, Achieve. The learning tool includes an eBook with embedded and adaptive quizzes. The eBook displays text in sections with various tools to support online reading, such as highlighting and note-taking. In order to encourage students to read the eBook and then practice what they have learned, students are periodically directed to adaptive quizzes with the goal of “filling in the bar” on a progress meter—combining formative assessment and retrieval practice with the gamification of having to “meet the target”. Quiz questions change depending on the topic and difficulty level, giving a student more practice in areas where they most need it. If a student needs extra support on the quiz, they can access hints or request to be taken back to the eBook section where the answer can be found. Accessing a hint reduces the amount of points that a question is worth, requiring the student to answer additional questions in order to fill in the bar. However, other types of support (e.g. reviewing eBook sections) do not trigger a point deduction because the learning tool intentionally reinforces these types of study habits. Once students have filled in the bar, the learning tool provides them with a study plan to review the eBook section(s) with which they need the most support, and the students have the option of retaking the quiz to improve their knowledge of those sections. Throughout the experience, students can access visual metrics (i.e. individual analytic dashboard) to guide them on where to focus their learning, and instructors can access visual metrics (i.e. course-level and student specific analytics dashboards) to help them tailor their lectures or provide extra support



STUDY PROCEDURES

This research complied with APA ethical standards for research. It was approved by a third-party Institutional Review Board (IRB) prior to participant recruitment, and then approved by individual institutional IRBs where required.

In the Spring 2019 semester, 13 instructors teaching Introduction to Psychology agreed to participate in an evaluation of Read & Practice. Instructors and students received Read & Practice free of charge to use. All students were required to use Read & Practice in their course because it was the curricular material their instructor selected and completion of assigned activities contributed to student final course grades. Students were not required to participate in this study, those interested in participating were required to provide active informed consent. The 670 consenting students in the study made up 79% of all students enrolled in participating courses (note analytics samples were reduced to 658 due to data missingness, and then reduced as needed for individual analyses as described in the results section).

Prior to the beginning of the semester instructors were given the option of meeting with a customer experience specialist for a thirty-minute training on Read & Practice. Four out of the 13 instructors participated in training. During training instructors were offered suggestions for implementations, but implementation patterns were not mandated as part of the study for any instructors. The only implementation requirement was that another adaptive quizzing tool could not be used alongside Read & Practice that semester.

Instructors were not required to use a specific textbook with Read & Practice, as we wanted to be able to explore whether there were any differences in learning outcomes based on the eBook used with Read & Practice. The texts that instructors elected to use can be found in Table 1.

Instructors were asked to use Read & Practice in their teaching as they would elect to even if they were not in a study, quantitative data were collected throughout the study to answer the research questions, and qualitative data were collected to contextualize the quantitative findings.

Table 1. Textbooks used with Read & Practice

Count of Instructors	Textbook
1	Abnormal Psychology Comer & Comer
1	Developing Person through Childhood and Adolescence Berger
2	Development Person through the Lifespan Berger
2	My Psychology Pomerantz
1	Psychology Hockenbury & Nolan
3	Psychology in Everyday Life Myers & DeWall
1	Psychology in Modules Myers & DeWall
2	Scientific American: Psychology Lict, Hull, & Ballentyne
1	The Development of Children Lightfoot



RESEARCH QUESTIONS

The study investigated four research questions, each with a specific goal of supporting teaching and learning in higher education Psychology courses.

- 1.) *How is Read & Practice implemented in higher education Psychology courses, and does student use vary by implementation model and/or educational context?*
- 2.) *What are instructor and student perceptions of Read & Practice? Do perceptions vary by educational context and/or implementation model?*
- 3.) *What is the relationship between use of Read & Practice and performance on in-class exams and does student motivation, prior academic performance or who the instructor was moderate that relationship?*
- 4.) *Is voluntarily retaking quizzes in Read & Practice related to higher in-class exams, does student motivation, prior academic performance or who the instructor was moderate that relationship?*

DATA COLLECTION

Data were collected for a mixed-methods analysis. Student and instructor surveys were administered at the beginning and end of the semester, instructors completed weekly implementation logs, and instructor interviews were conducted mid-semester. Product usage data were extracted from the Achieve platform on a weekly basis and at the end of study, and student records were shared by instructors at the end of the semester. Data were matched across sources, and descriptive and empirical analyses were conducted. A complete description of the collected data follows.

STUDENT PRE-SURVEY. During the first two weeks of course, instructors shared a link to an online survey that asked students to first consent to participate in the study and then report their background and demographic characteristics, as well as experience and perceptions. The survey captured data on student comfort with technology, student sentiment toward technology use in the classroom, values of digital tools in the classroom, academic behaviors outside of class, classroom behavior, and sentiment toward the course. Students were also asked to report their major, whether they were taking this course as part of their major requirement, high school grade point average, whether they took the SAT and/or ACT, and their scores on each section, as well as various demographic data. These data were collected as potential moderators of the relationship between use of Read & Practice and academic performance and were used in the analyses of the research questions.¹

1: Although students self-report their measures of prior academic performance, we can have confidence in the reliability of the scores based on previous research. Shaw and Matter (2009) examined the reliability of self-reported HSGPA and found a correlation of 0.74 and in a 2005 meta-analysis. Kuncel, Credé, and Thomas found a correlation of .82 between actual and self-reported SAT scores.



INSTRUCTOR PRE-SURVEY. An online survey that asked instructors to report their background and demographic characteristics, as well as their experience and perceptions, was administered online during the first month of the courses. The survey included a scale that measured acceptance of technology and included items about comfort with technology, perceptions of technology in the classroom, intended implementation of Read & Practice, intended implementation of other publisher-provided digital learning tools or open educational resources, previous experience with Read & Practice, and general early perceptions of the tool. These data were used to control for instructor characteristics and to better understand intended implementation of Achieve Read & Practice.

INSTRUCTOR WEEKLY IMPLEMENTATION LOGS. An online survey was sent to instructors at the end of each week. The survey asked instructors to report how they implemented Read & Practice in the previous week (which features and components they used), how much time various activities took them, their perception of Read & Practice that week, any benefits or challenges of using Read & Practice, and any other information that would help us understand usage that week (e.g. whether class was canceled for inclement weather). These data were used to track ongoing implementation and how that was related to perception.

SITE VISIT. Researchers visited a sample of instructors' classrooms to document the environment, observe the instructor's pedagogy, and conduct a focus group with willing students. Focus group questions were based on the research questions and probes were developed in real-time based on responses to questions in the focus group protocol.

INSTRUCTOR INTERVIEWS. An instructor interview protocol was developed that gathered information on how an instructor was implementing Read & Practice, why they decided to implement it in that way, their perceptions of the tool, and their perceptions of how their students were accepting the tool. Probes were developed based on the responses provided in implementation logs and in real-time based on responses to questions in the interview protocol. These data were used to contextualize implementation patterns observed in the platform data.



INSTRUCTOR POST-SURVEY. An online survey was administered during the last two weeks of the course. The survey included a scale that measured instructor ability to assess student understanding; a scale to assess active learning in the classroom; a scale to measure classroom challenges; a scale to measure student behavior, their implementation of Read & Practice and their perceptions of Read & Practice; a System Usability Scale;² and the net promoter score, and a likelihood of adoption scale. These data were used to measure whether there were systematic differences between instructors who assigned pre-class activities and those who didn't.

STUDENT POST-SURVEY. Instructors shared an online survey with their students during the last two weeks of the semester. The survey asked students to share demographic data, personal device data, how they used Read & Practice, their perceptions of the tool, their engagement in the course³, their satisfaction with the course, a System Usability Scale, and a Net Promoter Score. These data were used to measure student perceptions of Read & Practice and whether outcomes varied among implementation pattern or educational contexts.

PRODUCT USAGE DATA. The following student data were extracted from the Achieve Read & Practice platform for consenting students: student name, student email, each activity that an instructor assigned, assignment date and due date, whether student accessed each activity, student progress on each assigned activity, student completion of each assigned activity, student performance on each assigned activity, student access of unassigned activities, student progress on unassigned activities, student completion of each unassigned activity, student performance on each unassigned activity.

STUDENT RECORDS. Instructors were asked to share the following course performance data for consenting students: homework scores, quiz scores, exam scores, final exam scores, final course grades and percentages, attendance rate, and participation scores. Instructors were not asked to change their regular course performance methods, so not every variable was available for each instructor. For example, some instructors did not score homework or give quizzes, so they only reported exam scores and final course grades. And, not all class records were provided in the same metric, so only grades that could be reliably compared were included in the analysis.

2: Brooke, J. (1986). System Usability Scale. Digital Equipment Corporation.

3: Handelsman, M. M., Briggs, W. L., Sullivan, N., & Towler, A. (2005). A measure of college student course engagement. *The Journal of Educational Research*, 98 (3), 184-191.



Sample Descriptions

INSTITUTIONAL. Students and instructors from 13 institutions participated in this study. In total, eight institutions are two-year and five institutions are four-year. The institutions in the study are geographically diverse with seven located in the Northeast, four located in the South, and one located in the Midwest.

INSTRUCTORS. Thirteen instructors of varying backgrounds, experiences, and titles participated in the study. In total, six instructors had been teaching between six and ten years, three instructors had been teaching between one and five years, two have been teaching for more than 11 years, and it was the first year teaching for one instructor. Four instructors had the title Professor, four had the title Assistant Professor, three had the title Adjunct Professor, one reported their title as Associate Professor and one as Lecturer. Most (11) were employed full-time at the institution and four reported being employed part-time. Most (eight) instructors in the sample had used a digital education product in their higher education course in the past - five of those instructors had used a different adaptive, gamified quizzing tool previously. Class sizes represented in this study varied as well, with the largest group (seven) being between 30 and 50 students, four courses having fewer than 30 students, one course having between 51 and 100 students enrolled, and one course had more than 100 students enrolled. Instructors in the sample tended to have strong positive perceptions of digital learning tools in the classroom and reported being comfortable implementing them in their course.



STUDENTS. Across instructors, 847 students were enrolled during the semester this study was conducted. In total, 670 consented to participate in the study representing a 79% participation rate. Participation rates varied by instructor and ranged from 60% to 94%. Most students (52%) were in the second semester of their first year of college. The average self-reported HSGPA at graduation was 3.35. Slightly less than one third of the sample (31%) reported being the first in their family to attend college. In total, 75% of students had used digital learning tools in a college course before and 62% of those students reported enjoying using them. About 30% of students reported having taken Psychology in high school, and 15% were majoring in Psychology at the time of the study.

Although there were no survey data available to compare the group of participating students to the 21% of students who elected not to participate, we did compare aggregate Read & Practice usage and performance between consenting students and anonymous data from the other students in the participating courses. There were no meaningful differences between the group of participating students and non-participating students in terms of engagement (71% and 72%, respectively), completion (70.3% and 69.6%, respectively), or performance (70.3% and 69.6%, respectively). Thus, it is reasonable to assume that the group of participating students is a good representation of the target population of students as it relates to use and performance.



Results

RESEARCH QUESTION 1

How is Read & Practice implemented in higher education Psychology courses, and does student use vary by implementation model and/or educational context?

We hypothesized that instructors would implement Read & Practice regularly throughout the semester. We further hypothesized that instructors would tend to assign Read & Practice activities pre-class—that is they would assign activities within Read & Practice as a way to promote reading and to introduce students to material before they came to class. Research has demonstrated the benefit of promoting higher-order cognition in the classroom, so we expected that instructors would choose to apply Read & Practice as a tool to transfer direct knowledge before students came to class, thus enabling active learning during class time. And, since Read & Practice offers instructors analytic reports of student performance, we expected that they would use those insights to adapt or focus their in-class lectures and/or planned activities. Similarly we hypothesized that students would use Read & Practice at a high rate because of its engaging game-like features and because prior research conducted by the authors suggested that students had positive perceptions of Read & Practice and believed that it supported their academic performance in Psychology (Macmillan Research Note, 2018).



Table 2. Proportion of weeks in the semester that instructors assigned at least one Read & Practice activity to be completed, when it was assigned, and proportion of final grade it accounted for

Instructor	% weeks Read & Practice assigned	When Read & Practice was assigned to be completed	% of final course grade Read & Practice accounted for
Two-year institutions			
1	90	post-class	10
2	90	pre-class	20
3	90	pre-class	48
4	92	post-class	10
5	92	pre-class	17
6	100	combination	100
7	100	combination	35
Four-year institutions			
8	27	combination	77
9	73	pre-class	5
10	75	combination	25
11	77	combination	30
12	92	post-class	10
13	100	pre-class	6

INSTRUCTOR IMPLEMENTATION During the Spring 2019 semester, instructors in the sample assigned an average of 23 Read & Practice quizzes (i.e. activities) with the proportion of activities assigned by each instructor ranging from 11 to 35. On average, instructors assigned Read & Practice activities in 86% of the weeks in the semester, and activity assignment by individual instructors ranged from 27% to 100% of weeks in the semester. Instructors who assigned Read & Practice regularly reported that the only times they did not assign Read & Practice activities were during weeks in which classes were not held. One instructor reported assigning Read & Practice at a low

rate due to substantial technical difficulties with the tool. However, researchers classified difficulties as user experience and usability issues rather than technical difficulties; they also noted that the challenges were specific to that instructor.

Instructors teaching at two-year institutions tended to assign Read & Practice activities during more weeks of the semester than instructors teaching at four-year institutions, and instructors newer to teaching in higher education tended to implement Read & Practice in fewer weeks of the semester than instructors with more experience in higher education. The four instructors teaching



five years or less assigned activities in either 27%, 77%, 75%, or 92% of the weeks of the semester. While the three instructors teaching more than ten years assigned it either 90% or 100% of the semester. Most instructors in the sample reported being extremely comfortable (on a scale of 1 = “extremely uncomfortable” through 4 = “extremely comfortable”) with technology. Two of the three instructors who rated themselves a 2 (“somewhat comfortable”) had high implementation rates—with the exception of one instructor (Instructor 8)—who only used Read & Practice in slightly more than a quarter of the weeks of the semester.

The proportion of a student’s final course grade that was determined by Read & Practice ranged from 5% to 54% with the average across instructors being 20%. There were no meaningful differences in the weight of the final course grade based on educational context or instructor characteristics.

All instructors who assigned Read & Practice activities pre-class reported doing so to promote students staying on track with the reading and getting a basic understanding of the material. Instructors who assigned Read & Practice post-class reported doing so as a comprehension check of material taught during the in-class lecture. The five instructors who reported assigning Read & Practice pre- and post-class were asked to elaborate on their use of the tool in this way. They tended to respond that they assigned quizzes pre-lecture to prepare students for class and post-lecture to reinforce concepts. A quote representative of these responses is “pre-lecture for overview of what will be discussed; post-lecture for review of what they did at beginning of week to end of week - comprehension check [sic]”.

None of the instructors who had been teaching higher education courses between six and ten years were “combination” assigners. However, three of the four instructors who had taught five years or fewer and two out of three that had taught more than ten years were “combination” assigners.

ANALYTICS DASHBOARDS. Most instructors (76%) reported reviewing the instructor-facing analytic dashboards available in Read & Practice at least once a week. The most frequently reported reason for reviewing the dashboards was to identify students that were struggling, for example one instructor reported, “primarily I used the analytics to determine how well each class performed in

comparison to the other classes. I also used them to determine which students appeared to be having problems with understanding the readings (or who seemed to be “blowing off” those readings).” Instructors also reported using the dashboards to understand comprehension and determine whether their lecture should be modified, for example, “[I checked the dashboards] to see where the students that were participating were struggling so that I could tailor my lecture accordingly.” On their weekly implementation logs instructors reported accessing the insights an average of 80% of the weeks of the semester. The three instructors who did not access dashboards reported this was because (1) there was not enough time to do so, (2) technical difficulties (which were actually user experience and usability challenges) and/or (3) they found the information to be too general to be helpful.

Of the ten instructors that reported checking the dashboards, 70% (7) reported that they used the insights provided in the reports to intervene with a student or group of students, most often to probe about why they weren’t keeping up with the reading. For example, one instructor reported, “I talked with struggling students and asked ‘why’ not completing the work? [sic]” Six of these ten instructors (60%) reported that they used the insights to modify or focus their classroom lecture. For example, one instructor reported, “I reviewed the overall performance for each previous week with them at the beginning of the subsequent week—not 100% of the time, but for more than 50% of the term.”

STUDENT USAGE. Students were coded as either “pre-class only” (n=272, 41%), “post-class only” (n=151, 23%), or “combination” (n=235, 36%) based on who their instructor was, and how they decided to implement Read & Practice. None of the participating instructors required students to retake quizzes after first submitting them for credit. However, some students voluntarily re-took quizzes. Platform usage data were examined to code students as “retakers” or “non-retakers”. A student was coded as a “re-taker” if they had launched and met the target on at least one activity that they had previously submitted for credit. In total, 237 students (43%) were coded as a “re-taker” and 312 (57%) were coded as a “non re-taker”. The comparison descriptive statistics can be found in Table 3.

**Table 3. Student group and descriptive comparison by percentage of students**

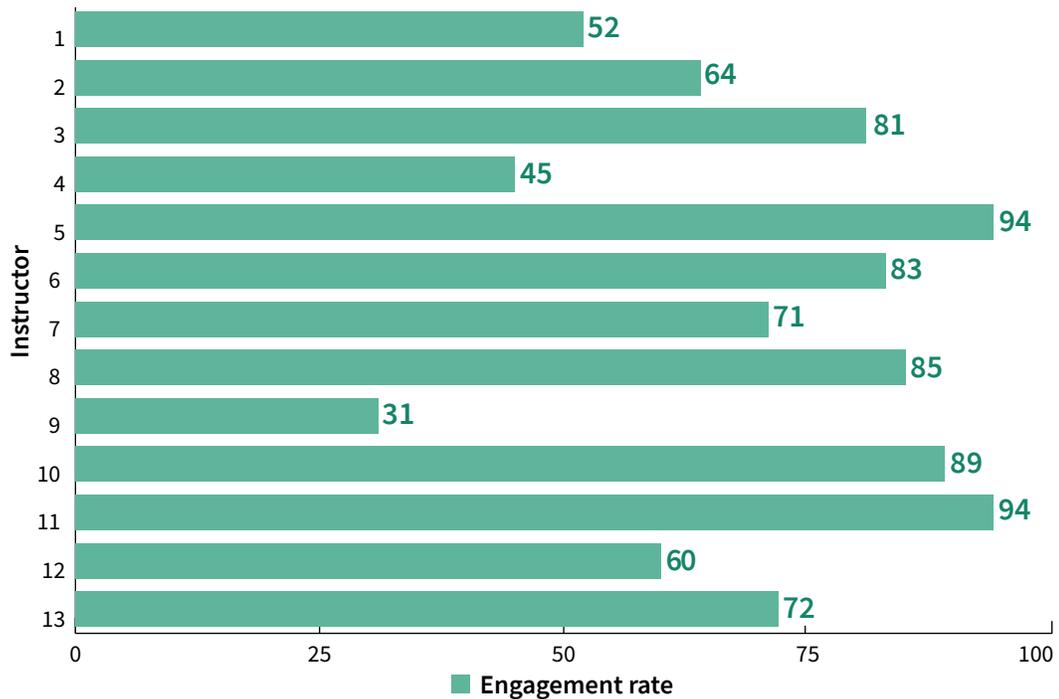
	All (n=658)	Pre-class only (n=231)	Post-class only (n=151)	Combination User (n=235)	Retaker (n=231)	Non-retaker (n=309)
Year in school						
Dual Enrolled	6.22	13.22	1.69	1.90	4.68	8.33
First	52.44	47.70	55.93	5.06	50.88	53.70
Second	26.22	30.46	25.42	22.15	25.15	24.07
Third	9.33	4.02	10.17	14.56	15.20	5.56
Fourth	3.33	<1	5.08	5.06	2.92	4.63
Fifth	<1	1.72	0.00	0.00	<1	<1
Other	1.78	2.30	1.69	1.27	<1	2.87
Eligible for federal financial aid (yes)	67.41	60.57	80.51	65.19	67.44	66.67
First generation (yes)	30.60	34.29	32.20	25.32	33.14	29.17
Majoring in psychology	14.73	9.25	14.53	20.89	15.79	12.50
Took psychology in high school	29.46	27.17	25.64	34.81	28.65	30.09
HSGPA	3.35	3.37	3.06	3.51	3.47	3.31
Less motivated to succeed	30.28	21.63	32.79	38.86	28.64	32.22
Less academically prepared	49.19	60.29	61.63	32.00	51.97	40.78
Average exam scores	80.12	80.56	76.15	84.37	82.26	78.97
Average course grade	83.50	85.69	77.84	84.72	86.46	82.49

All students engaged in Read & Practice quizzes at a relatively high rate. Of all activities assigned, students engaged in an average of 71.2%. Average engagement rates ranged from 31% to 94% across instructors as presented in Figure 1.

There were differences in engagement rates by when Read & Practice was assigned. “Combination” students had the highest average engagement rate at 81.9%, followed by “pre-class only” students at 74.4%. Students assigned Read & Practice post-class only had the lowest engagement rate, which was 50.6%.



Figure 1. Average rate of students engaging in assigned Read & Practice activities by instructor



The differences in engagement rates by implementation model were statistically significantly different: $F_{521} = 54.00$, ($p < 0.0001$). The engagement rate among “post-class only” was so different that we connected with the instructors of those students to inquire whether there was something about their implementation models that would influence such a low engagement. While there wasn’t anything about the implementation that would influence engagement rates, note from Table 3 that the proportion of a student’s final course grade that Read & Practice accounted for is lower for “post-class only” students than for the other two implementation models, which likely contributed to lower engagement rates. Another hypothesis is that students who completed activities after class may have felt that they had already received the instruction that they needed from the class lecture, but this is a hypothesis that requires testing.

A student meeting the target score set by the instructor is a good proxy for “performance”, as a student has to persist through the appropriate number of correct responses before meeting the target score. Once they meet the target score, they earn a 100% on that activity. If they do not persist

to the target score, they earn a 0 for that activity. In this study students met the target score at a relatively high rate (70.3%). Like engagement rate, there was variability in the proportion of students who met the target score by implementation model. “Combination” students had the highest average engagement rate at 80.5%, followed by “pre-class only” students at 74.2%. Students assigned Read & Practice post-class only had the lowest engagement rate at 49.5%. The distribution of the proportion of students meeting the target score can be found in Figure 2.

Surprisingly, the proportion of the final course grade that Read & Practice made up did not substantially influence the average completion rate. Note, for example, that the class taught by Instructor 5, who determined 17% of the final course grade by Read & Practice activities, realized a 93% average completion rate. Instructor 11, who had Read & Practice account for 30% of the final course grade, had about the same completion rate. Then note, Instructor 6’s class wherein a student’s entire course grade was attributed to Read & Practice activities and there was only an 80% completion rate. These insights suggest there is something else influencing completion.

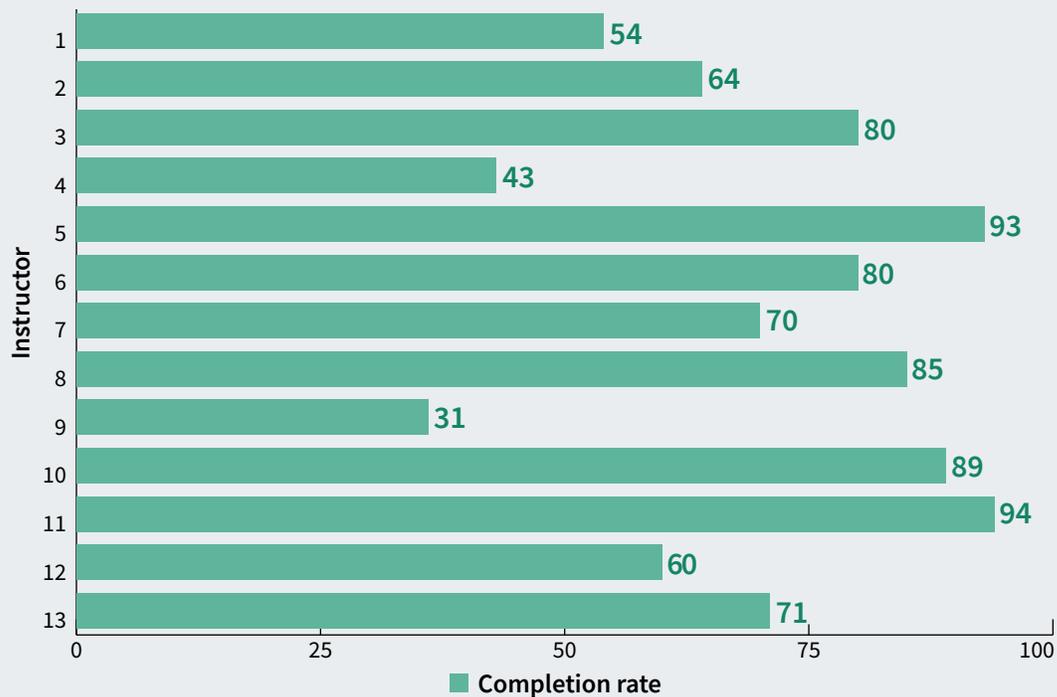


To further investigate whether there was a difference in rates of engagement and meeting the target score among subgroups of students, the data were disaggregated by baseline level of motivation to perform in this course and level of academic preparation. The sample was reduced to those students who had completed the baseline survey and data were disaggregated by those more motivated to succeed (n=339) and less motivated to succeed (n=151). There was no meaningful difference in the engagement rate between students more and less motivated to succeed (71% and 73%, respectively). Nor was there a meaningful difference in the average rate of meeting the target score by students more or less motivated to succeed (70% and 72%, respectively). This finding suggests that Read & Practice is as engaging among more and less motivated students.

The sample was then reduced only to students who had taken the ACT and/or SAT and reported their scores by section (n=306). Level of academic

preparedness was established by whether or not students had met the college readiness benchmarks set on the SAT and ACT. If they had met all section benchmarks they were categorized as “more academically prepared” (n=167) and if they had fallen below one or more readiness benchmarks they were categorized as “less academically prepared” (n=139). Students categorized as more academically prepared engaged in assigned activities at a significantly higher rate than those less academically prepared to succeed (78% and 65%, respectively). Students more academically prepared also met the target score at a higher rate than those less academically prepared (78% and 64%, respectively). This finding suggests that students less academically prepared need additional support to engage and persist. Qualitative data disaggregated by student’s level of motivation suggests that students less academically prepared had to spend more time in Read & Practice before they could meet the target score and this became frustrating, sometimes influencing disengaging from the activity.

Figure 2. Average rate of meeting the target score by instructor





RESEARCH QUESTION 2

What are instructor and student perceptions of Read & Practice? Do perceptions vary by educational context and/or implementation model?

We hypothesized that instructors would have a strong, positive perception of Read & Practice because it was built on core learning science principles and because features built into the tool were expected to increase instructor efficiency in their course. Features were also expected to support positive student behaviors, like helping students self-regulate, come to class more prepared to participate, and actively participate in class. It was further hypothesized that students would have strong positive perceptions of Read & Practice because they would find it easy to use, they would observe that it helped them come to class prepared to participate and that using it had positively influenced their academic performance in the course.

INSTRUCTOR PERCEPTIONS. Instructors in this sample had positive perceptions of Read & Practice. When asked to rate, on a scale of 0-10, how likely they were to recommend Read & Practice to a colleague, the average rating was 7.08. Ratings ranged from two to ten. Two outlier ratings of 2 and 3 did pull the average down. When asked to explain why they provide the rating that they did, those two participants noted, “I like the book and adaptive learning quizzes, but there are still a number of glitches that need to be resolved” and “I don’t teach everything that’s in the book, and I include material that’s not in the book, so it would be helpful to have more control over what questions are on the quizzes. I could see being able to add questions (maybe with a difficulty value to help with the adaptive aspect) as a useful feature. Being able to remove questions would be nice too.”, respectively. The other 11 instructors’ ratings ranged from six to ten and the comments left with the ratings were coded as “easy to use”, “track student reading”, and “students come to class prepared”.

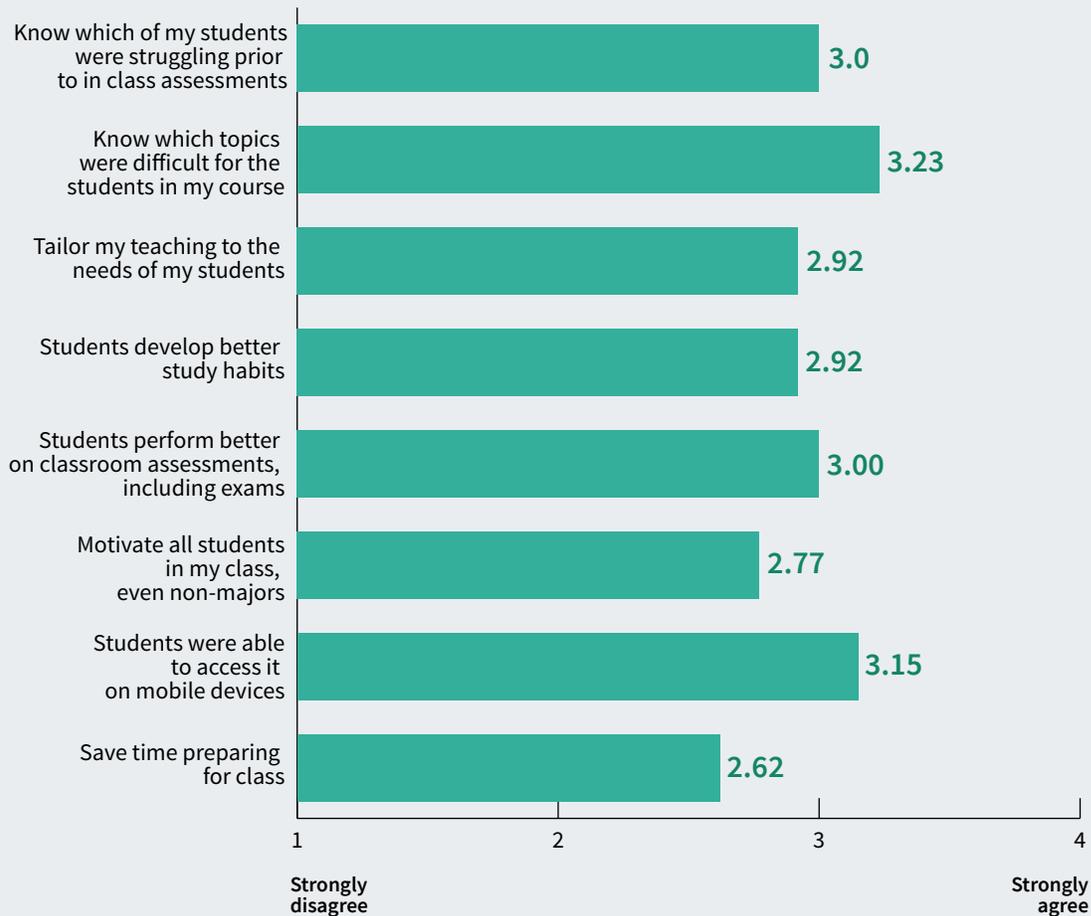
USABILITY. Instructors reported that Read & Practice was easy to use, giving the tool an average System Usability Scale (SUS) rating of 70 (considered above-average) and rating the tool a 3.3 on a single ease question with response options ranging from 1 = “very difficult” to 4 = “very easy”. One instructor noted, “It was easy to set up and use the students had no complaints regarding its functionality.”

When the data were disaggregated by how an instructor used Read & Practice, the highest average rating (7.9) came from the group of instructors who used assigned Read & Practice activities as pre-class only. Instructors who assigned quizzes as post-class activities only and those who assigned a combination of both pre- and post- class had similar average ratings (6.6 and 6.7, respectively). There was a significant, positive correlation between an instructor’s rating of likelihood to recommend the tool and the rating they gave Read & Practice on the SUS, suggesting that ease of use influences likelihood to recommend. There were no meaningful differences in likelihood to recommend based on an instructor’s background, experience, title, or the type of institution where they taught.

CLASSROOM MANAGEMENT AND STUDENT BEHAVIORS. Instructors tended to agree that Read & Practice supported classroom management and student behaviors. Instructors rated their level of agreement on a scale of 1 = “strongly disagree”, 2 = “disagree”, 3 = “agree”, and 4 = “strongly agree” with eight statements about Read & Practice. Instructors most strongly (3.23) rated “Achieve Read & Practice helped me know which topics were difficult for the students in my course” and gave the weakest average rating (2.62) to “Achieve Read & Practice helped me save time preparing for class”. Average responses to all items can be found in Figure 3.



Figure 3. Instructor perception of Read & Practice supporting classroom management and student behaviors



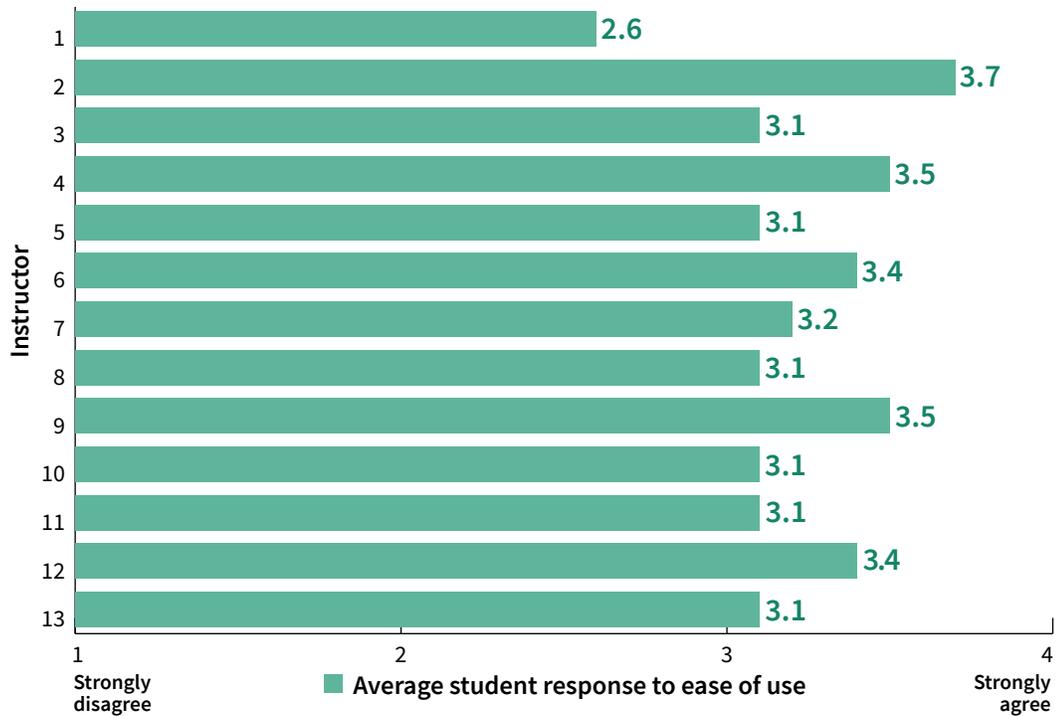
Instructors had strong, positive perceptions that Read & Practice supported student preparation for class. They were asked to rate the extent to which they agreed that Read & Practice supported their students completing assignments that were due and coming to class prepared to participate on a scale of 1 = “strongly disagree” through 4 = “strongly agree”. Overall, instructors rated these two items an average of 3.50 and 3.41. When considered by use case, instructors who used Read & Practice pre-class rated both items more highly on average than those who assigned the tool post-class and those who used it a combination of pre- and post-class (completed assignments = 3.8, 3.0, 3.5, came to class prepared = 3.5, 3.0, 3.4, respectively). Instructors also had positive perceptions of Achieve in supporting their student’s taking ownership over their own learning. Of all instructors who reported at least one benefit to Achieve, 65% commented on

student ownership. For example, “The best part of Read & Practice is student agency, taking my hands off the process.”

STUDENT PERCEPTIONS. Students in this sample had strong, positive perceptions of Read & Practice. Students were asked to rate, on a scale of 0 to 10 how likely they would be to recommend a course to a friend if they knew that Read & Practice was going to be used in that class. The overall average rating was 8.2. When examined by the usage model in their course, students whose instructors used Read & Practice pre-class had a higher rating, on average, than those whose instructors used Read & Practice post-class and those whose instructors used a combination of pre- and post-class (8.5, 8.0, 8.1, respectively). Students also completed a System Usability Scale and the score, on average, was 69.



Figure 4. Average student rating of single ease question by instructor



EASE OF USE. Students were asked to rate, on a scale of 1-4, their agreement with the statement “Achieve Read & Practice is easy to use”. The average rating on the single ease question was 3.3. Ease of use ratings by instructor can be found in Figure 4. Students also tended to comment about ease of use in the open response item that asked, is there anything else you would like to tell us about your use of Read & Practice this semester? For example, “it was very easy to use. For example it was very easy to click on a section of the chapter and read solely that section. It made it easier to go over course material”.

STUDENT BEHAVIORS. Students were also asked to rate the extent to which they agreed that Read & Practice supported them completing assignments that were due and coming to class prepared to participate on a scale of 1 = “strongly disagree” to 4 = “strongly agree”. Average students ratings were very similar to average instructor ratings for both items (3.51 and 3.42, respectively). When considered by use case, students in courses where their instructor

used Read & Practice pre-class rated both items more highly, on average, than those who assigned the tool post-class and those who used it a combination of pre- and post-class (completed assignments = 3.71, 3.42, 3.51, came to class prepared = 3.68, 3.05, 3.58, respectively). Average responses by instructor can be found in Table 5.

STUDENT ENGAGEMENT. To measure students’ perceived levels of engagement in the course, they were asked to report on a scale of one through five (1 = “a lot less engaged”, 2 = “less engaged”, 3 = “about the same level of engagement”, 4 = “more engaged”, 5 = “much more engaged”) their level of engagement in this course as compared to other courses they were taking that semester. The average rating was 3.4. Students in courses where Achieve was used pre-class had a higher average rating (3.55) as compared to students in courses where Achieve was used post-class (3.39) or a combination of pre- and post-class (3.42). Average rating by instructor can be found in Figure 6.



Figure 5. Rating of perception of students coming to class prepared and having completed assignments, by instructor

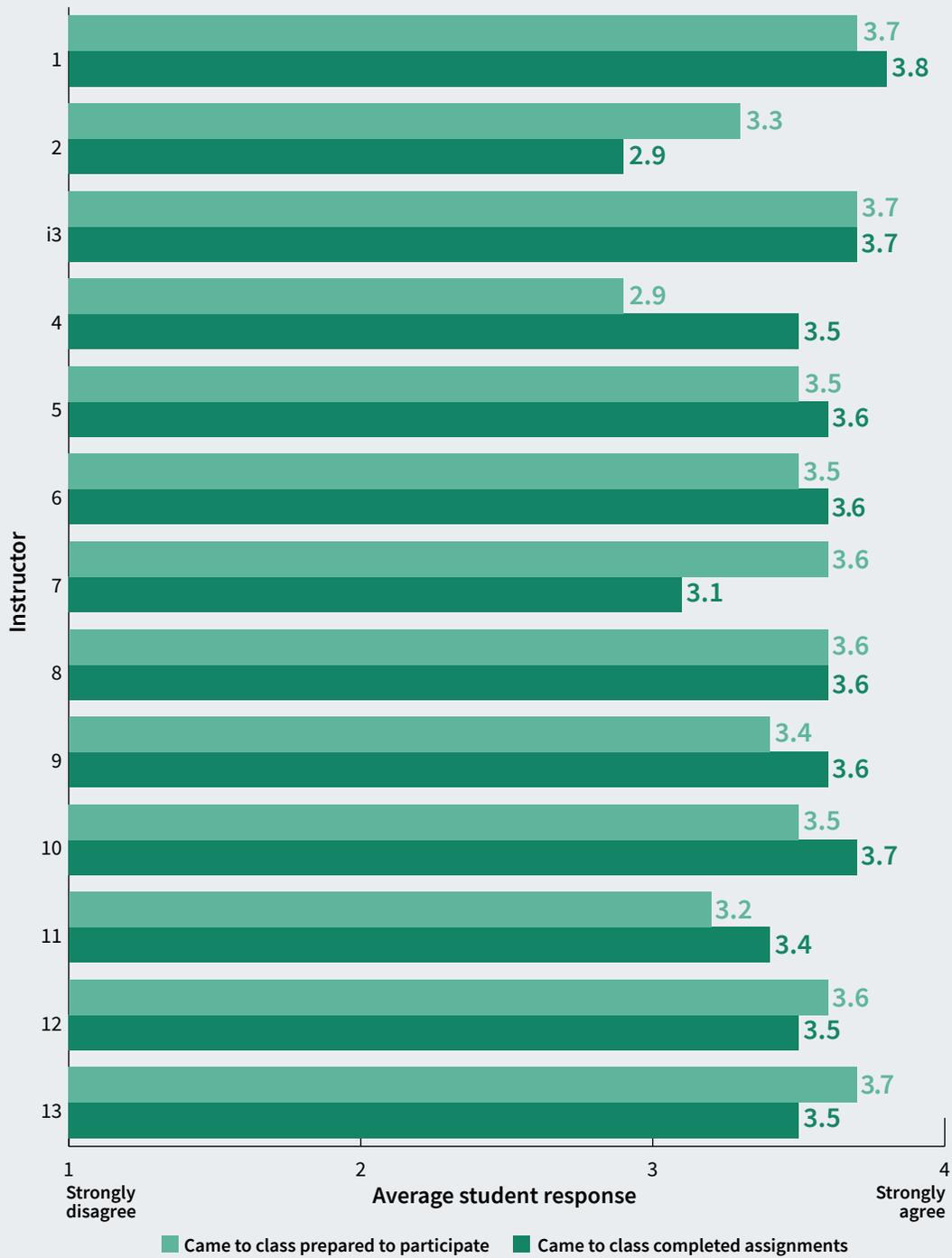
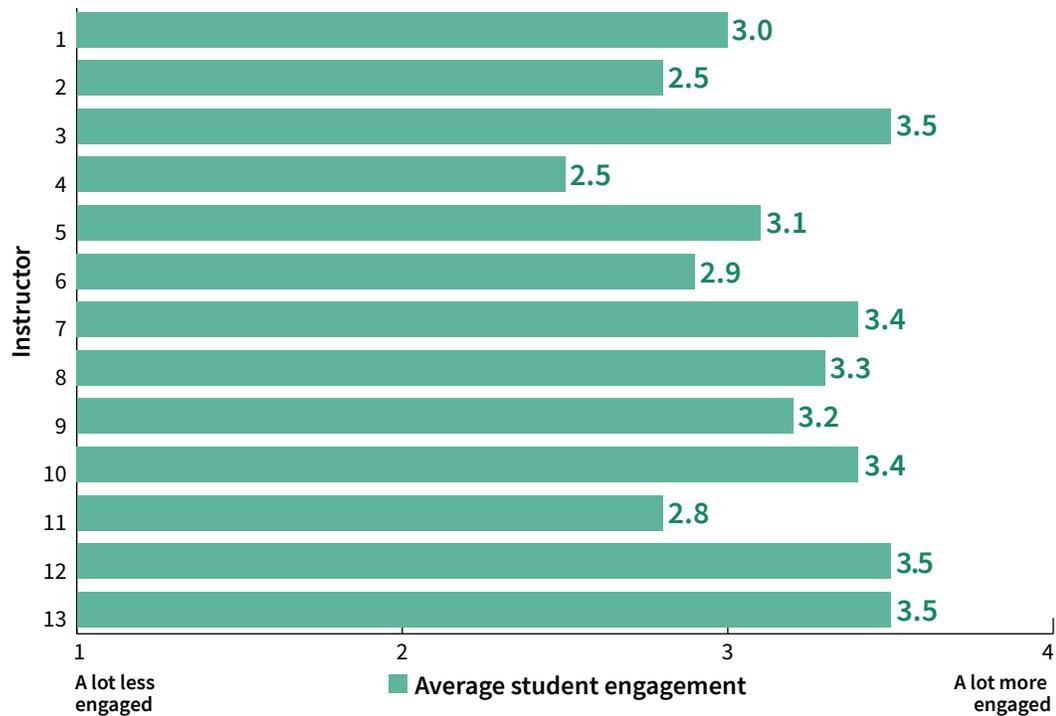




Figure 6. Student-perceived level of engagement in the course in which Read & Practice was being used as compared to other courses they are taking that semester



MOST BENEFICIAL. Students were asked to respond to one open response question that asked, what, if anything, was the most beneficial part of using Read & Practice. Many students in the sample (43%) reported that the most beneficial part of using Read & Practice was that it helped them stay on track with the reading, for example, “Read & Practice helped me to stay on track with the reading and helped me to really understand the content”. About one-third of the student respondents (34%) reported the most beneficial part of using Read & Practice was that it helped them prepare for in-class summative assessments like midterms and finals, for example, “I liked being able to take a mini quiz to help prepare for the major test for the week”. And, 17% reported the most beneficial part being using it to prepare for in-class discussions, for example, “Read & Practice really prepared me for class”. In total, 3% were coded as some “other” benefit (e.g. not having to carry a textbook, it being free, etc) and 3% were coded as “there was no benefit”.

MOST CHALLENGING. We were also interested in understanding how Read & Practice could be

further optimized to better meet student needs and hopefully influence higher engagement and completion. When students were asked the open response question, “what was the most challenging part about using Read & Practice?”, the largest proportion (48%) had a response coded as “nothing” (e.g. N/A, none, nothing was challenging, it was all easy, etc). The next largest proportion (22%) commented on the difficulty level of the quizzes—either that the individual items were too challenging or that the instructor had set a target score that took a long time to reach, for example, “the questions got harder the better you did”. In total, 10% commented on usability challenges, particularly navigating from the quizzes to the eBook, for example, “it was difficult to figure out how to get back and forth to the quiz.” Eight percent commented on performance issues like slow load times, for example, “it takes a little bit to load chapters sometimes or a technical difficulties”, and five percent commented on getting themselves motivated to do the activities, for example, “the most difficult part was getting myself to log in every day, even when I was tired”.



RESEARCH QUESTION 3

What is the relationship between use of Read & Practice and performance on in-class exams and does student motivation, prior academic performance or instructor moderate that relationship?

We hypothesized that more use of Read & Practice would positively influence assessment scores. Based on previous research, we suggested that the following factors would also contribute to this relationship: a student's level of academic preparedness when entering college, their level of motivation to succeed at the beginning of the semester, and which instructor was teaching the course.

Although final course grade or final exam score might seem like the more obvious outcomes to measure academic performance in the course, the outcome of average exam score was selected for two reasons: (1) because of the proportion of a student's final course grade that engagement in Read & Practice accounted for, the relationship would introduce multicollinearity, so it was not appropriate to use it as an outcome, and (2) the majority of instructors in this sample did not give a final exam; rather they gave a set of exams throughout the course of the semester. We determined that the average of the scores a student received on their exams over the course of the semester was the most appropriate outcome to measure the relationship between engagement in Read & Practice and academic performance. However, this restricted our sample because we did not include the four instructors who only gave a final exam. The final analytics sample for this research question included nine instructors and 488 students. The descriptive differences between the overall sample of students and the sample of students used in this analysis is presented in Table 4.

The main effect (whether rate of engagement influenced average exam scores) was tested by first examining if the relation was moderated by a student's level of academic preparedness when entering college, their baseline level of motivation to succeed, and which instructor was teaching the course. We wanted to investigate, for example, whether the relation of engagement and exam performance is stronger for students who are more academically prepared as compared to those who were less academically prepared. If the relation is not moderated by academic preparedness, level of motivation, or instructor, than these variables will be used as controls to examine the main effects.

VARIABLES EXAMINED IN THIS ANALYSIS

STUDENTS' AVERAGE RATE OF ENGAGEMENT IN READ & PRACTICE.

All instructors assigned Read & Practice activities for credit. If a student had launched an activity, he or she was recorded as having "engaged" in that activity. An average engagement rate per student was calculated by summing all the activities a student had engaged in and dividing that number by all of the activities assigned. The overall engagement rate across students was 71.2% and ranged from 18% to 100% among students in this sample.

STUDENTS' IN-CLASS EXAM AVERAGE. To operationalize each student's academic performance in the course, the scores on all exams were summed and divided by the number of assessments the instructor had given. The resulting calculation was the student's in-class exam average. The overall in-class exam average was 81.60 (n=469) and ranged from 34.6 to 106.5 among students in the sample (19 students had missing data on average exam score and were removed from the sample).

BASELINE LEVEL OF MOTIVATION. Students completed four items on the pre-survey that were all on a four-point scale, and their ratings were averaged across the four items. The resulting calculation was the student's baseline level of motivation. The average rating of baseline level of motivation was 3.03 (n=481). Motivation level was gathered to assess if the relation of engagement in Read & Practice and exam score was moderated by motivation. If the relation did not differ by motivation level (i.e., homogeneity of the regression slopes), we could examine the relation between engagement in Read & Practice and exam scores when holding the motivation level constant (i.e., relation between engagement and exam as if all students were at the mean of motivation).



HIGH SCHOOL GRADE POINT AVERAGE (HSGPA).

Students self-reported their high school grade point average on the pre-survey and again on the post-survey. The average HSGPA was 3.38 and ranged from 1.00 to 5.00 among students in this sample. HSGPA was gathered to assess if the relation of engagement in Read & Practice and exam score was moderated by academic preparedness. If the relation did not differ by HSGPA level (i.e., homogeneity of the regression slopes), we could examine the relation between engagement in Read & Practice and exam when holding HSGPA constant (i.e., relation between engagement and exam as if all students were at the mean of HSGPA).

MODEL BUILDING

We first examined the interaction effects. Before testing the interactions, HSGPA, motivation, and engagement were centered to reduce multicollinearity between the predictors, and instructors were dummy coded (Aiken & West, 1991).

Multiple regression analyses were conducted to evaluate if the relation between engagement (predictor) and their average in-class exam score (outcome) was moderated by their academic preparedness coming into college (as measured by HSGPA). When testing the main effects, engagement had a significant positive relationship in general ($F_{(385)} = 17.05, p < .001$) with an R_2 of 0.0814. However, it was hypothesized that this relationship may change as a function of academic preparedness so the interaction of HSGPA and proportion of activities engaged in was predicted to be statistically significant. A second model was built including the interaction term. The interaction term (between HSGPA and engagement) was positive (5.54) but was not statistically or practically significant ($t_{384} = 1.29, p = 0.1944$) indicating that there is no meaningful interaction between HSGPA and engagement when predicting exam performance.

Table 4. Descriptive comparison between overall sample and research question three analytic sample by percentage of students

	All (n=658)	Research question three analytic sample (n=488)
Year in school		
Dual Enrolled	6.22	6.42
First	52.44	53.74
Second	26.22	25.67
Third	9.33	8.56
Fourth	3.33	3.21
Fifth	<1	<1
Other	1.78	1.87
Eligible for federal financial aid (yes)	67.41	69.60
First generation (yes)	30.60	29.33
Majoring in psychology	14.73	15.01
Took psychology in high school	29.46	30.56
HSGPA	3.35	3.89
Less motivated to succeed	30.28	30.55
Less academically prepared	49.19	47.51
Average exam scores	80.12	81.60
Average course grade	83.50	85.18



When testing the interaction of baseline level of motivation and rate of engagement, engagement and motivation were centered. The main effects were tested first ($F(338) = 13.93, p < .001$) before including the interaction term in the second model. The interaction was positive (1.55) but was not statistically or practically significant ($t_{337} = 0.21, p = 0.8331$) indicating that there is no meaningful interaction between baseline level of motivation and engagement.

To test the interaction between engagement and instructor, instructors were dummy coded and the main effects were tested. The model was significant ($F(339) = 15.09, p < .001$). When the interactions were included in the model, two of the interaction terms were positive and seven were negative. But, none of the interactions were statistically or practically significant.

Results of these models showed that the relationship between engagement in Read & Practice and exam scores was not moderated by prior academic performance, baseline level of motivation, or who the instructor was, so it was appropriate to control for them when testing the main effects.

A multiple linear regression was calculated to predict average exam score based on proportion of engagement controlling for HSGPA, baseline level of motivation, and instructor. A significant regression equation emerged ($F(382) = 17.77, p < 0.0001$). When considered individually the proportion of activities that a student engaged in was significant ($p < 0.0001$), as was HSGPA ($p < 0.0001$), as were each of the dummy coded instructor variables. Baseline level of motivation to succeed was not significant ($p = 0.1001$). The overall model R^2 was 0.3940 and the overall model adjusted R^2 is 0.3718, suggesting that the model including baseline level of motivation to succeed, HSGPA, who the instructor was, and rate of engagement accounted for 37% of the variance in average in-class exam scores.

Squared semi-partial correlations were examined to understand incremental contribution to the prediction. HSGPA accounted for 4.7% of the variance, baseline level of motivation accounted for <1% of the variance, instructor contribution ranged from <1% to 8.3%, and engagement accounted for 7.4% of the variance in exam scores.

To further visualize this relationship and for ease of interpretation, we graphed average exam score by rate of engagement with Read and Practice among students who fell below the HSGPA average and those who met or exceeded it. Figure 7 suggests that if a student falls below a HSGPA of 3.38 (the mean in this sample), and they engaged in at least 80% of assigned Achieve Read & Practice activities, they could expect more than a seven-percentage point increase in their average in-class exam scores—representing more than half of a letter grade in most higher education institutions. Students who graduated high school with at least a 3.38 grade point average and engaged in at least 80% of their assigned Read & Practice activities could expect about an eight-percentage point increase in their in-class exam score average.

The same analysis was conducted within baseline level of motivation. Figure 8 suggests that if a student fell below a baseline motivation rating of 3.03 (the mean in this sample) and they engaged in at least 80% of assigned Achieve Read & Practice activities, they could expect a nine-percentage point increase in their average in-class exam scores—representing nearly a full letter grade in most higher education institutions. Students with an average baseline motivation rating of at least a 3.03 who engaged in at least 80% of their assigned Read & Practice activities could expect nearly the same difference in performance at 80% engagement.



Figure 7. Relationship between the rate of engagement in assigned Read & Practice activities and average exam scores by HSGPA.

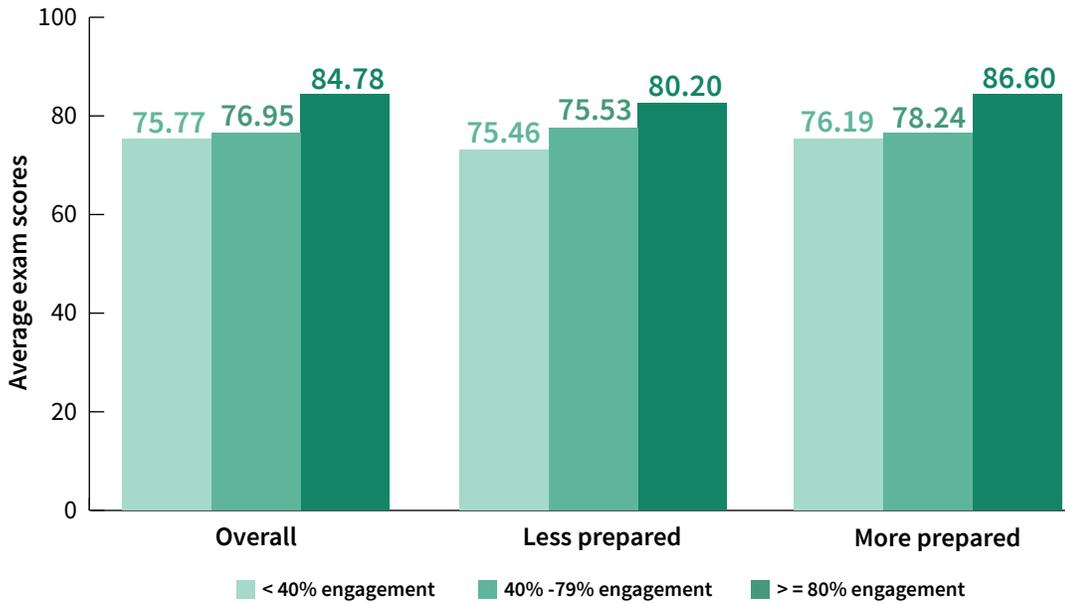


Figure 8. Relationship between the rate of engagement in assigned Read & Practice activities and average exam scores by baseline level of motivation.





RESEARCH QUESTION 4

Is voluntarily retaking quizzes in Read & Practice related to higher in-class exams, does student motivation, prior academic performance or who the instructor was moderate that relationship?

We hypothesized that some students would voluntarily retake quizzes for practice and that voluntarily re-taking quizzes would be positively related to a student's average in-class exam score based on the retrieval research around the testing effect. However, we suggested that a student's level of academic preparedness and baseline level of motivation might have a moderating effect on the relationship.

In total, 540 students had valid data to be included in this sample. Of those students, 231 (43%) were coded as having retaken at least one assigned quiz and met the target score and 309 (57%) students never retook an assigned quiz.

Because we hypothesized the academic preparedness and motivation would moderate the relationship between group status and exam performance, we first examined the interaction effects of HSGPA and baseline level of motivation.

The interaction term of HSGPA and retaker status was negative (-3.89) but was not statistically or practically significant ($t_{401} = -1.30$, $p=0.1959$) indicating that there is no meaningful interaction between HSGPA and retaker status when predicting exam performance. When testing the interaction between retaker status and baseline level of motivation, the model was not statistically significant ($F(350) = 1.39$, $p=0.2458$), indicating that there is no meaningful interaction between baseline level of motivation and retaker status when predicting exam performance.

Based on the results of the interaction analyses it was appropriate to include academic preparedness and baseline level of motivation as covariates in a multiple linear regression examining the relationship between being a

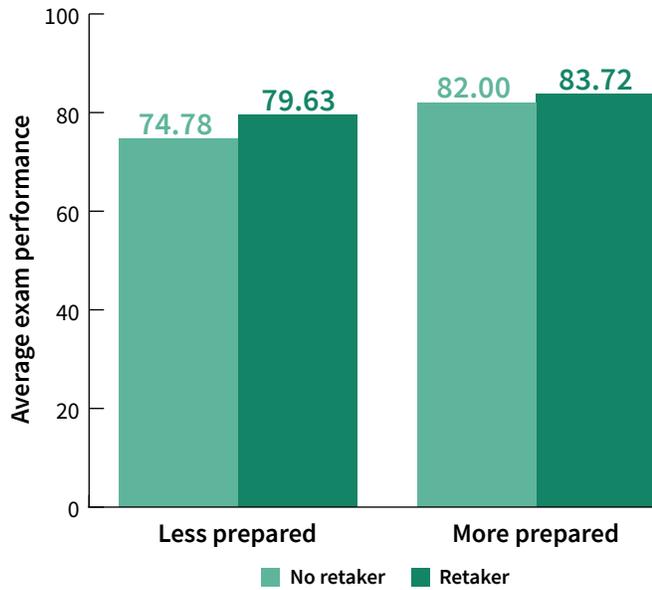
retaker and average exam scores. A significant regression equation was found ($F(350) = 3.53$, $p=0.0151$), with an R^2 of 0.029 and an adjusted R^2 of 0.021. When examined independently, retaker status and HSGPA were significant and baseline level of motivation to succeed was not, however, we decided to leave it in the model but there is strong theory that motivation would contribute to exam average. These results indicate that whether a student is a retaker, their HSGPA account for only about 2% of the variance in exam scores.

After statistically controlling for student prior academic performance and baseline level of motivation we wanted to illustrate the differences observed in this sample for ease of interpretation. Figure 9 illustrates that students who are less academically prepared to succeed and retake quizzes for practice realize a near half letter grade improvement in their average in-class exam scores, and reduce the gap with their more academically prepared peers at a rate of five percentage points. Put another way, if they didn't retake quizzes there would be a gap between more and less academically prepared students of nine percentage points, nearly a full letter grade. By retaking quizzes for practice less academically prepared students have closed that gap to less than half a letter grade.

Figure 10 illustrates the difference in final exam scores for less and more motivated students, by whether they retook quizzes for practice. Note that there is a similar effect realized by both less and more motivated students when they retake quizzes for practice. That is, both cohorts of students - those less and more motivated to succeed realize benefits when retaking quizzes for practice.

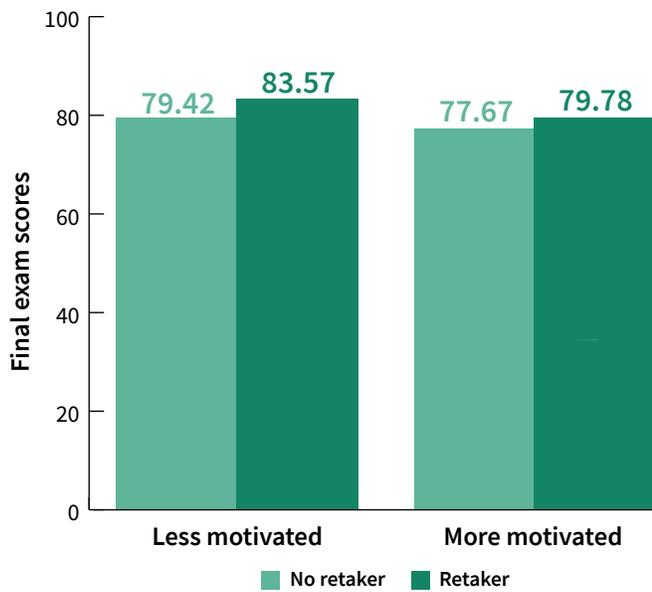


Figure 9. Average exam performance by retaker status and academic preparedness



Note:
nLess Prepared, no retaker = 161 nMore Prepared, no retaker = 148
nLess Prepared, retaker = 142 nMore Prepared, retaker = 89

Figure 10. Average exam performance by retaker status and baseline level of motivation



Note:
nLess Motivated, no retaker = 84 nMore Motivated, no retaker = 225
nLess Motivated, retaker = 82 nMore Motivated, retaker = 149



Discussion

Implementation, usage, and instructor and student perceptions of Read & Practice had been systematically investigated by the authors throughout the development lifecycle. Early evaluation findings influenced development and optimization, and when Read & Practice was released to market the authors began to measure effectiveness by partnering with a representative sample of instructors and their students.

Like previous studies of Read & Practice the results of this study conducted with a more representative sample of instructors suggested that educators have strong positive perceptions of the tool. They perceive Read & Practice to be easy to use and report that it supports students keeping up with assigned readings and coming to class prepared to participate. Instructors also value the insights that Read & Practice provides, enabling efficient pedagogical decisions and identifying students who may benefit from intervention. Student perceptions were similar, and many students articulated that Read & Practice helped them perform better on in-class assessments.

Because Read & Practice is a new tool, this is the first study to systematically examine how instructors from a variety of educational contexts and backgrounds chose to use it. We were surprised that only five of the 13 instructors assigned Read & Practice before the course in which the content would be taught given the research on the effectiveness of pre-class activities. However, interestingly, the same proportion of instructors assigned Read & Practice pre- and post-class, as this was not an implementation pattern that emerged when we were evaluating the tool in beta. And, even more interesting was that the students who were



in the “combination” implementation group realized the highest average in-class exam scores among the three implementation groups, even when holding rate of engagement with the tool constant. These findings suggest that assigning Read & Practice pre-class was more effective than assigning it only post-class, but that a model of offering students an introduction to the material, discussing the material in-class, and then reinforcing it with another Read & Practice activity was most effective to positively influence in-class exams. This insight expands the current literature, which primarily examines the impact of pre-class only.

All instructors assigned Read & Practice activities for credit nearly every week of the semester and students engaged in and completed assigned Read & Practice activities at a relatively high rate. These findings, paired with qualitative data from students suggests that they find the tool valuable and perceive that it supports their level of class preparedness and their performance on in-class assessments.

Like Carpenter et al. (2019) suggested, we were interested in further investigating whether the relationship that emerged between retrieval practice and performance on in-class exams was moderated by student prior academic performance. The interaction was not statistically significant so we controlled for HSGPA to test the main effects. Findings suggested that students who were both more and less academically prepared to succeed benefited from using Read & Practice. We wanted to further expand the literature by understanding whether motivation was a moderating factor. Again, findings suggested that it did not modify the relationship and that when controlled, the model was statistically significant. We suspect that students who were classified as less motivated to succeed engaged in Read & Practice at a high rate because of their strong, positive perceptions of its ease of use and perceived value.

Although none of the instructors in our sample assigned retaking quizzes for practice like we had hypothesized that some would, we were nevertheless able to expand the current literature by examining students who voluntarily retook quizzes for practice. The findings suggest that retaking quizzes positively influenced assessment performance regardless of prior academic performance or baseline level of motivation to succeed. Qualitative data suggested that students realized early in the semester that Read & Practice supported exam preparation, and thus 43% voluntarily retook at least one quiz to help prepare for exams.

This study is the third in a series of studies conducted on Achieve Read & Practice as it was being developed and optimized for use at scale. The first study, conducted with a small, controlled group of instructors and students built early, directional evidence that instructors and students valued Achieve Read & Practice and found it to be of high quality. The primary insights that emerged from that study were areas for tool optimization. Results from the second study, conducted with the same small, controlled set of instructors and students across a full semester suggested that there may be a relationship between use of Achieve Read & Practice and student learning outcomes among instructors who are comfortable with technology and tolerant of product gaps expected in a beta. The results from this study validated those findings with a larger sample more representative of the target population of users.

Taken together, the findings from this study provide evidence that Read & Practice is engaging among various cohorts of students (i.e less and more motivated to succeed and less and more academically prepared to succeed). The results also provide evidence that use of Read & Practice supports academic performance among subgroups of interest. Thus, there is evidence to suggest that use of Read & Practice can help bridge the gap between more and less academically prepared students and that this finding would persist at both two- and four-year institutions.



Implications for Instruction

Two key implications for instruction emerged from this study.

IMPLEMENTATION PATTERNS. Achieve Read & Practice is a flexible solution that can be used in the way that best compliments an instructor’s pedagogy in their unique educational context. However, results from this study suggest that when Read & Practice was used pre-lecture, students had a more positive perception of the tool, they perceived themselves to be more prepared to participate in class than students in courses when it was used pre- and post-lecture, and the strongest relationship between use and student outcomes emerged when considered among use cases. Together, these findings suggest that instructors and students might realize the strongest outcomes when Achieve Read & Practice is used pre-lecture.

PROMOTING RE-TAKING. The findings from this study suggest that retaking quizzes may positive influence student learning outcomes among students more and less motivated to succeed, more and less academically prepared when entering college, and among students of all instructors. These findings, paired with the qualitative data from students reporting that re-taking quizzes for practice helped them prepare for in-class exams (21% of student open response items asking what, if anything, students liked most about Achieve Read & Practice, were coded as “re-taking quizzes helped me prepare for in-class assessments) suggest that instructors might consider positively influencing student’s retake of quizzes—perhaps by offering extra credit or re-assigning a previously submitted activity in preparation for an in-class assessment.



Limitations and Future Research

Like most applied research in educational settings there are limitations to this work that are important to document. The research presented in this report is correlational and thus causal inferences cannot be made based on the results. Nevertheless, the results are significant, as they validate early findings that emerged from earlier studies with a representative sample of instructors and students; and they contribute to a growing body of efficacy evidence of a new digital learning tool.

While there were important controls used in this study at the instructor and student levels, we cannot statistically account for all differences among 13 courses at 13 different institutions. Therefore it is likely that there are factors contributing to the relationships that we did not observe in this study.

Lastly, while we had a sample that covered various types of instructors and students, the results cannot be generalized to the entire population of users.

To further contribute to the evidence of the efficacy of Achieve Read & Practice, a quasi-experimental study is being conducted at the time of this report publication. In the Fall 2019 semester, 13 instructors and their students are participating in quasi-experimental study of the impact of Achieve Read & Practice. In that study they are using Read & Practice in one section of their Introductory Psychology course and they are not using it in another section of their Psychology course. The results of this study will allow the isolation of the impact of Read & Practice and enable more generalizable claims.



Note on Data Privacy

Prior to data collection, this study and the associated consent forms and instruments were reviewed and approved (found exempt) by the Human Resources Research Organization (HumRRO). HumRRO is a third-party Institutional Review Board organization with no affiliation with Macmillan Learning (federal wide assurance number 00009492 and IRB number 00000257). Macmillan Learning seeks independent and unfunded third-party review to eliminate any bias in decision of exemption. Macmillan Learning then seeks local Institutional Review Board approval at each participating institution, where required. The data collected in this study, which are provided by the instructor and consenting students, are initially identifiable. However, once a random identifier is generated identifiable data are destroyed. Data are provided in secure storage locations, and access is permitted only to the primary investigator in the study. For full details of our data handling and storage privacy procedures, contact Kara McWilliams, Vice President Impact Research at Macmillan Learning at kara.mcwilliams@macmillan.com.



REFERENCES

- Agarwal, P. K. (2018). Retrieval Practice & Bloom's Taxonomy: Do Students Need Fact Knowledge Before Higher Order Learning? *Journal of Educational Psychology*. Advance Online Publication. https://www.researchgate.net/publication/325639446_Retrieval_Practice_Bloom's_Taxonomy_Do_Students_Need_Fact_Knowledge_Before_Higher_Order_Learning
- Barata, G., Gama, S., Jorge, J., & Gonçalves, D. (2013). Improving participation and learning with gamification. Proceedings of the First International Conference on gameful design, research, and applications, 10-17. Retrieved from https://www.researchgate.net/publication/259821680_Improving_Participation_and_Learning_with_Gamification
- Brooke, J. (1986) SUS—A Quick and Dirty Usability Scale. Usability Evaluation in Industry, 189-194. <https://hell.meiert.org/core/pdf/sus.pdf>
- Brown, P.C., Roediger, H.L. & McDaniel, M.A. (2014). *Make it Stick: The science of successful learning*. London: The Belknap Press of Harvard University Press. <https://www.tandfonline.com/doi/abs/10.1080/00220671.2015.1053373>
- Carpenter, S., Lund, T.J., Coffman, C.R., Armstrong, P.I., Lamm, M., & Reason, R.D. (2016). A Classroom Study on the Relationship Between Student Achievement and Retrieval-Enhanced Learning. *Educational Psychology Review*, 28, 353-375. <https://www.semanticscholar.org/paper/A-Classroom-Study-on-the-Relationship-Between-and-Carpenter-Lund/c9143c842b6c5184273084f026cecd3db0142ce4>
- Cassady J. C., Johnson R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, (2), 270-295. https://www.researchgate.net/publication/223427593_Cognitive_Test_Anxiety_and_Academic_Performance
- Endres, T., Carpenter, S., Martin, A., & Renkl, A. (2017). Enhancing learning by retrieval: Enriching free recall with elaborative prompting. <https://www.semanticscholar.org/paper/Enhancing-learning-by-retrieval%3A-Enriching-free-Endres-Carpenter/9c657004516b027c06ec428aee192adb90ad125>
- Freeman S., Haak D., Wenderoth M. P. (2011). Increased course structure improves performance in introductory biology. *CBE—Life Sciences Education*, (2), 175-186. <https://www.lifescied.org/doi/full/10.1187/cbe.10-08-0105>
- Grimaldi, P. J., & Karpicke, J. D. (2014). Guided retrieval practice of educational materials using automated scoring. *Journal of Educational Psychology*, 106(1), 58-68. <http://dx.doi.org/10.1037/a0033208>
- Gross D., Pietri E. S., Anderson G., Moyano-Camihort K., Graham M. J. (2015). Increased pre-class preparation underlies student outcome improvement in the flipped classroom. *CBE—Life Sciences Education*, (4), ar36. <https://www.ncbi.nlm.nih.gov/pubmed/26396151>
- Handelsman, M. M., Briggs, W. L., Sullivan, N., & Towler, A. (2005). A measure of college student course engagement. *The Journal of Educational Research*, 98 (3), 184-191. <http://dx.doi.org/10.3200/JOER.98.3.184-192>
- Hodges LC, Anderson EC, Carpenter TS, et al. Using reading quizzes in STEM classes – the what, why, and how. *J College Sci Teach*. 2015;45(1):49. http://dx.doi.org/10.2505/4/jcst15_045_01_49
- 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(2), 151-162. <http://dx.doi.org/10.1016/j.jml.2006.09.004>
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborate studying with concept mapping. *Science*, 331(6018), 772-775. <http://dx.doi.org/10.1126/science.1199327>
- Kuncel, N.R., Credé, M., & Thomas, L. L. (2005). The Validity of Self-Reported Grade Point Averages, Class Ranks, and Test Scores: A Meta-Analysis and Review of the Literature. *Review of Educational Research*, 75(1), 63-82. Retrieved from <https://doi.org/10.3102/00346543075001063>
- Pape-Lindstrom, P., Eddy, S., & Freeman, S. (2018). Reading Quizzes Improve Exam Scores for Community College Students. *CBE life sciences education*, 17(2), ar21. www.ncbi.nlm.nih.gov/pmc/articles/PMC5998316/
- 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. <http://dx.doi.org/10.1037/a0026252>
- Roediger H. L., III, Butler A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Science*, (1), 20-27. [10.1016/j.tics.2010.09.003 http://europepmc.org/abstract/MED/20951630](http://europepmc.org/abstract/MED/20951630)
- Roediger, H.L., & Karpicke, J.D. (2006). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on psychological science : a journal of the Association for Psychological Science*, 1 3, 181-210. <https://journals.sagepub.com/doi/abs/10.1111/j.1745-6916.2006.00012.x?journalCode=ppsa>
- Roediger, H.L., III, & Karpicke, J.D. (2006). Test enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249-255. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>
- Shaw, E. J., & Mattern, K. D. (2009). Examining the accuracy of self-reported high school grade point average. (College Board Research Report No. 2009-5). New York: The College Board. Retrieved from <https://eric.ed.gov/?id=ED562616>
- Toppino, T. C., & Cohen, M. S. (2009). The testing effect and the retention interval: Questions and answers. *Experimental Psychology*, 56(4), 252-257. <http://dx.doi.org/10.1027/1618-3169.56.4.252>

About the Authors

Dr. Kara McWilliams

Vice President Impact Research

Kara is passionate about researching the impact of digital technologies in higher education, and how insights can inform teaching and learning. She has ten years of experience conducting qualitative and quantitative investigations of how course and classroom interventions can improve learner outcomes and influence learning gains. She holds a doctorate in Educational Research, Measurement and Evaluation and a master's degree in Curriculum & Instruction from Boston College.

Alexandra Stone,

University of Connecticut

Alexandra is a graduate student at the University of Connecticut working on her PhD. in Measurement, Evaluation, and Assessment where she earned her M.A in the same program. At the time of this study Alexandra was an Associate Impact Research Scientist in the Learning Science and Insights team at Macmillan Learning, she continues to contract for the Learning Science team. Alexandra has broad experience in using qualitative and quantitative methods to examine the effectiveness of instructional programs on student achievement. She also has experience in developing and managing large databases using SQL and performing data analyses in SAS and R. Alexandra received a B.A. in Statistics from Elon University.

About Macmillan Learning

Macmillan Learning improves lives through learning. Our legacy of excellence in education continues to inform our approach to developing world-class content with pioneering, interactive tools. Through deep partnership with the world's best researchers, educators, administrators, and developers, we facilitate teaching and learning opportunities that spark student engagement and improve outcomes. We provide educators with tailored solutions designed to inspire curiosity and measure progress. Our commitment to teaching and discovery upholds our mission to improve lives through learning. To learn more, please visit <http://www.macmillanlearning.com> or see us on Facebook, Twitter, LinkedIn or join our Macmillan Community.

About the Learning Science and Insights Team

As the Learning Insights company, we are passionate and scientific about helping students, instructors, and institutions to achieve their full potential. We use a unique combination of user-centered design, research from the learning sciences, and empirical insights from extensive data mining and Impact Research. To learn more about this approach, please visit <http://www.macmillanlearning.com/catalog/page/learningscience>