ACHIEVE MORE

The learning engineering of Achieve and insights into instructor implementations and instructor and student outcomes

Kara McWilliams PhD, Vice President Impact Research, Macmillan Learning

Jeff Bergin PhD, Vice President, Learning Research

Adam Black PhD, Chief Strategy and Learning Officer

CONTRIBUTING AUTHORS | Marcy Baughman, Senior Director of Impact Research
Becca Runyon, Director, Assessment and Reporting

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.

macmillan learning

Contents

Foreword

At Macmillan Learning we are committed to developing learning solutions that help institutions, instructors, and all students to achieve their full potential. We go about this by co-designing with students, collaborating with leading educators and learning scientists, deriving cutting-edge insights from responsible data mining, and partnering with colleges and instructors to research impact and share insights for success. For impact research, we strive to provide instructors and students with practical, actionable, and timely insights derived from studies that meet the highest standards for educational and psychological testing. Our goal is to help advance teaching and learning by enabling evidence-based decision making and to contribute to research into educational technology. To these ends, we take a comprehensive approach to measuring the effectiveness and efficacy of the digital learning tools that we produce. Starting 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 comprehensive and rigorous evidence to understanding the efficacy of that tool. Studies also codify usage, engagement, and outcome patterns in differing educational contexts to provide instructors with insights into how they may choose to implement the tool in their own courses. This report represents one study that makes up the larger body of efficacy research into Achieve. We are confident in the results presented here but acknowledge that measuring efficacy is complex, and that we are always learning. The authors of this report, and the learning science 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 presented 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 educational tools we are developing meet the needs of instructors and students, and contribute positively to driving stronger outcomes in 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. And, 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 Learning 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 Impact Research team who worked tirelessly to establish rigorous study designs, collect valid and reliable data, and review drafts of this report. Marcy Baughman, Billie-Jo Grant, Larisa Nachman, and deKoven Pelton, your work is exemplary.

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:

Alvin Community College, Austin Community College - Cypress, Bethel University, Bunker Hill Community College, Butler County Community College. California Polytechnical Institute - San Luis Obispo, Cerritos College, Delgado Community College, Drexel University, Florida State College at Jacksonville, Georgia State University - Georgia Perimeter College, Gettysburg College, Hillsborough Community College, Howard College, Northern Arizona University, Owens Community College, Pellissippi State Community College, Pomona College, Richland College, Rockhurst University, Rutgers University Busch Campus, San Francisco State University, San Jose State University, Santa Clara University, Santa Monica College, Somerset Community College, Spelman College, State University of New York College at Oneonta, University of Arizona, and the University of North Florida.

Executive summary

Advances in the learning sciences combined with the fast evolution of powerful digital technologies and advances in the design of user experience have the potential to transform the future of higher education. There is now a large body of research demonstrating that learning science principles such as self-regulation, formative assessment, and active learning support the development of lifelong learners (Bell & Kozlowski, 2008). Digital tools can increasingly offer content, resources, and assessments that are personalized, adaptive, and relevant. (Cook et. al, 2013). And, seamless, efficient, engaging digital experiences that make the lives of students and instructors easier can meet the needs of busy educators and a diverse student population (Herrington & Oliver, 2000). However, despite a host of digital learning tools available to instructors there is a lack of rigorous and relevant evidence researching effectiveness and this has often led to false starts and frustrations of what to use and how to use it to best effect to improve student success.

When building Achieve, a new digital learning platform, Macmillan Learning began by founding the solution on research-based learning science principles, and co-designing the platform with instructors and students. Researchers at Macmillan Learning then took the unusual approach of beginning to investigate the effectiveness and efficacy of Achieve in its infancy by conducting rapid-cycle evaluations of tool features as they were developed and implementing increasingly rigorous validity and efficacy studies as the platform matured through beta testing.

ACHIEVE

Achieve is a digital learning solution developed for higher-education courses (at the time of this report publication Achieve was being studied in five disciplines: Biology, Calculus, Chemistry, Composition, and Economics). Achieve provides a connected suite of course content and tools designed to give instructors choice, with flexible recommendations for optimal pedagogical structures based on the learning sciences.

ACHIEVE WAS CONCEIVED BASED ON SIX LEARNING DESIGN PRINCIPLES.

1. Develop Learner Motivation. When students are highly motivated, they are able to tackle challenging problems and strive to accomplish goals that will improve their abilities.

2. Provide Personalized and Adaptive Experiences. Students enter the classroom with a variety of cultures and psychological traits, thus, personalization and adaptation of instruction and assessment can have positive effects for all learners.

3. Target Cognitive and Memory Elements. Today, there are numerous methods that learning scientists have researched to enhance learner cognition and transfer. These begin with learning objectives, which describe “the intended change in knowledge” and can enable a mastery approach which has positive impacts on conceptual learning, attitudes toward learning, and performance.

4. Build on Well-Constructed Learning Models. Being cognitively engaged stimulates learning, specifically, learning that “sticks”. Active learning, which can be fostered through models including Project-Based Learning (PjBL) and Problem-Based Learning (PBL), lead to the growth of complex-reasoning skills, critical-thinking processes, perceived learning, engagement, attitudes towards and perceived usefulness of subjects, self-directed learning, exam performance, motivation, and autonomy.

5. Create Interactive and Constructive Opportunities. The development of critical-thinking skills and higher-order learning benefit from collaborative learning, which lead to enhancement in academic performance and intellectual development.

6. Enable Metacognition and Self-Regulation. Metacognition and self-regulation are critical for academic success.

ACHIEVE HAS BEEN CHOREOGRAPHED AROUND TEACHING PRACTICES THAT PROMOTE ACTIVE LEARNING.

The tools and content that Achieve provides are choreographed around a pedagogical model that promote active learning. The model provides a variety of end-to-end structured courses that increase instructor efficiencies and support student success. The active learning model has built-in opportunities to support student outcomes beyond course instruction and assessment — like motivation, self-regulated learning, relevance, and study skills. The active learning model also enables metacognition by providing preflection and reflection activities that prompt students to evaluate their developing knowledge.

ACHIEVE WAS DEVELOPED BASED ON THREE FOUNDATIONS OF LEARNING SCIENCE

1. Effective learning objectives. Effective learning objectives enable instructional alignment across all instructional and assessment content/components via backward design.

2. Impactful assessment practice. Research shows that an evidence-based, learning-objective-driven assessment strategy addressing cognitive and noncognitive aspects of the learning experience can drive better learner engagement, motivation, self-regulation, and performance.

3. Empowering analytics for instructors and students. The analytics provided to instructors in Achieve provide timely and actionable insights to support teaching and learning.

ACHIEVE'S EFFECTIVENESS AND IMPACT ARE MEASURED BASED ON A FRAMEWORK PUBLISHED BY THE AUTHORS IN 2017: UNPACKING THE BLACK BOX OF EFFICACY

The portfolio of evidence of the effectiveness and efficacy of Achieve has been built through a trajectory of increasingly deep and rigorous partnerships with instructors and students during the development, optimization, and ongoing use of Achieve.

1. *Summer 2017.* Solution co-design with instructors and students
2. *Fall 2017*. User and outcome learning research
3. *Fall 2018.* Formative evaluation
4. *Spring 2019.* Implementation study in early beta testing
5. *Fall 2019.* Replication implementation study in later beta testing
6. *Spring 2020.* Quasi-experimental efficacy study

THIS REPORT PRESENTS RESEARCH FROM SPRING 2019 INTO HOW INSTRUCTORS CHOSE TO IMPLEMENT ACHIEVE

Research procedures

In the Spring 2019 semester 41 instructors teaching one of five disciplines (Biology, Calculus, Chemistry, Composition, and Economics) agreed to participate in an evaluation of Achieve before it was to be used at scale. Before the start of the semester, Instructors completed a thirty-minute training on how to use Achieve. During training instructors were offered suggestions for best-practice implementations based on learning science research. However, specific implementations were not mandated as part of the evaluation. The only requirement was that Achieve be the primary curricular material used that semester.

Data were collected for a mixed-methods analysis. Students and instructors completed surveys 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 course records — quiz, test, exam grades, attendance records, etc. — were shared by instructors at the end of the semester. Data were matched across sources, and descriptive and empirical analyses were conducted.

RESEARCH QUESTION 1

How did instructors implement the beta version of Achieve and how did students engage with it? How did implementation and/or use case vary by educational context and discipline?

- During the Spring 2019 semester, instructors in the sample assigned an average of 76.75 assignments within Achieve, spanning a range from 10 to 336.
- On average, instructors assigned Achieve activities in 92% of the weeks in the semester, spanning a range from 36% to 100%.
- Platform data were mined to examine empirical implementation patterns which survey and interview data were used to contextualize. Four implementations emerged empirically and were validated qualitatively.
    - **Post-class summative (16% of instructors).** Instructors who assigned only post-class summative activities.
    - **Pre-class formative and some post-class (43% of instructors).** Instructors who assigned pre-class formative activities and either post-class formative activities or post-class summative activities.
    - **Pre-class formative and all post-class (31% of instructors).** Instructors who assigned pre-class formative activities and post-class formative and post-class summative.
    - **Pre-class, in-class, post-class (6% of instructors).** Instructors who assigned pre-class formative activities, in-class activities, post-class formative activities, and post-class summative activities.
- There were differences in implementation pattern by educational context. For example, instructors at two-year institutions were more likely to adopt implementation one and less likely to adopt implementation pattern two. Additionally, all instructors adopting implementation one teach at large institutions. And, instructors adopting the most comprehensive implementation (four) were more likely to be teaching at highly selective institutions.
- The majority of instructors in this sample (n=25, 71%) accessed the dashboard reports in Achieve in at least 90% of the active weeks of their semester during which reports were available.
    - 24 of the 25 instructors who accessed dashboards regularly took action based on the insights provided.
    - The primary action that instructors reported taking (92%) was to modify their upcoming class lecture based on what students were struggling with or mastered in a preceding assignment.
- Students in this sample engaged in assigned activities in Achieve at a high rate (80.3% total engagement, as calculated by sum of assignments engaged in/sum as assignments assigned)
    - The rate at which students engaged with activities varied by activity type, with the highest rate being for pre-class activities (85.6%). This finding was found to be true across disciplines and within a single discipline.
- Students in this sample completed assigned activities at a relatively high rate too (73.2% total completion, as calculated by sum of assignments completed in/sum as assignments assigned). Individual student overall total completion rates ranged from 2.4% to 100%.
    - Completion rates varied by activity type, with the highest completion rate being for pre-class activities.
- There were differences in engagement and completion rates for students of different levels of academic preparedness.
    - Students more prepared to succeed had an overall engagement rate of 83% as compared to 78% among students less academically prepared to succeed. This significant difference was largely influenced by students in Economics courses. In other disciplines studied significant differences in engagement rates did not emerge.
    - Students less prepared to succeed had significantly higher engagement rates in in-class activities (62%) than students more prepared to succeed.

RESEARCH QUESTION 2

How did instructors and students perceive the beta version of Achieve?

Instructors had high perceptions of Achieve, despite the product being in a beta and evolving versions during the first of three semesters of beta testing. The results that follow should therefore be interpreted in the context of Achieve's stage in the development lifecycle during this study.

- Instructors were asked to rate on a scale from 1 to 10 ("Would definitely not recommend" to "Would definitely recommend," the extent to which they would recommend the Spring 2019 beta version of Achieve to a colleague based on their experience with it through the semester. The average rating of likelihood to recommend was 6.8.
- Instructors were also asked to rate on a scale from 1 = "Will definitely not adopt" to 10 = "Will definitely adopt" based on their experience with the Spring 2019 beta version of Achieve through the semester how likely they were to adopt Achieve when it is launched commercially in Fall 2020 (if the price was reasonable and it was approved by their department). The average rating of likelihood of adoption was 6.6.
- Instructors teaching Chemistry had the strongest positive perception of Achieve (likelihood to recommend = 8.1), with Economics instructors reporting similarly positive perceptions (likelihood to recommend (Mean = 7.2).
- In an effort to measure the extent to which the beta version of Achieve is better than their previous product experience, we asked instructors on the baseline survey to rate, on a scale of 1 = "strongly disagree," 2 = "disagree," 3 = "agree," 4 = "strongly agree," the extent to which they agreed with a set of statements about their current approach to teaching their course. At the end of the semester we asked instructors to rate, on the same scale, the extent to which they agreed with the same statements about Achieve.
    - None of the differences between the average rating of perception of current approach and use of Achieve were meaningfully different.
- Instructors were asked to rate, on a scale of 1 = "very difficult," 2 = "difficult," 3 = "easy," 4 = "very easy" how difficult a set of activities were in their course the last time that they taught it using another product. At the end of the semester we asked instructors to rate, on the same scale, how difficult the same set of activities were in their course this semester using Achieve.
    - In all cases instructors rated the behaviors that Achieve was designed to support less difficult, on average, during the semester that they were using Achieve as compared to the last semester that they were teaching the course using another product.
        - Implementing active learning strategies (+0.88*)
        - Assessing how well students are comprehending the material (+0.85*)
        - Promoting students coming to class prepared to participate (+0.82*)
        - Fostering ability to remember information (+0.60*)
        - Promoting student collaboration (+0.54*)
        - Fostering deep insights (+0.39*)
- 8% of the instructors in this sample either agreed or strongly agreed that Achieve supported mastery more than if Achieve had not been used (M=3.03, SD = 0.54).

Students also had moderately high perceptions testing the beta version of Achieve.

- Students were asked to rate, on a scale of 0 to 10, their likelihood to recommend the same course to a friend if they knew that Achieve was going to be used. Average student rating was 6.83 out of 10.

- Differences in ratings between implementation groups were statistically significant with students assigned only post-class summative activities rating it the highest (7.63).

- One question on the post-survey asked "please rate the extent to which you agree that Achieve was easy to use" (scale: 1 = "strongly disagree," 2 = "disagree," 3 = "agree," 4 = "strongly agree"). In general, students agreed (M=3.06, SD=0.71) that Achieve was easy to use.

- On average, students had moderately more positive perceptions of Achieve than the previous digital learning tool(s) used (scale: 1 = "strongly disagree," 2 = "disagree," 3 = "agree," 4 = "strongly agree").

  - Achieve helped me gain better mastery of course content (2.78)

  - Achieve was easier to use (2.69)

  - Achieve is more engaging (2.68)

  - Achieve increased efficiencies (2.64)

  - Achieve motivated me to learn more (2.63)

- Students reported that actively engaging in classroom discussion was significantly less difficult during the semester that Achieve was used than it typically is for them ($t(845)=2.13$, $p=0.034$).

RESEARCH QUESTION 3

Is the use of the beta version of Achieve related to academic performance in the course?

We hypothesized that more use of Achieve would positively influence assessment scores. Based on previous research however, we suggested that a student's level of academic preparedness when entering college would also contribute to this relationship. We also suggested that students being nested within instructor in this sample would influence the relationship. So, we proposed that high school grade point average (HSGPA) should be controlled for and a hierarchical linear model employed to account for prior academic performance and nesting in the data.

- The dependent variable final exam score was first examined descriptively. Valid final exam score data were available for 1,703 students and scores ranged from 0.00 to 104.40 with an average score of 74.92 (SD = 20.12). The skewness of the distribution was -1.88 and kurtosis was 3.90.
- Then the correlation between student engagement in Achieve (calculated by the total number of assigned activities engaged in/total number of assigned activities) and student final exam score was calculated — a significant correlation was found of .54 ($p<.0001$).
- "Rate of engagement" bands were developed and the average final exam score for the students that fell in each band was calculated. Students who engaged in 0%-20% of assigned activities (n=68) earned an average final exam score of 44.43. Those who engaged in 21%-40% (n=52) earned an average of 47.79. Those who engaged in 41%-60% (n=126) earned an average of 58.85. Those who engaged in 61%-80% (n=423) earned an average of 73.48 and those who engaged in 81% to 100% of assigned activities (n=1,034) earned an average of 80.84.
- To account for prior academic performance, baseline level of motivation to succeed, and instructor, a hierarchical linear model was calculated using PROC MIXED in SAS. We evaluated the change in AIC and BIC and concluded that Model four (the model including rate of engagement) was the best fitting model. And, given that the inclusion of student level of engagement in Achieve emerged as significant we have evidence to conclude that use of Achieve is predictive of academic performance in the student's course. More specifically, the more assigned activities that a student engages in, the higher they can expect their final exam score to be in the course, regardless of their level of academic preparedness coming into the course.
- For every ten percent increase in a student's engagement in assigned activities, they can expect a 5.7 percentage point increase on their final exam score.
- Within each discipline, the same analyses were repeated and a significantly positive relationship emerged.
- Due to a limitation of data, correlations (rather than hierarchical linear models) were calculated to examine the relationship within implementation pattern. Results suggest that the more comprehensive the implementation (see use cases 1-4 previously described), the more overall variance in final exam score use of Achieve accounted for.

LIMITATIONS

- This study was conducted in the first semester of testing a beta version of Achieve, meaning that the tool was in a formative state and still evolving. As such, instructors may have used Achieve in a different way than if it was fully developed. Implementation patterns that emerged in this study will be compared with those observed in a replication study the following semester to understand if different user cases are more robust with a more fully developed tool.
- Instructors who agree to participate in an early test of a beta version of a new digital learning tool are likely to have more positive perceptions because they are looking for innovation and are more comfortable with technology. As we discussed in the description of our sample, nearly all instructors were comfortable with technology and have positive perceptions of it.
- The design and analyses presented in this study are correlational and therefore causal statements cannot be made based on the results. Although we controlled for student prior academic performance and baseline level of motivation, there are a myriad of other factors that could be contributing to the outcomes measured.

FUTURE RESEARCH

- In the Fall 2019 semester a replication implementation study is being conducted with a more developed beta version of Achieve. In the replication study we are partnering with a larger and more representative sample of instructors using a more evolved version of Achieve. We will replicate the analyses to investigate whether the trends that are presented in this report persist when studying a more evolved version of Achieve.
- In the Spring 2020 semester a quasi-experimental study will be conducted with a fully developed version of Achieve. In that study we will investigate the impact that Achieve has on teaching and learning outcomes including any causal inferences of efficacy.

Introduction

The higher education landscape has undergone seismic shifts in the past decade. The make-up of the student population is changing as institutions realize growing enrollments of adult learners returning to higher education while juggling competing demands like careers and families. Administrators and educators are reconsidering pedagogical approaches as they work to develop lifelong learners who have the cognitive and non-cognitive skills needed to succeed in the global economy. And, as women and men enter the workforce with mounting student debt, a return on a student's investment in higher education is becoming heavily scrutinized. In the wake of these changes the need for highly effective curricular material that supports the success of a diverse student population has never been more important.

Advances in the learning sciences combined with the fast evolution of powerful digital technologies and the evolution of the design of user experience have the potential to transform the future of higher education. There is now a large body of research demonstrating that learning science principles such as self-regulation, formative assessment, and active learning support the development of lifelong learners (Bell & Kozlowski, 2008). Digital tools can increasingly offer content, resources, and assessments that are personalized, adaptive, and relevant (Cook et. al, 2013). And, seamless, efficient, engaging digital experiences that make the lives of students and instructors easier can meet the needs of busy educators and a diverse student population (Herrington & Oliver, 2000). However, despite a host of digital learning tools available to instructors there is a lack of rigorous and relevant evidence researching effectiveness and this has often led to false starts and frustrations of what to use and how to use it to best effect to improve student success (EdTech Efficacy Symposium, 2017).

Significant progress has been made in the work toward understanding educational technology efficacy and the standards for conducting efficacy studies. However, critical gaps in providing timely and actionable insights that support instructors persist. Three such gaps are:

1. *Evidence of efficacy is established after instructors have made adoption and implementation decisions*. Most studies investigating educational technology efficacy, while strong studies, were conducted after the tool being researched was already being used at scale (e.g Tingley, 2017; Furr & Williamson, 2019; Mojarad, Essa, Mojarad, Baker, 2019; Parker & Loudon, 2012). The insights derived from these studies can support the decisions made by instructors who are considering using the tool, but many instructors have already made adoption and implementation decisions without evidence to support their choices.

2. *Design, development, and iteration process as part of the efficacy argument*. Most efficacy research appropriately focuses on outcomes, but often very little attention is paid to the design, development, and iteration process before the tool is used at scale. Outcome research provides insights into the "what" (i.e does the tool impact retention) but the upstream efficacy helps to explain the "why" (i.e. this tool impacts retention because it was co-designed with a representative sample of students who communicated which student support tools would help them persist in their courses). The efficacy of educational technology begins at ideation and is part of the evidence instructors should be able to evaluate when deciding whether to use a tool. An understanding of the design, development, and iteration process can help instructors evaluate whether the tool will meet their needs and resonate with their students.

3. *Effectiveness and/or Impacts are investigated within specific educational contexts and/or a specific implementation pattern*. Effective digital learning tools are flexible, and can be implemented in a way that most enhances an instructor's pedagogy based on the needs of their students and their educational context. When an efficacy study is conducted with one instructor, in a unique educational context, who adopts a specific implementation pattern, those results are only relevant for that use case. For example, it is unlikely that results from an efficacy study conducted at a large two-year institution where the tool was used only as an in-class engagement tool will be relevant for an instructor teaching at a four-year institution using the tool as a pre-class and in-class solution. Efficacy studies should be conducted in partnership with instructors teaching in various educational contexts and among a set of naturally occurring implementation patterns to document how instructors choose to use the tool and the results their students realize.

When building Achieve, a new digital learning platform, Macmillan Learning began by founding the solution on research-based learning science principles, and co-designing the platform with instructors and students. Researchers at Macmillan Learning took the unusual approach of beginning to investigate the effectiveness and efficacy of Achieve in its infancy by conducting rapid cycle evaluations of tool features. Then, implementing increasingly rigorous validity and efficacy studies as the platform matured through beta testing. This approach allowed researchers to provide insights to development teams so that they could make evidence based decisions around development and optimization. And, it enabled valid, reliable, and timely insights to be delivered to instructors so they could make informed adoption and implementation decisions as soon as Achieve is available for adoption in Fall 2020. This manuscript discusses the learning science foundation of Achieve and presents the results from an implementation study conducted in the first semester of beta testing. The results presented should be interpreted as exploratory and directional, given that the tool was in development during the study, but they are important early insights, nonetheless. At the time of this manuscript publication a replication implementation study is underway and a quasi-experimental study is planned for Spring 2020. Taken together the results from all studies conducted in beta will provide administrators and instructors a robust portfolio of evidence to consider when deciding whether to use Achieve.

This report begins by providing a complete description of Achieve and the multi-year, multi-partnership ongoing efficacy agenda that was designed to mitigate the limitations of most efficacy research. The paper then focuses on presenting the technical documentation and results of the first in-context Achieve efficacy study conducted during the Spring 2019 semester, the implications for instructors based on the findings, and the limitations of the study. The report concludes with a forward look at the study underway in the Fall 2019 semester and other planned research that will be presented in updates of this efficacy report.

Achieve

Achieve is a digital learning solution developed for higher education courses (at the time of this report publication Achieve was available in five disciples: Biology, Calculus, Chemistry, Composition, and Economics). Achieve provides a connected suite of course tools designed to give instructors choice, with flexible recommendations for optimal pedagogical structures based on the learning sciences. The key principles that Achieve is built on include: everyone has the potential to learn, each learner starts at a different place and learns at their own pace, cognition can be enhanced through technology, an instructor's pedagogy matters, learning is a social activity, and students should be empowered to manage their learning.

Achieve was conceived based on six learning design principles that underpin all Macmillan products, as well as a series of robust learning science foundations that support active learning, objective-driven instruction, formative assessment, and actionable analytics. And, it has been optimized based on the findings of research conducted in close partnership with instructors and students.

ACHIEVE WAS CONCEIVED BASED ON SIX LEARNING DESIGN PRINCIPLES

1. Develop Learner Motivation. When students are highly motivated, they are able to tackle challenging problems and strive to accomplish goals that will improve their abilities. However, there is no one way to motivate students — there is no "magic bullet" (Conley, 2012, p. 44). Rather, instructors and instructional technologies can support motivation by providing opportunities for success, and by framing errors and struggles as important elements in the processes of growing and learning (Bjork, Dunlosky, & Kornell, 2013; Joët, Usher, & Bressoux, 2011; Wolters, 2004). Motivation is an influential mediator of learning as it regulates cognitive processing and affect (Mayer, 2014). Being in a positive affective state and possessing high levels of autonomy can enhance motivation (Pekrun, Elliot, & Maier, 2009; Schumacher

& Ifenthaler, 2018). Thus, it's important to encourage motivation through learner autonomy, goal setting, and positive feedback that focuses on the task, learner process, and/or self-regulation.

2. Provide Personalized and Adaptive Experiences. Students enter the classroom with a variety of cultures and psychological traits, thus, personalization and adaptation of instruction and assessment can have positive effects for all learners (Alexander, Schallert, & Reynolds, 2009; Sternberg, 2007). Personalized learning environments such as dashboards, which can be adapted by learners, can help students to modify their learning strategies and foster skills in managing, monitoring, reflecting, and motivating their own learning (Knox, 2017; Park & Jo, 2015; Roberts, Howell, & Searman, 2017; Schumacher & Ifenthaler, 2018). Within a course, dynamic testing can improve predictive models of student success and enhance learners' metacognition, learning efficacy, and performance while providing immediate feedback, scaffolding questions, and hints (Feng, Heffernan, & Koedinger, 2009; Tseng, Chu, Hwang, & Tsai, 2008). These systems and tools must be developed through a process considering and involving students' needs at all stages and time-periods of the course (Santos, Boticario, & Pérez-Marín, 2014).

3. Target Cognitive and Memory Elements. Today, there are numerous methods that learning scientists have researched that enhance learner cognition and transfer — all of which begin with learning objectives, which describe "the intended change in knowledge" (Mayer, 2008, p. 762) and can enable a mastery approach which has positive impacts on conceptual learning, attitudes toward learning, and performance (Pekrun, Elliot, & Maier, 2009). Learning objectives enable instructors and instructional technologies to foster desirable difficulties, interleaving and/or spaced practice which increase storage strength and long-term retention and ultimately aid in performance (Bjork & Bjork, 1992; Bjork, Dunlosky, & Kornell, 2013; Credé, Roch, & Kieszcznkya, 2010; Ehrlinger, Mitchum, & Dweck, 2016). Retrieval-based learning, exercised through certain study methods and during assessments, enhances later performance (Agarwal, Bain, & Chamberlain, 2012; Bjork et al., 2013) and frequent quizzes can support a "testing effect," strengthening students' memories for the retrieved information (Delozier & Rhodes, 2017). Immediate feedback on assessments can lead to high procedural knowledge (Fyfe & Rittle-Johnson, 2016), improve low confidence on correct answers and enhance later performance (Agarwal et al., 2012), and can revise misunderstandings through the use of causal explanations (Hattie & Timperley, 2007; Kendeou, Walsh, Smith, & O'Brien, 2014).

4. Build on Well-Constructed Learning Models. Being cognitively engaged stimulates learning, specifically, learning that "sticks" (Hirsh-Pasek, Zosh, Golinkoff, Gray, Robb, & Kaufman, 2015, p. 9). Active learning, which can be fostered through models including Project-Based Learning (PjBL) and Problem-Based Learning (PBL), leads to the growth of complex reasoning skills, critical thinking processes, perceived learning (e.g., better conceptual understanding of material, retention of knowledge, transfer of knowledge to new problems), engagement, attitudes towards and perceived usefulness of subjects, self-directed learning, exam performance, motivation, and autonomy (e.g., Crouch & Mazur, 2001; Goedert, Pawloski, Rokooeisadabad, & Subramaniam, 2013; Kay & Kletskin, 2012; Muehlenkamp, Weiss, & Hansen, 2015; Akinoğlu & Tandoğan, 2007; Sawyer, 2014; Tseng, Chang, Lou, & Chen, 2013). PjBL and PBL are considered active learning models because students are required to take responsibility for their own learning processes (English & Kitsantas, 2013).

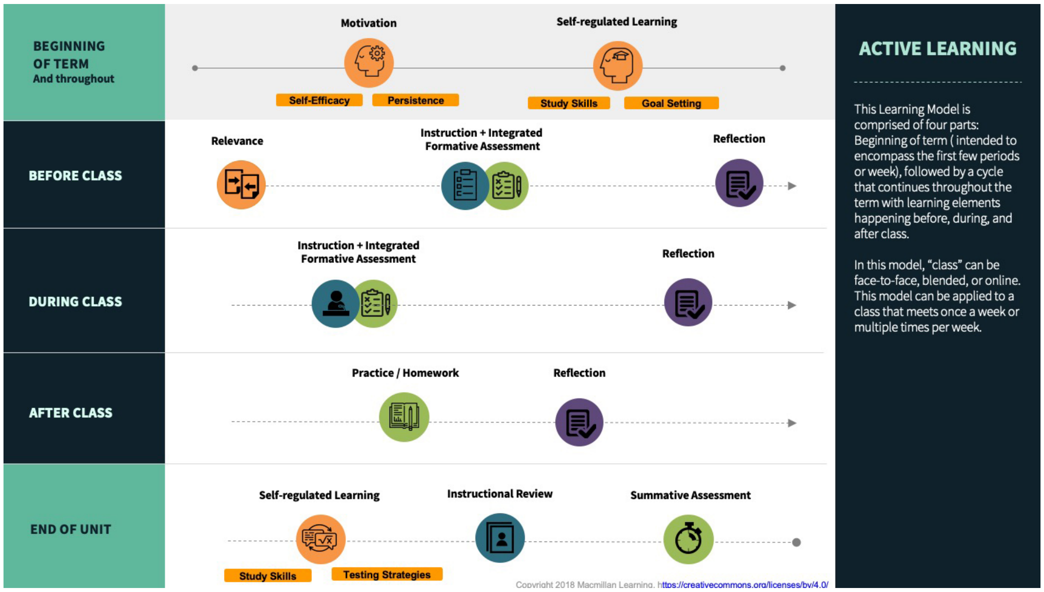
5. Create Interactive and Constructive Opportunities. The development of critical-thinking skills and higher-order learning benefit from collaborative learning, which leads to enhancement in academic performance and intellectual development (Bai & Chang, 2016; DeLozier & Rhodes, 2017; Hirsh-Pasek, Zosh, Golinkoff, Gray, Robb, & Kaufman, 2015). When students are able to take an active role in discussions, choose their own topics, and ask questions, they co-construct knowledge and engage in high-level co-regulation through making their thinking explicit and evaluating their peers' and instructors' perspectives (Do & Schallert, 2004; Volet, Summers, & Thurman, 2009). With peers, students can work together to revise misunderstandings (Crouch & Mazur, 2001), and engage in self-explanations which promote "prior knowledge activation, inference generation, and revision of existing knowledge" (Richey & Nokes-Malach, 2015, p. 203). Ultimately, social relationships are important to develop in classrooms as they have strong impacts on student performance, persistence, and retention (Bernardo, Esteban, Fernández, Cervero, Tuero, & Solano, 2016). When instructors interact directly with students, it reduces the transactional distance between them, thereby increasing student retention (Simpson, 2013).

6. Enable Metacognition and Self-Regulation. Metacognition and self-regulation is critical for academic success. Students typically spend time studying items they do not know well, thus, a metacognitive judgment can lead to the decision to terminate learning or to continue. If the judgment is inaccurate, revision time will not be allocated effectively (Dunlosky & Metcalfe; McDaniel & Butler, 2010). Metacognitive illusions, most frequently that learning strategies that feel difficult are not as productive as those that feel easy, can lead to low levels of achievement (Bjork, Dunlosky, & Kornell, 2013). Accurate judgements, however, can lead to the correction of misconceptions and an increase in academic performance (Richey & Nokes-Malach, 2015). Similarly, if students struggle with regulating their learning processes, they will likely become less engaged, make poor study choices, and become less successful in their courses (Ehrlinger, Mitchum, & Dweck, 2016; Kizilcec, Pérez-Sanagustín, & Maldonado, 2017). Learning success is "predominantly attributed to students' self-regulation capabilities that are relevant for initiating and sustaining learning processes" (Schumacher & Ifenthaler, 2018, p. 397), which are especially relevant for achieving learning goals (Cho, Kim, & Choi, 2017). In hypermedia environments, self regulation skills are imperative in navigating and learning from multiple representations, especially as those who do not regulate their learning tend to become more easily overwhelmed (Green & Azevedo, 2009).

ACHIEVE HAS BEEN CHOREOGRAPHED AROUND A PEDAGOGICAL STRUCTURE TO PROMOTE ACTIVE LEARNING

A pedagogical structure developed to promote active learning acted as a blueprint for the choreography of Achieve. The model provides an end-to-end structured course that increases instructor efficiencies and supports student success. The active learning model has built-in opportunities to support student outcomes beyond course instruction and assessment — like motivation, self-regulated learning, relevance, and study skills. The active learning model also enables metacognition by providing preflection and reflection activities that prompt evaluation of developing knowledge. And, a host of proven instructional content, like publisher provided materials, lecture slides, and instructional reviews offer opportunities to provide new, or review learning-objective aligned instructional information. Integrated formative assessment, practice activities and homework, and end of unit or term summative assessments provide an ongoing assessment of learning and feedback for increased learning.

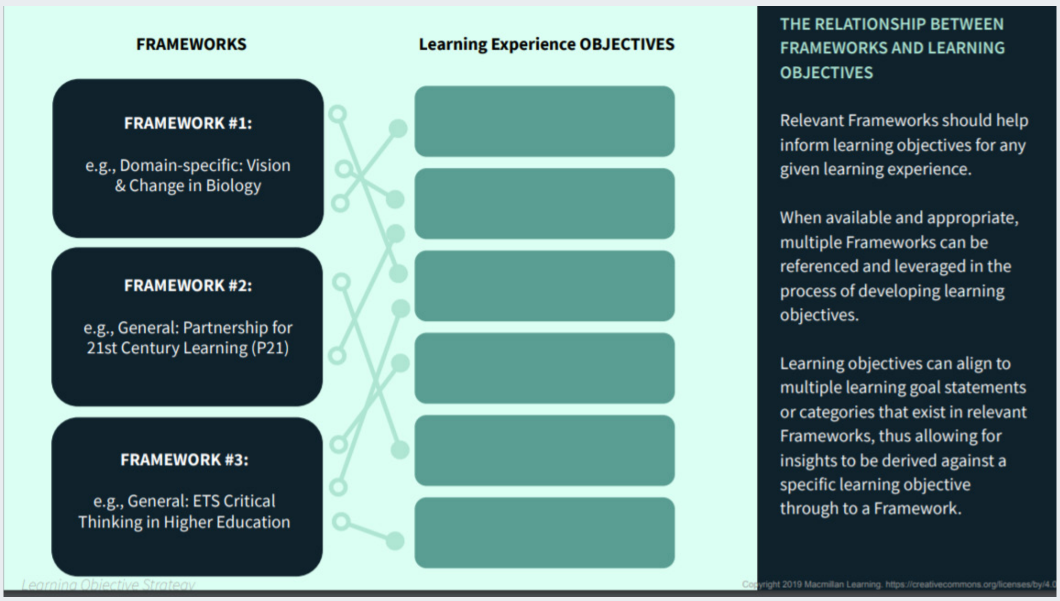
Image 1. The learning model underpinning Achieve



ACHIEVE WAS DEVELOPED BASED ON THREE CORE LEARNING SCIENCE FOUNDATIONS

1. Effective learning objectives. Effective learning objectives enable instructional alignment across all instructional and assessment content/components via backward design. They drive assessment task development and implementation and enable high-quality, targeted reporting of learning progress and performance for learners, instructors, institutions, and meet accreditation and employability requirements. They also facilitate personalization and adaptive learning capabilities. The learning objectives in Achieve support learner engagement and improved learning outcomes because they are comprehensive and derived from relevant frameworks. They support self-regulated learning by clearly and concisely describing criteria for success, providing transparency into all aspects of the learning experience, and by challenging learners but also providing the appropriate level of scaffolding to achieve broader, more cognitively complex learning goals. The assessments in Achieve are enhanced by learning objectives because they are measurable and support the creation and implementation of high-quality assessment tasks, facilitate formative and summative assessment best practices, and support intervention whereby instructors and students can remediate gaps in understanding.

Image 2. The relationship between frameworks and learning objectives in Achieve

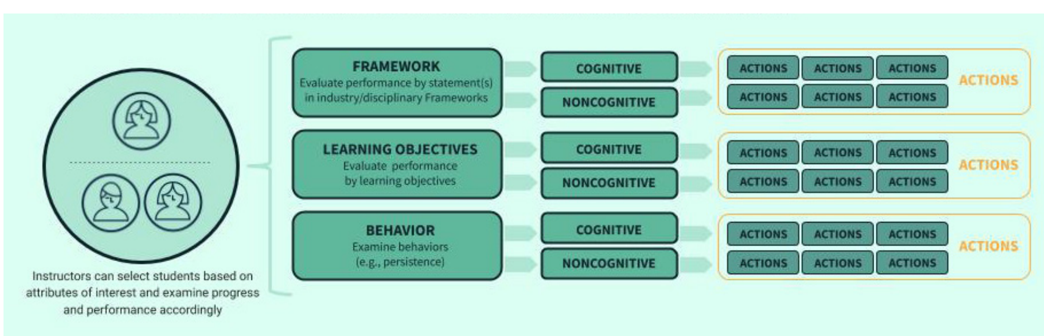


2. Impactful assessment practice. Research shows that an evidence-based, learning-objective-driven assessment strategy addressing cognitive and noncognitive aspects of the learning experience can drive better learner engagement, motivation, self-regulation, and performance. Research also shows that active, constructive, and interactive learning activities provide opportunities for learners to be more engaged and support a deeper, more impactful learning experience. The assessments in Achieve are effective because they take a learning-objective driven, integrated approach to assessment which improves transparency of learning goals to stakeholders, supports instructional alignment, and enables monitoring of learner progress, and timely, targeted intervention. Achieve assessments also provide high-quality feedback using an evidence-based approach to the nature, tone, and timeliness of feedback provided to learners in both formative and summative contexts. This approach improves motivation, affect, and metacognitive abilities, can support self-regulated learning strategies, and can improve performance by addressing gaps in understanding.

Image 3. An assessment strategy built on sound measurement principles

3. Empowering analytics for instructors and students. The analytics provided to instructors in Achieve provide actionable insights to support teaching and learning. The information provided in instructor-facing dashboards report against learning objectives to enable monitoring and improving mastery of concepts, application of skills, and development of attributes. They provide strategic feedback to enable learners to better understand their current performance, how they should be performing, and how they can close the gap between the two. The analytics further support metacognition and self-regulation by helping learners more accurately and efficiently gauge their progress and adjust their practices accordingly. The insights further have the potential to enhance interaction and collaboration by fostering productive instructor-to-learner interaction and collaboration increases learner engagement and performance. Finally, the analytics enable effective interventions by providing valid insights that are visualized in ways that reduce extraneous cognitive load.

Image 4. A framework for empowering analytics



THE CONTENT, ASSESSMENTS, AND COURSEWARE THAT MAKE UP THE STRUCTURED LEARNING EXPERIENCE INCLUDE:

eBook. There is an eBook included with each Achieve offering. eBook content is delivered in two ways. The complete eBook can be downloaded to read or instructors can assign eBook "snippets" if they want to direct their students to specific content delivered in manageable sections. The eBook offers highlighting, note-taking, offline access, and read-aloud functionality. If assigned, eBook snippets are worth one or more points. If unassigned but made available, eBook snippets are not worth any points but can be accessed by students.

Video-tutorials. Students are asked to view brief, five to seven-minute videos. In most cases there was an associated set of non-adaptive quiz items related to the content in the video. After playing the video once, students were permitted to view the content in subsequent replays without any penalty. Where quiz questions were provided, students were given multiple attempts to complete the questions correctly and could refer back to the video while responding to the questions. The total points that a student received for each question decreased with each incorrect attempt. Students were also given the option to receive hints about the answer to the question which decreased the total points available to them per item.

Adaptive, gamified reading quizzes. Instructors identify a "target score" for a reading quiz that students are expected to reach, and students have to continue answering questions until they reach that score. If a student answers a question correctly on the first try, they earned the total number of points for that item; each subsequent attempt decreases the total number of points earned for that item. If a student gets a quiz item wrong they are presented another item with similar content and similar level of difficulty; if they answered the item correctly they are presented an item assessing different content with slightly more difficulty. Students are given the option to receive hints about the answer to the question which decreases the total points available to them per item. Students can also request to be taken directly to the portion of the ebook where the content needed to correctly answer the question can be found. Student scores did not decrease if they requested to "refer to the text". Students can also select the "show answer" tab, but earn no points for the item if they selected this option. The adaptive reading quiz is graded for completion. Students receive full credit (100) if they reach the target score and no credit if they didn't (0).

In-class activity guides. The in-class activity guides provide suggested implementations of activities that are intended to increase student engagement. Activity guides suggest preparatory materials (e.g. pre-class activities) that are aligned to the suggested in-class activities, have associated in-class iClicker questions, and homework activities to reinforce the concepts.

Lecture slides. Powerpoint lecture slides are included for each chapter of the associated eBook to complement classroom lectures. Slides can be delivered as presented in Achieve or modified.

iClicker | REEF. Instructors have the option of integrating iClicker, a student response system, into Achieve. Pre-written iClicker questions are included in Achieve that can be used to reinforce concepts and increase student engagement in the classroom.

Formative homework. Instructors set the post-class assessment to the "multi-response" setting. Instructors choose questions from a bank that include multiple item types — such as clickable area, ranking, sorting, labeling, multiple choice, multiple select, graphing, and numeric entry. If a student answers a question correctly on the first attempt they receive the total number of points that the question is worth. If a student answers a question incorrectly they are provided wrong-answer specific feedback targeted to their misconceptions. Then, students can try again, however the number of points earned is reduced. Students are given the option to receive hints about the answer to the question which decreases the total points available to them per item. Fully worked out solutions are provided for each question and referred to as "in-tool study guides." Formative homeworks are graded for performance. Student scores can range from 0 to 100.

End of chapter problems. Many chapters include end of chapter questions. End of chapter questions can be assigned as formative (allowing students multiple attempts to answer the question correctly) or summative (permitting only one attempt to answer the question correctly). In other scenarios, students can ask for a hint, reducing the total number of points they can receive for the question and they are provided fully worked out solutions for each question. End of chapter problems are graded for performance. Student scores can range from 0 to 100.

Summative assessments. Summative assessments include multiple item types — such as clickable area, ranking, sorting, labeling, multiple choice, multiple select, graphing, and numeric entry. If a student answers a question correctly on the first attempt they receive the total number of points that the question is worth. If a student answers a question incorrectly they receive zero points for that item. Summative assessments are graded for performance. Student scores can range from 0 to 100.

Content and assessments are housed in Achieve courseware which also provides resources that support instructor and student efficiencies, instructor course management, and student self-regulation.

Course planner. Instructors set up and manage their courses through the course planner which includes assignment, scoring, and due date management.

Insights and reporting. Instructor facing dashboards provide insights specific to the activities or assessments assigned in the course. Insights are provided at the unit, subunit, learning objective and student-level.

Gradebook. Performance on all activities or assessments assigned in Achieve is recorded in the gradebook which can be exported or synced to a campus LMS.

ACHIEVE'S EFFECTIVENESS AND IMPACT IS MEASURED BASED ON A FRAMEWORK PUBLISHED BY THE AUTHORS IN 2017

From ideation through use of scale, instructors and students have been an integral part of the design, development, and optimization of Achieve. Close attention has been paid to partnering with representative samples of instructors and students to ensure all backgrounds, experiences, needs, and expectations were reflected in Achieve. Table 1 outlines the trajectory of increasingly deep and rigorous partnerships with instructors and students during the development and optimization of Achieve.

A full report of the Spring 2019 Achieve efficacy study follows the framework (the Fall 2018 Formative Evaluation report can be found here).

Table 1. The framework for evaluating the effectiveness and measuring the impact of Achieve

	PREVIOUS			CURRENT	FORTHCOMING	
	Summer 2017	Fall 2017-Spring 2018	Fall 2018	Spring 2019	Fall 2019	Spring 2020
Research performed	Co-designing with instructors and students	User and outcome learning research	Formative evaluation	Implementation studies in early beta testing	Replication implementation studies in later beta	Quasi-experimental efficacy studies
Partnerships forged	12 instructors and 10 students attended summer workshops held collaboratively with product developers, learning researchers, and user experience designers	88 students join the Student Codesign Group, 5 Learning Research experts partner on the design of the learning model	39 instructors representative of five disciplines, teaching at two- and four-year institutions among a variety of educational contexts partnered with validity researchers	40 instructors and 2,206 students representative of the target educational contexts	59 instructors and 3,500 students across six disciplines, representative of the target educational contexts and class size and user segments partnered with efficacy researchers	70 instructors and 5,000 students across six disciplines acting as treatment and control to examine impact in a controlled partnership
Questions explored	What are the real life journeys of instructors and students who will be using Achieve? What are the personas of the instructors and students who will be using Achieve?	What educational research are we using as the foundation of the product design? What learning science foundations should we base development on?	Is there evidence to validate our hypotheses about Achieve's core value propositions? What are instructor perceptions of Achieve? How do instructors expect that they would use Achieve?	How do instructors implement Achieve in their courses? Do implementation patterns vary among educational context? Is use of Achieve related to student outcomes? What are student and instructor perceptions of Achieve?	When Achieve is used with a larger, more representative sample, do the same implementation patterns emerge? Is there differential efficacy of Achieve? That is, do different sub-populations realize different outcomes?	What is the impact of using Achieve on student engagement, retention, and learning? Can Achieve help close the achievement gap in higher education? Does Achieve impact the outcomes of the groups of students instructors most often ask about?
Impact on Achieve	Created the user journeys that would help structure the learning model underpinning Achieve	Formed the learning science and design principles that Achieve would be developed based on	Derived formative insights to inform the product development roadmap	Derived implementation and effectiveness insights to support product development and instructor decision-making	Will enable valid and reliable insights for instructor-decision making in all educational contexts	Will provide instructors with rigorous evidence of impact

Spring 2019 Achieve efficacy study

PROCEDURES

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

In the Spring 2019 semester 41 instructors across five disciplines (Biology, Calculus, Chemistry, Composition, and Economics) agreed to participate in an evaluation of Achieve before it was being used at scale. Instructors and students received Achieve free of charge to use. All students were required to use Achieve in their course because it was the curricular material their instructor selected, but they were not required to participate in this study. Interested students were required to actively consent to participate. The 2,206 students in the evaluation study made up 74% of all students enrolled in participating courses.

Prior to the beginning of the semester instructors were required to complete a thirty minute training on Achieve. During training instructors were offered suggestions for best-practice implementation based on learning science research but implementation patterns were not mandated as part of the evaluation. The only implementation requirement was that Achieve had to be the primary curricular material used that semester.

Specific textbooks were available to be used with Achieve based on an instructor's discipline. Table 2 outlines the textbooks used by discipline.

Table 2. Textbooks used with Achieve

Discipline	Textbook	Instructor count
Biology	How Life Works 3rd Edition; Morris	7
Calculus	Calculus ET 4th Edition; Rogawksi	4
	Calculus 2nd Edition; Sullivan	5
Chemistry	Interactive General Chemistry with General Chemistry Readiness	8
Composition	Writer's Help 3.0, Hacker	3
	Everyday Writer 7th Edition, Lunsford	3
Economics*	Macroeconomics 5th Edition, Krugman and Wells	6
	Microeconomics 5th Edition, Krugman and Wells	**7**

*Some Economics instructors taught Macroeconomics and Microeconomics, total 11 Economics instructors

RESEARCH QUESTIONS

This study addressed three research questions designed to help educators better understand whether Achieve would be an effective solution in their educational context and how they might implement it to best effect.

1. *How are instructors using Achieve and how are students engaging with it?*
2. *What are student and instructor perceptions of Achieve?*
3. *Does use of Achieve influence final-exam scores? Is there a difference in the relationship between usage and exam scores by implementation pattern? Do student characteristics 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

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 was shared by their instructor during the first two weeks of the course. 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 pre-class activities 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

A survey that asked instructors to report their background and demographic characteristics 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 Achieve, intended implementation of other publisher-provided digital learning tools or open educational resources, previous experience with Achieve, and general early perceptions of Achieve. These data were used to control for instructor characteristics and to better understand intended implementation of Achieve.

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 Achieve in the previous week (which features and components they used), how much time various activities took them, their perception of Achieve that week, any benefits or challenges of using Achieve, 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 actual implementation and how that was related to perception.

Instructor interviews

An instructor interview protocol was developed that gathered information on how an instructor was implementing Achieve, why they decided to implement it in that way, their perceptions of Achieve, and their perceptions of how their students were accepting Achieve. 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 better understand why an instructor chose to assign pre-class activities or not assign pre-class activities.

Instructor post-survey

A survey was administered online during the last two weeks of the course to instructors. 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 Achieve and their perceptions of Achieve; 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

A survey was administered online during the last two weeks of the semester. The survey asked students to share demographic data, personal device data, how they used Achieve, their perceptions of Achieve, their engagement in the course[3], their satisfaction with the course, a System Usability Scale, and a Net Promoter Score. These data were used to measure whether there were systematic differences between students who engaged in pre-class activities and those who didn't.

Product usage data

The following data were extracted from the Achieve 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. These data were used to measure actual instructor implementation and student usage.

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 some data were not available for all students. 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. In this report, final exam scores were the only student record used.

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 DESCRIPTION

Although the total study sample was derived from instructors and students in Biology, Calculus, Chemistry, Composition, and Economics courses, the analyses and findings presented in this report reflect the usage, engagement, and performance of instructors and students in Biology, Calculus, Chemistry, and Economics courses only. These courses were selected for the Achieve efficacy report because they are all structured similarly and can be analyzed at the aggregate. Achieve for Composition is a different "product model" and its efficacy is reported in the Achieve for Composition efficacy report. Therefore, to collect evidence of the efficacy of Achieve in these disciplines, we investigated data collected from 1,991 undergraduate students enrolled at 35 institutions in 38 courses among 35 instructors (79% of the total population of students enrolled in those discipline courses).

Institutional

Across institutions represented in this sample 42% were two-year and 58% were four-year, the majority of institutions (76%) had more than 2,000 undergraduate students enrolled, and based on the Carnegie Classification of Institutions of Higher Education, 45% were more selective, 34% were moderately selective, and 16% were less selective.

Instructor

To ensure we engaged a representative sample, and provide context by which to interpret results, instructor background and pedagogical data for the instructors in this sample are included in Appendix Table A1. Biology, Calculus, Composition, and Economics instructors totaled 35 instructors. The largest proportion of instructors in the sample (40%) have been teaching for more than 15 years, with 26% teaching 6-10 years, 23% teaching 11-15 years, and 11% teaching 1-5 years. The majority of instructors (80%) reported being either comfortable or extremely comfortable with technology and nearly all either agreed or strongly agreed that technology can enhance their pedagogical framework. About 66% of the instructor sample had used publisher provided digital learning tools during the last semester that they taught this course (48% had used other Macmillan Learning tools), their reported reasons for using them varied with 35% reporting increased efficiencies from publisher provided tools, 30% reporting wider breadth of resources, 22% reported that they supported more positive student behaviors and 13% reporting that publisher provide digital tools were more effective than print alone.

The instructors in the sample were also asked to report on their typical pedagogical approaches in this course. The majority (74%) reported that they typically implement active learning strategies including peer-to-peer engagement, in-class activities, case studies and current events investigated in groups, student response systems, and games. The 16% not implementing active learning strategies reported this is because they either don't have enough time to prepare the activities, don't have enough time in class to implement the activities, or they do not think active learning is as effective as direct instruction.

Most instructors in this sample (82%) measure learning through formative and summative assessments, while 12% assign only summative assessments and 6% assign only formative assessments. When asked to report the primary method they use to monitor ongoing learning, 68% reported in-class assessments, while 15% replied on homework assignments, 9% on in-class activities, 6% on online assessments, and 3% on work in a learning platform.

Student

To ensure we engaged a representative student sample, self-reported demographic, experiential, and perception information were provided on the baseline and end of semester student surveys and is presented in Appendix Table A2. The largest proportion of students in this sample were enrolled in Economics courses (41%), followed by Chemistry (37%), with the smallest proportions in Calculus and Biology (each 11%). Most (59%) were first year students, though a moderate proportion (20%) were second year students and the average age of a student in the sample was 19.8. The majority (54%) of students in this sample were female students and most (73%) were taking the course as a discipline requirement. A large majority (86%) reported being either comfortable or extremely comfortable with technology and 90% reported that publisher provide tools could support their learning. The average HSGPA reported was 3.70 and 72% of students who reported having taken the SAT and/or ACT were classified as "college ready" (i.e. met or exceeded the college readiness benchmark on each of the sections of either the SAT or ACT). Based on a scale completed on the pre-survey most students (61%) were coded as less motivated to succeed in the course.

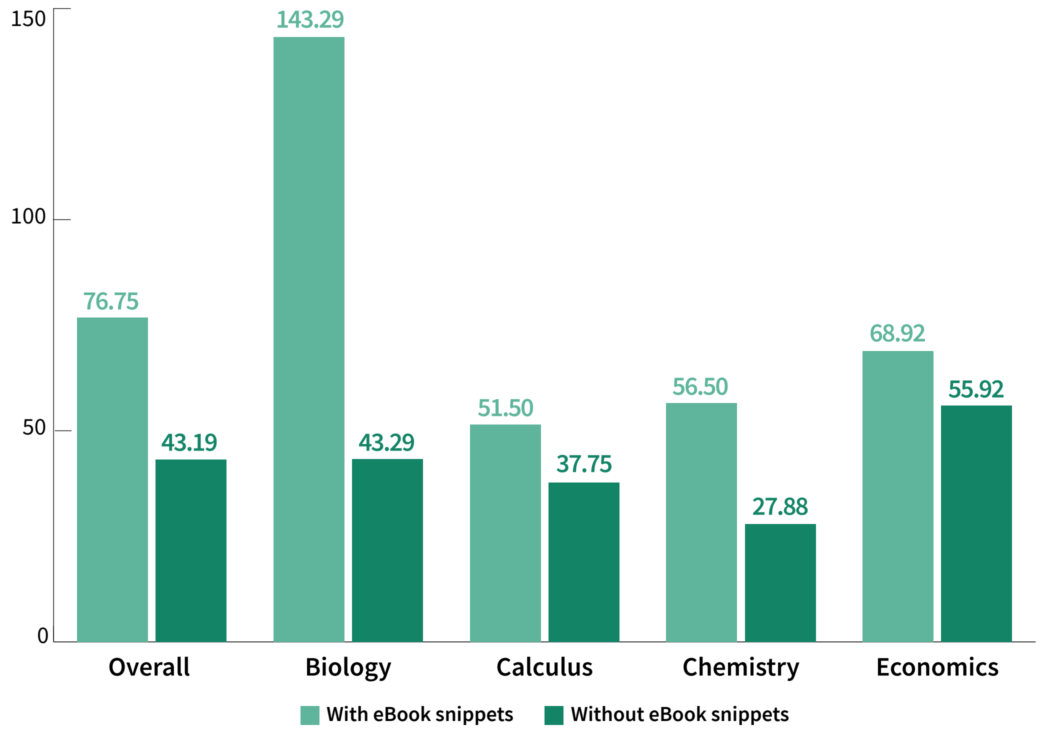
Results

RESEARCH QUESTION 1

How did instructors implement Achieve and how are students engaging with it? Did implementation and/or use case vary by discipline or educational context?

We hypothesized that instructors would implement Achieve regularly throughout the semester. Because of the robustness of the program, we further hypothesized that instructors would assign activities from Achieve to be completed before and after the class where the content would be lectured on. We expected that instructors would choose to apply the tools in Achieve that were designed to be implemented "pre-class" to transfer direct knowledge before students came to class, thus enabling higher order cognition and active learning during class time. We then expected instructors to assign formative homework in Achieve to reinforce concepts. We hypothesized that instructors would assign "in-class" activities (i.e. activity guides or case studies) at a lesser extent because they were only available in an average of four chapters in the disciplines studied. Lastly, we expected to see some implementation of summative assessments, but because Achieve was primarily developed as a formative tool, we hypothesized that the majority of summative assessments would be developed by the instructor and delivered in class. Since Achieve offers instructors analytic reports of student engagement and performance, we expected that they would use those insights to adapt or focus their in-class lectures and/or planned activities. Among students, we hypothesized that they would use Achieve at a high rate because of its engaging features, the wealth of feedback they receive, and because prior research conducted by the authors suggested that students had positive perceptions of Achieve and believed that it supported their academic performance in Biology, Calculus, Chemistry and Economics. We did not hypothesize about implementation or engagement by educational context as we had no prior research to base a theory, so those analyses are meant as exploratory.

Graph 1. Average number of activities assigned for credit in Achieve



During the Spring 2019 semester, instructors in the sample assigned an average of 76.75 assignments within Achieve, with the proportion of assignments ranging from 10 to 336. It is important to consider that readings are assignment in chunks (or "snippets") for consumability, so assignment of eBook snippets inflates the average assignment rate. When calculated without considering eBook snippets, on average, instructors assigned 43.2 activities (range 10 to 134). There was variability in average rate of assignment by discipline which can be seen in Graph 1.

On average, instructors assigned Achieve activities in 92% of the weeks in the semester, and activity assignment by individual instructors ranged from 36% to 100% of weeks in the semester. Instructors who assigned Achieve regularly reported that the only times they did not assign Achieve were either weeks in which classes were not held or when they were experiencing technical challenges with Achieve (reported by four instructors in one week each, due to an update that was released).

The platform usage data were mined to examine empirical implementation patterns and survey and interview data were used to contextualize those use cases. First, each asset that had been assigned was coded to establish consistency across disciplines. The variables housed in the platform of "tool name" (e.g assessment, reading, test, etc.), "assignment classification" (e.g. pre-class tutorial, Learning Curve, etc.), "assignment name" (e.g. EconEd in-class activity, Chapter 4 test, etc.), and "learning path" (e.g. pre-class, in-class, post-class) were used to develop the algorithm to classify an activity as either (1) reading (2) diagnostic (3) pre-class formative assessment, (4) in-class activity, (5) post-class formative assessment or (6) post-class summative assessment. Assignment types are defined by:

1. **Reading.** Any reading "snippet" assigned.

2. **Diagnostic.** Any assignment that was assigned prior to the content being taught and that provided a pre-test, study plan, and post-test with associated diagnostic information (i.e Pathfinder diagnostic activities).

3. **Pre-class formative assessment.** Any assessment where students were given more than one opportunity to answer a question correctly or where the goal was persistence to completion and when the assignment was due prior to the class in which the content would be introduced. These activities tended to be video tutorials and LearningCurve activities.

4. **In-class activity.** Any activity that was assigned to be completed during in-class time. These activities tended to be case studies, current event assignments, and iClicker student response system questions.

5. **Post-class formative assessment.** Any assignment where students were given more than one opportunity to answer a question correctly or where the goals was persistence to completion and when the assignment was due after the class in which the content was introduced. These activities tended to be end of chapter problems, data analysis activities, etc.

6. **Post-class summative assessment.** Any assignment where students were given one attempt to answer a question correctly and there were a discrete number of items presented to a student, or the activity was weighted more heavily in a student's grade. These activities tended to be quizzes, tests, and homework.

Patterns of activity assignment were examined descriptively and we noted that instructors assigned activities in a cadence. We then examined the proportion of weeks in their semester that each instructor assigned a specific asset type and coded them as a type of activity assigner. An instructor was considered an activity assigner if the activity type was assigned to be due in at least 10% of the active weeks (not including break weeks) during their semester. For example, if an individual instructor's semester was 19 weeks long with one inactive week for Spring break, they had to have assigned an activity type in at least 16 weeks that semester to be considered an assigner. The decision was made because we knew from qualitative data that in ten percent of the weeks Achieve was not used because of in-class assessments.

Table 3. Number of instructor in sample by implementation pattern and discipline

	Post-class summative	Pre-class formative and some post-class	Pre-class formative and all post-class	Pre-class, in-class, and all post-class
Biology (n=7)	0	3	4	0
Calculus (n=8)	3	3	1	1
Chemistry (n=8)	2	5	0	1
Economics (n=11)	1	4	6	0
Total (n=35)	6 (17%)	15 (43%)	11 (31%)	2 (6%)

Note: one Calculus instructor (3% of sample) was an outlier and is not represented in these patterns, they assigned chunks of Achieve immediately before assessments and their data were analyzed separately.

Assignment patterns loaded on a set of implementations. However, the inclusion of reading as a separate activity type conflated the implementation patterns. In total, 21 instructors assigned any reading snippets to be completed, but of those instructors there was only an average of 3.4 snippets assigned per week (range 0.25 to 18.12) translating to only about four paragraphs of assigned reading per week. We explored qualitatively with instructors why the assignment rate or reading was so low and learned that either (1) instructors believed that students got the information from other activities or (2) readings had to be assigned for points and instructors did not want readings to count toward a student's grade so they did not assign them but did make the snippets available to students. Consequently, we explored the rate at which students access unassigned readings and found that only 13% of unassigned reading snippets were accessed. Given the low reading rate, we concluded that readings did not substantively change the implementation patterns so removed them from the categorization algorithm.

Four meaningful implementation patterns emerged empirically and were validated qualitatively.

1. **Post-class summative.** Instructors who assigned only post-class summative activities
2. **Pre-class formative and some post-class.** Instructors who assigned pre-class formative activities and either post-class formative activities or post-class summative activities.
3. **Pre-class formative and all post-class.** Instructors who assigned pre-class formative activities and post-class formative and post-class summative.
4. **Pre-class, in-class, post-class.** Instructors who assigned pre-class formative activities, in-class activities, post-class formative activities, and post-class summative activities.

Table 4. Proportion of final course grade each Achieve activity type accounted for

Activity type	# of Instructors	% Final course grade
Pre-class tutorial	12	6.5
Pre-class videos	6	5.7
Adaptive reading quizzes (LearningCurve)	9	40.4
Homework	11	19.9
Quiz	8	17.4
Test	5	23

The largest proportion of instructors in this sample (43%) assigned pre-class activities and some post-class (either formative or summative) while the next largest proportion (31%) assigned pre-class and post-class formative and summative activities. Our hypothesis that few instructors would assign post-class summative assessments in Achieve could not be confirmed due to the high assignment rate regardless of the lower availability. These findings suggest that instructors do have a desire to use assessments in Achieve as summative measures, indicating that content development teams should populate more high quality summative assessments in the platform across disciplines. The Calculus instructor with the most highly structured course noted on the post-survey, “The concept of pre-class in-class and post-class assignments is very well supported in math pedagogy. There are real benefits to using this model, but it has to be done right.”

Our hypothesis that very few instructors would emerge as having the most highly structured course in Achieve (pre-class, in-class, and post-class) was confirmed when only two instructors were coded that way. Interviews across all instructors revealed three important findings to this end: (1) three instructors did utilize the in-class activities that were available but did not assign them for points, which is why they did not emerge in the algorithm. However, they were not re-coded as in-class activity assignors because they noted only having used them in one of two weeks of the semester (2) instructors requested more in-class activities be made available in Achieve and (3) at the time of this study iClicker was not yet integrated in Achieve which instructors noted would influence greater use of the iClicker questions available in Achieve.

Instructors were asked to report the proportion of a student’s final course grade that Achieve would account for. In total 18 instructors in the sample combined all activity in Achieve to account for 25% of a student’s grade, on average (SD = 14.8). And, 13 instructors included activity averages separately. Table 4 presents the number of instructors who counted that activity average as part of a student’s final course grade, and the average proportion is accounted for.

Table 5. Number of instructors in sample by implementation pattern and institutional categorization

	% Post-class summative assigners	% Pre-class formative and some post-class assigners	% Pre-class formative and all post-class assigners	% Pre-class, in-class, and all post-class assigners
Institution type				
Two-year (n=13)	23	31	31	8
Four-year (n=22)	14	50	32	4
Institution size				
Large (n=20)	30	45	20	5
Medium (n=9)	0	44	44	0
Small (n=6)	0	33	50	17
Acceptance rate				
Highly selective (n=4)	0	25	50	25
Moderately-selective (n=14)	14	57	29	0
Less selective (n=3)	0	67	33	0
Open enrollment (n=13)	31	31	31	9

Note: outlier implementation pattern is not represented, instructed at a medium-sized, two-year, open enrollment institution.

It was also important to evaluate whether use cases varied by educational context, so we disaggregated the implementation patterns by institution type, size, and selectivity (size and selectivity categorized by the Carnegie classification). These results can be found in Table 5.

When considered by institution-type, instructors assigning pre-class formative and either formative or summative post-class activities were overrepresented in four-year institutions, and instructors only assigning summative assessments were slightly overrepresented in two-year institutions. Only instructors at large institutions fell into the "post-class summative only" category and instructors at small institutions were more likely to have the most highly structured course assigning pre-class, in-class, and post-class formative and summative activities. Instructors at highly selective institutions were more likely to have structured courses, assigning pre-class activities in every case as were instructors at less selective institutions. Instructors assigning only summative assessments were overrepresented in open enrollment institutions.

Table 6. Implementation patterns by instructor characteristics

	% Post-class summative assigners	% Pre-class formative and some post-class assigners	% Pre-class formative and all post-class assigners	% Pre-class, in-class, and all post-class assigners
Years teaching				
1-5 years (n=4)	25	75	0	0
6-10 years (n=10)	30	0	70	0
11-15 years (n=8)	0	63	25	13
More than 15 years (n=14)	14	50	21	7
Comfort with educational technology				
Extremely uncomfortable (n=1)	0	0	100	0
Uncomfortable (n=6)	17	33	33	17
Comfortable (n=13)	23	46	23	8
Extremely comfortable (n=15)	13	47	33	0
Agreement that publisher provided digital learning tools enhance pedagogy				
Agree (n=21)	19	33	38	5
Strongly agree (n=14)	14	57	21	7
Used a published provide learning tool last time they taught this course?				
Yes (n=23)	9	52	30	4
No (n=12)	33	25	33	9

Note: the outlier implementation pattern is not represented, instructor had been teaching more than 15 years, was extremely comfortable with educational technology, agreed it could enhance their pedagogy, and did use a publisher provide tool the last time they taught this course.

Finally, we wanted to examine implementation pattern by years teaching, comfort with educational technology, perception of publisher provided digital learning tools, and experience with publisher provided digital learning tools to explore whether these factors influenced how an instructor chose to use Achieve. These results are presented in Table 6.

Instructors newer to their role tended to have less structured implementation patterns, while those who had been teaching longer had more highly structured courses, but the majority of instructors in most ranges tended to implement pre-class and either formative or summative post-class activities. An instructor's reported level of comfort with technology did not appear to influence their implementation pattern nor did their perception of whether publisher provided digital learning tools could enhance their pedagogy (note the high perceptions across the sample). Finally, whether an instructor had used a publisher provided digital learning tool did not appear to influence an instructor's chosen implementation pattern.

Survey and interview data provided contextual information to provide further insight into why instructors chose to implement Achieve in the way that they did. A summary of qualitative responses by implementation pattern are provided below.

Post-class summative only. Three of the six instructors who chose to assign only post-class summative assessments reported doing so primarily because they were using Achieve as a homework system — the same way they had used their previous publisher provide platform. They noted that the previous use case was fine so they continued. When asked what would help them consider assigning other assets, they all noted that some evidence that the other assets would support student success would help. One noted that they would need to know that adding additional activities to their implementation pattern wouldn't create an unreasonable amount of work for students. Two reported that they did not realize there were pre-class activities available in Achieve.

Pre-class formative and post-class assessment (either formative or summative). Nine of the 15 instructors who assigned pre-class formative assessments and some post-class activity indicated they selected that pattern because they felt the pre-class formative assessment helped their students come to class prepared to participate and engage more in class and the post-class activities helped reinforce the concepts discussed in class. Three instructors noted that they had used this structured model before and believed it supported performance in the course, and three indicated that this implementation pattern was suggested to them during training and that they otherwise would not have assigned pre-class activities.

Pre-class formative, post-class formative, and post-class summative. All eleven instructors who used this implementation pattern reported doing so because pre-class formative activities helped them come to class prepared and engaged, post-class formative activities reinforced concepts, and post-class summative assessments helped them measure student comprehension. When these instructors were asked why they did not implement in-class activities they reported that they either used their own in-class activities, that they did not know in-class activities were available in Achieve, or that they did use them one or two times but did not assign them so they did not emerge in the empirical data.

Pre-class formative, in-class, post-class formative, post-class summative. The two instructors who adopted the most highly structured implementation pattern reported doing so because it provided students with a complete cycle of learning. They both reported that they have wanted to implement in-class activities but have been unable to because of time constraints. But, that the pre-class formative reduced the amount of direct instruction time needed in the class and that the activities and iClicker questions in Achieve were high quality so they didn't have to create their own. One instructor reported that they had used this highly structured course before but had to use resources from a variety of places — rather than them all being housed in one system. The other instructor reported this being the first time they used this highly structured of a course but that they spent a good amount of time exploring the system and experimenting with various activities, perceived effectiveness early and continued implementing them.

Table 7. Student engagement rates by activity type and discipline

	% Overall	% eBook	% Pre-class formative	% In-class	% Post-class formative	% Post-class Summative
Overall (n=1,623)	80.3	100	85.6	46.5	71.2	74.5
Biology (n=159)	83.0	100	84.6	78.6	83.4	84.7
Calculus (n=203)	85.2	72.3	88.4	47.4	94.7	84.1
Chemistry (n=655)	73.4	100	86.1	NA	38.8	57.3
Economics (n=642)	84.7	100	85.1	36.0	81.5	85.6

Note: "engagement" is defined by a student launching an activity

Data insights

Instructors were presented with data analytics that provided insights in reports at the course, unit, and learning objective level. The majority of instructors in this sample (n=25, 71%) accessed the reports in Achieve in at least 90% of the active weeks of their semester during which reports were available (reports were released after the start of some instructors' courses). Each week instructors were asked in their implementation logs whether they took action based on the insights provided in the reports. In total, 24 of the 25 instructors who had accessed dashboards regularly took action based on the insights provided in at least one week of their semester. Trends in those data suggest that instructors did not begin taking action based on insights available until their students were approaching the midterm, however once students passed midterms instructors began to take action based on the dashboards. The primary action that instructors reported taking (92%) were modifying their class lecture based on what students were struggling with or what they had already mastered. A smaller proportion, but half (50%) reported intervening with specific students based on the student-level insights provided.

An evaluation of the survey and interview data provided insight into dashboard patterns. In total, 64% of instructors either agreed or strongly agreed that the reports were easily understood and the same 64% either agreed or strongly agreed that they understood from where the information in the reports was derived. When instructors who did not regularly access the dashboards were asked why they didn't they reported that they had looked at them but it appeared that they would have to spend a great deal of time analyzing the data to understand how to use it. The instructor who accessed the dashboard but did not take action reported it was because the data were not easily understood. The combination of the survey results and the interview data led to more systematic testing of the reports and ultimately a redesign and deployment of reports optimized based on participant feedback in the Fall 2019 semester.

Student usage

Relative to the research literature on average student engagement in digital learning tools, the students in this sample engaged in assigned activities in Achieve at a high rate. Average total engagement was calculated by dividing the total number of activities engaged in over the total number of activities assigned and the overall total engagement rate was 80.3%. Individual student overall total engagement rates ranged from 2.4% to 100%. Table 7 presents the student engagement rates overall and by activity type and discipline.

Table 8. Student completion rates by activity type and discipline

	% Overall	% eBook	% Pre-class formative	% In-class	% Post-class formative	% Post-class Summative
Overall (n=1,621)	73.2	100	93.3	61.2	73.8	72.1
Biology (n=159)	80.6	100	95.8	100.0	92.2	83.7
Calculus (n=203)	60.6	100	74.0	54.8	61.8	64.2
Chemistry (n=655)	65.7	100	93.9	67.0	42.2	62.7
Economics (n=642)	82.7	100	92.1	36.0	96.1	82.2

Completion rates were also high relative to the research literature on completion of activities in digital learning tools. Average total completion was calculated by dividing the total number of activities completed over the total number of activities assigned and engaged in and the overall total completion rate was 73.2%. Individual student overall total completion rates ranged from 2.4% to 100%. Table 8 presents the student completion rates overall and by activity type and discipline.

Among activity types, overall and in each discipline, students had the highest engagement rate in pre-class formative activities. This was not surprising given the survey and qualitative feedback receive around these activities. On the post-survey students were asked to report their perceptions of pre-class activities in Achieve by rating their level of agreement with five statements on a scale of 1 = "strongly disagree" through 4 = "strongly agree." On average students agreed that pre-class activities helped them stay on track with reading (Mean = 2.96, SD = 0.71); that they gave them a basic understanding of what would be covered in class (M=3.08, SD = 0.68); that pre-class activities helped them achieve a basic understanding of concepts (M=3.05, SD=0.67); that they helped them actively learn in class (Mean = 2.92. SD = 0.72); and that pre-class activities helped them participate more than they normally would in this course (Mean = 2.74, SD = 0.75).

ENGAGEMENT AND COMPLETION BY STUDENT SUBGROUPS

To understand whether students are varying levels of motivation and academic preparedness were engaging and completing activities in Achieve at varying rates, the data were disaggregated by these subgroups and compared.

Students were asked to report on the baseline survey whether they had taken the SAT and/or ACT, if they had they were asked to report their section scores. A college readiness variable was created by coding all students who met or exceeded the college readiness benchmarks (based on SAT and ACT concordance work) as "college ready" and all students who fell short of at least one benchmark as "not college ready." There was a significant difference in the total engagement rates between students coded as "not college ready" (M=78%, SD=20%) and "college ready" (M=83%, SD=19%); $t(1,263)=-4.07$, $p < .0001$. There was also a significant difference in completion rates, "not college ready" (M=69%, SD=22%) and "college ready" (M=76%, SD=22%); $t(1,262)=-5.16$, $p < .0001$.

Table 9. Differences in student engagement rates by activity type, discipline, and college readiness

	% Overall		% eBook		% Pre-class formative		% In-class		% Post-class formative		% Post-class Summative	
	CR	NCR	CR	NCR	CR	NCR	CR	NCR	CR	NCR	CR	NCR
Overall (n=1,621)	83	78	100	100	89	87	35	62	78	63	80	70
Biology (n=159)	88	82	100	100	88	86	86	71	88	85	89	82
Calculus (n=203)	85	88	74	87	85	97	51	59	95	85	83	88
Chemistry (n=655)	72	74	100	100	86	88	58	60	37	37	56	58
Economics (n=642)	87	79	100	100	88	76	23	63	84	76	88	82

Note: boxes colored dark green are statistically significantly different with higher rates among students coded "not college ready" and boxes colored light green are statistically significantly different with higher rates among students coded "college ready"

Differences in engagement rates were examined by discipline and activity type by baseline level of college readiness, results can be seen in table 9. Results suggest that the overall differences are being influenced by differences within discipline. For example, the low overall implementation of in-class activities and the large difference in rates in Economics have influenced the overall differences. The same is likely true of post-class formative in Calculus and post-class summative in Economics.

The results suggest that students less and more academically prepared engage in all activity types in Achieve for Biology and Achieve for Chemistry at the same rates. And, that it might be worth considering including supports in post-class formative activities for less academically prepared students in Calculus courses and Economics courses.

> "The best part about Achieve is that there are so many different ways to get students engaged in the course material"

Table 10. Differences in student engagement rates by activity type, discipline, and baseline level of motivation

	% Overall		% eBook		% Pre-class formative		% In-class		% Post-class formative		% Post-class Summative	
	More	**Less**	**More**	**Less**	**More**	**Less**	**More**	**Less**	**More**	**Less**	**More**	**Less**
Overall (n=1,621)	81	78	100	100	89	85	39	33	73	65	76	74
Biology (n=159)	83	86	100	100	85	86	75	100	83	87	85	91
Calculus (n=203)	84	71	69	62	86	77	43	39	96	58	83	72
Chemistry (n=655)	71	74	100	100	87	84	23	57	40	36	59	54
Economics (n=642)	87	82	100	100	87	82	35	43	84	82	88	84

Note: boxes colored green are statistically significantly different with higher rates among students coded "college ready"

Differences in total engagement rates by baseline level of motivation were also examined. Students were asked to respond to a set of questions that, when combined enabled a coding of "less motivated to succeed" and "more motivated to succeed." There was a small, but significant difference in the total engagement rates between students coded as "less motivated to succeed" (M=78%, SD=21%) and "more motivated to succeed" (M=81%, SD=19%); t(1,471)=2.84, p=0.0046). There was no significant difference between in completion rates between "less motivated" (M=71%, SD=24%) and "more motivated" (M=74%, SD=22%); t(1,469), p=0.1064).

Interestingly, as shown in Table 10, when examined by activity type, specific differences by discipline influenced the overall difference. For example, in Biology and Chemistry, there are no differences in engagement rates in any activity type, and in Calculus the differences are in post-class activities only, while in Economics there is a significant difference in pre-class formative activity engagement rate and post-class summative activity engagement rate.

The results suggest that students less and more motivated to succeed engage in all activity types in Achieve for Biology and Achieve for Chemistry at the same rates. And, that it might be worth exploring including nudges for students in Economics classes less motivated to succeed during pre-class activities. For example, comment boxes that read *"don't give up — research shows that for every 10% increase in questions completed, students earn a higher final exam score!"* The Product and User Experience teams have implemented these data as part of their ongoing optimization of Achieve efforts.

RESEARCH QUESTION 2

What are instructor and student perceptions of Achieve?

Instructor perceptions

In general, instructors had moderately high perceptions of Achieve. Important context is that this study was conducted in the first of three semesters of beta testing, thus Achieve was still in a formative state. The results that follow should be interpreted based on Achieve's stage in the development lifecycle when this study occurred.

Instructors were asked to rate on a scale from 1 = "Would definitely not recommend" to 10 = "Would definitely recommend," the extent to which they would recommend Achieve to a colleague based on their experience with it this semester. The average rating of likelihood to recommend was 6.8. Instructors were also asked to rate on a scale from 1 = "Will definitely not adopt" to 10 = "Will definitely adopt" based on their experience with Achieve this semester how likely they were to adopt Achieve when it is available for adoption in Fall 2020 (if the price was reasonable and it was approved by their department). The average rating of likelihood of adoption was 6.6. We further examined these ratings by discipline, years teaching, perception of educational technology, and experience with publisher provided digital learning tools. Results are presented in Table 11.

Instructors teaching Chemistry had the strongest positive perception of Achieve (likelihood to recommend = 8.1), with Economics instructors reporting similarly positive perceptions (likelihood to recommend (Mean = 7.2). Calculus instructors had the lowest rating of likelihood to recommend (Mean = 5.5). Qualitative data suggest that this rating is due to the ongoing development of math modules that created technical challenges for some students when they were submitting their assignments. These features have since been developed and higher ratings are expected in the update to this efficacy report following the Fall 2019 semester.

Table 11. Average rating of likelihood of recommendation and adoption by discipline and instructor characteristics

	Likelihood to recommend	Likelihood of adoption
Overall	6.8	6.6
Discipline		
Biology (=7)	6.3	6.3
Calculus (n=9)	5.5	5.3
Chemistry (n=8)	8.1	7.4
Economics (n=11)	7.2	7.2
Years teaching		
1-5 years (n=4)	6.8	6.0
6-10 years (n=9)	7.1	6.7
11-15 years (n=8)	6.1	6.3
More than 15 years (n=14)	7.0	6.8
Comfort with technology		
Extremely uncomfortable	3.0	3.0
Uncomfortable	6.7	6.0
Comfortable	7.0	7.0
Extremely comfortable	6.9.	6.7
Digital learning tools enhance pedagogy		
Agree (n=21)	6.3	5.9
Strongly agree (n=14)	7.8	7.8
Used a published provide learning tool last time they taught this course?		
Yes (n=23)	6.8	6.6
No (n=12)	6.7	6.5

Table 12. Differences in instructor ratings between Achieve and their previous approach to teaching their course

	Previous approach	Achieve	Difference	Significance
Easy to set up and use	2.89	2.76	-.12	t(32), p=0.423
Saved time in the classroom	2.97	2.76	-.18	t(32), p=0.161
Enhanced my pedagogical framework	3.09	3.06	0	t(32), p=1.000
Flexible enough to meet my pedagogical needs	2.94	2.85	-.09	t(32), p=0.4470
Offered me the data and analytics needed to inform instruction	2.91	2.73	-.21	t(32), p=0.1285
Helped me understand where my students had content gaps	2.06	2.39	+.39*	t(32), p=0.0072
Promoted active learning in the classroom	2.89	2.83	-.06	t(32), p=0.6441

Note:
1 = strongly disagree 3 = agree
2 = disagree 4 = strongly agree

Note: *represents a statistically significant difference at the .05 level.

Years teaching did not appear to influence instructor perceptions, nor did comfort with technology with the exception of the outlier instructor that reported a likelihood of 3. Interviews with this instructor provided context for this rating. The instructor noted *"I don't think I was comfortable enough with technology to participate in a beta, but I look forward to participating next semester when Achieve is more fully developed."* Whether an instructor had used a digital learning tool the last time they taught this course did not influence their likelihood to recommend Achieve, suggesting that instructors both new and veteran to using digital learning tools had moderate to high perceptions of Achieve.

Instructor comparison to current approach

In an effort to measure the extent to which Achieve is providing an experience that enhances what instructors are currently experiencing, we asked instructors on the baseline survey to rate, on a scale of 1 = "strongly disagree," 2 = "disagree", 3 = "agree," 4 = "strongly agree" the extent to which they agreed with a set of statements about their current approach to teaching their course. At the end of the semester we asked instructors to rate, on the same scale, the extent to which they agree with the same statements about Achieve. Our hypothesis was that they would rate the statements higher when evaluating Achieve as compared to their current approach to teaching their course. Table 12 presents the average ratings on the pre-survey, the average rating on the post-survey, the difference, and whether the difference is statistically significant.

Table 13. Differences in instructor ratings between Achieve and their previous approach to teaching their course by use of digital

	No digital			Digital		
	Previous approach	Achieve	Diff	Previous approach	Achieve	Diff
Easy to set up and use	3.14	2.69	-0.45	2.73	2.64	-0.09
Saved time in the class-room	3.00	2.77	-0.23	3.03	2.64	-0.39*
Enhanced my pedagogical framework	3.14	3.00	-0.14	3.16	3.00	-0.16
Flexible enough to meet my pedagogical needs	2.79	2.92	0.13	3.00	2.72	-0.28*
Offered me the data and analytics needed to inform instruction	2.02	2.98	0.96*	2.80	2.72	-0.08
Helped me understand where my students had content gaps	2.50	2.93	-0.43*	2.73	2.80	+0.07
Promoted active learning in the classroom	2.74	2.72	-0.02	2.90	2.86	-0.04

Note:
1 = strongly disagree 3 = agree
2 = disagree 4 = strongly agree

Note: *represents a statistically significant difference at the .05 level.

As presented in the table, none of the differences between the average rating of perception of current approach and use of Achieve were meaningfully different. Here, again, it is important to note that this study was conducted in Achieve's first semester in beta. Given that context the non-significant findings were positive.

Note that these differences are calculated at the aggregate, that is instructors who did not use a digital learning tool before using Achieve are included in these analyses. Table 13 disaggregates instructors who previously used a digital learning tool and those who didn't.

Making instructors' lives easier

One important intended outcome of Achieve is to make instructors' lives easier. In an effort to measure the extent to which Achieve is accomplishing that, at the beginning of the semester we asked instructors who rate, on a scale of 1 = "very difficult," 2 = "difficult," 3 = "easy," 4 = "very easy" how difficult a set of activities were in their course the last time that they taught it. At the end of the semester we asked instructors to rate, on the same scale, how difficult the same set of activities were in their course this semester.

Table 14. Instructor perceptions of difficulties in their course with and without Achieve

How difficult is:	Without Achieve	With Achieve	Difference	Significance
Implementing active learning strategies	2.51	3.36	+.88**	t(32), p<0.001
Assessing how well students are comprehending material	2.49	3.33	+.85**	t(32), p<0.001
Promoting students coming to class prepared to participate	2.43	3.21	+.82**	t(32), p=0.001
Fostering ability to remember information	2.34	2.88	+.60**	t(32), p=0.001
Promoting student collaboration	2.69	3.15	+.54*	t(32), p=0.007
Fostering deep insights	2.06	2.39	+.39*	t(32), p=0.007

Note:
1 = strongly disagree 3 = agree
2 = disagree 4 = strongly agree

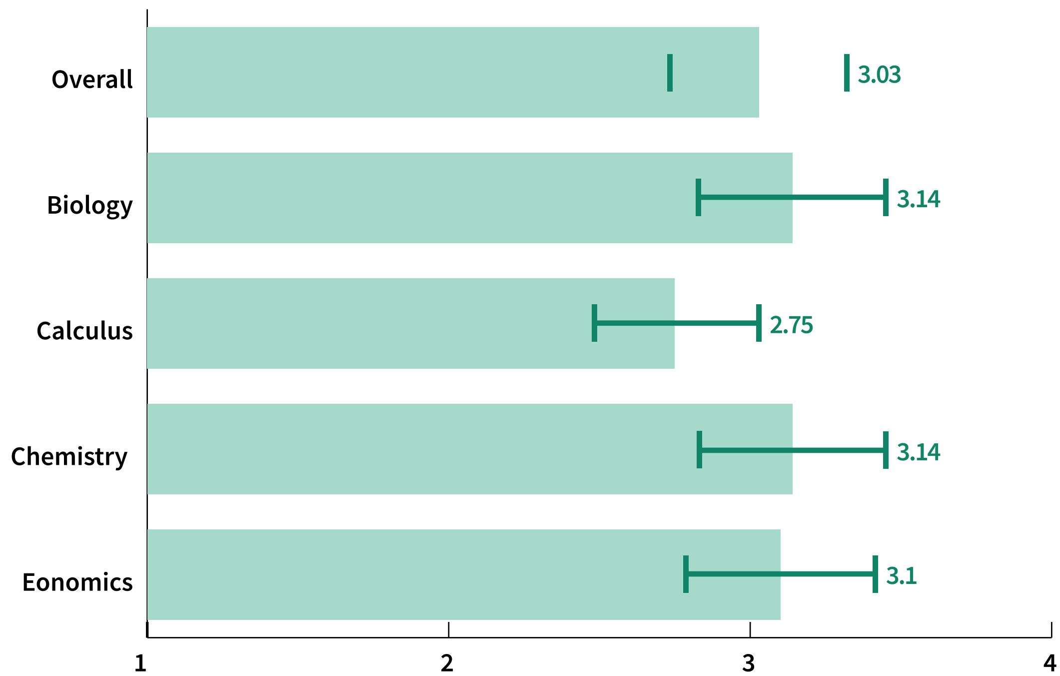
Note: **represents a statistically significant difference at the .001 level and *represents a statistically significant difference at the .05 level.

Our hypothesis was that the activities would be less difficult for instructors when they implemented Achieve in their course. Table 14 presents the average ratings on the pre-survey, the average rating on the post-survey, the difference, and whether the difference is statistically significant.

In all cases instructors rated the behaviors that Achieve was designed to support less difficult, on average, during the semester that they were using Achieve as compared to the last semester that they were teaching the course. The largest significant decrease in difficulty (.88) rating emerged when instructors were asked to rate how difficult it was to implement active learning strategies in the classroom. The difference of .88 on a four-point scale represents a near full average change in rating of difficulty. This finding is interesting in that only a portion of the active learning capabilities had been built into Achieve at the time of this study, but given that Achieve was built on the foundation of the active learning model, this finding offered strong validity evidence for the learning science built into Achieve.

The next largest significant decrease (.85) emerged when instructors were asked to rate how difficult it was to assess how well students were comprehending the material. The various insights that Achieve enables through student analytic reports was designed to offer actionable insights into student comprehension, therefore this difference is a strong piece of validity evidence. Instructors who accessed the dashboards had positive perceptions of them, one Biology instructor commented, *"I enjoyed the ability to quickly see how students were performing in the course. This allowed me to quickly modify lectures or activities in class to help address these issues."*

Graph 2. Average rating of perception that Achieve supports mastery more than if Achieve was not used by discipline



The third largest change emerged when instructors were asked how difficult it was to promote students coming to class prepared to participate. One Biology instructor noted, "*Students completing pre-class assignments meant they were coming to class much more prepared than in previous semsters.*" Interestingly, when examined by implementation pattern, instructors who implemented Achieve post-class summative only, did not have a statistically significantly different perception, on average, of the level of difficulty of promoting students coming to class prepared to participate. This finding suggests that a more highly structured course (with pre-lecture assignments) better supports students coming to class prepared to participate.

INSTRUCTOR PERCEPTION OF SUPPORTING STUDENT MASTERY

The ultimate goal of use of Achieve is to support student mastery in higher education courses. Instructors were asked to rate (scale 1 = "strongly disagree" to 4 = "strongly agree") the extent to which they agreed that using Achieve supported their students gaining mastery of the course content more than if they had not used Achieve. In total, 88% of the instructors in this sample either agreed or strongly agreed that Achieve supported mastery more than if Achieve had not been used (M=3.03, SD = 0.54). Average ratings by discipline can be found in Graph 2.

Most challenging and best experiences with Achieve

Instructors were asked to comment, in an open response item on the post-survey, "*what, if anything, was the best part about using Achieve in your course this semester?"* The responses were meant to support ongoing development efforts. In total, 32 of the 35 respondents left a comment. Of the 32 responses, four instructors (12.5%) commented that the best part was that they could see Achieve supported student mastery. Another four instructors (12.5%) reported the best part of Achieve was that students were coming to class prepared to participate as a result of the readings and pre-class assignments. Relatedly, two instructors (6%) noted the benefits of the structured course (i.e pre-, in-, and post-class) and three (9%) noted that students were actually doing the reading before class — unlike they had observed in other semesters. Three instructors (9%) valued the feedback students received while working through activities, and three other instructors reported that the best part was the breadth of resources available in Achieve. Two instructors (6%) commented that the best part of Achieve were the actionable insights provided in the analytics reports enabling. One instructor each reported that the best part was: ease of use, efficiency, flexibility, instructor resources, student resources, adaptive reading quizzes, varied item types, and support by Macmillan staff.

Another open response item on the post-survey asked instructors to report what, if anything, was the most challenging part of using Achieve this semester. Responses were meant to support remediation, optimization, and development roadmap decisions while Achieve was still in beta. In total, 32 of the 35 respondents left a comment.

> "In addition to increasing efficiency, students seemed to really like the system as I had a significantly higher retention rate in this class as compared to the same class in previous semesters and my other macroeconomics classes this semester"

Of the 32 responses, nine (28%) listed specific "bugs" in the system, for example a Chemistry instructor reported, *"When I made a change to an activity such as due date or when it would be opened, I had to wait a long time after submitting the change, and sometimes the details of the activity would be changed to something different from what I entered."* All nine reported bugs have been remediated in releases that came out post-data collection. Six instructors (19%) commented that the most challenging part was getting to know a new system, but after a few weeks that was no longer an issue. For example, *"The most challenging was the difference in the layout compared to Mastering Chemistry and WileyPlus. However, Achieve was fairly easy to navigate and learn from any errors."* Four instructors (12.5%) commented that the most challenging part of using Achieve was the usability and user experience of

Table 15. Average student adapted Net Promoter Score rating overall and by subgroup

		Net Promoter Score
Overall		6.83
Discipline		
	Biology	6.77
	Calculus	6.60
	Chemistry	6.88
	Economics	6.87
Taking course for major		
	Yes	6.81
	No	6.83
Level of academic preparedness*		
	More prepared	6.96
	Less prepared	6.53
Level of motivation		
	More motivated	6.96
	Less motivated	6.74

Note: Scale is 0-10
Note: * indicates a statistically significant difference

the gradebook. Each of these instructors were interviewed about how the gradebook could be optimized and their feedback was implemented, tested, and released for the Fall 2019 semester. Another four instructors each provided specific feedback about navigation challenges they experienced which the User Experience team began conducting deeper research into. Another four instructors (12.5%) reported the most challenging part of using Achieve being that it was in beta form so the system was not yet as robust as it will be when fully developed. Instructors noted gaps that they look forward to seeing when Acheive is fully developed like section management and batch assign (features which have since been released to Achieve). Two instructors (6%) offered feature requests for homework assignment management which have been added to the product roadmap. And one instructor found course set up to be the most challenging part of using Achieve, though based on the specific feedback the expectation is that when the instructor participates in the replication study the challenges will be curbed by the advances that have been made to the course set up process in Achieve.

Student perceptions

Students also had moderately high perceptions of Achieve, one question on the post-survey asked students to rate, on a scale of 0 to 10, their likelihood to recommend the same course to a friend if they knew that Achieve was going to be used. Average student rating waw 6.83 out of 10. It was important to examine the differences in ratings among subgroups of students to understand whether some groups have systematically higher or lower perceptions of Achieve, and why. These rtesults can be found in Table 15.

Table 16. Average student rating of likelihood to recommend course to a friend if Achieve was being used, by instructor implementation pattern

Implementation model		Average rating
Model 0	Instructor who only used Achieve as review (n=13)	3.67 (SD=1.97)
Model 1	Model Homework only (n=84)	7.63 (SD =2.08)
Model 2	Pre-class + formative or summative assessment (n=666)	6.69 (SD =2.29)
Model 3	Pre-class + post-class formative assessment + post-class summative assessment (n=431)	6.90 (SD = 2.02)
Model 4	Pre-class + in-class + post-class formative assessment + post-class summative assessment (n=31)	7.16 (SD = 1.79)

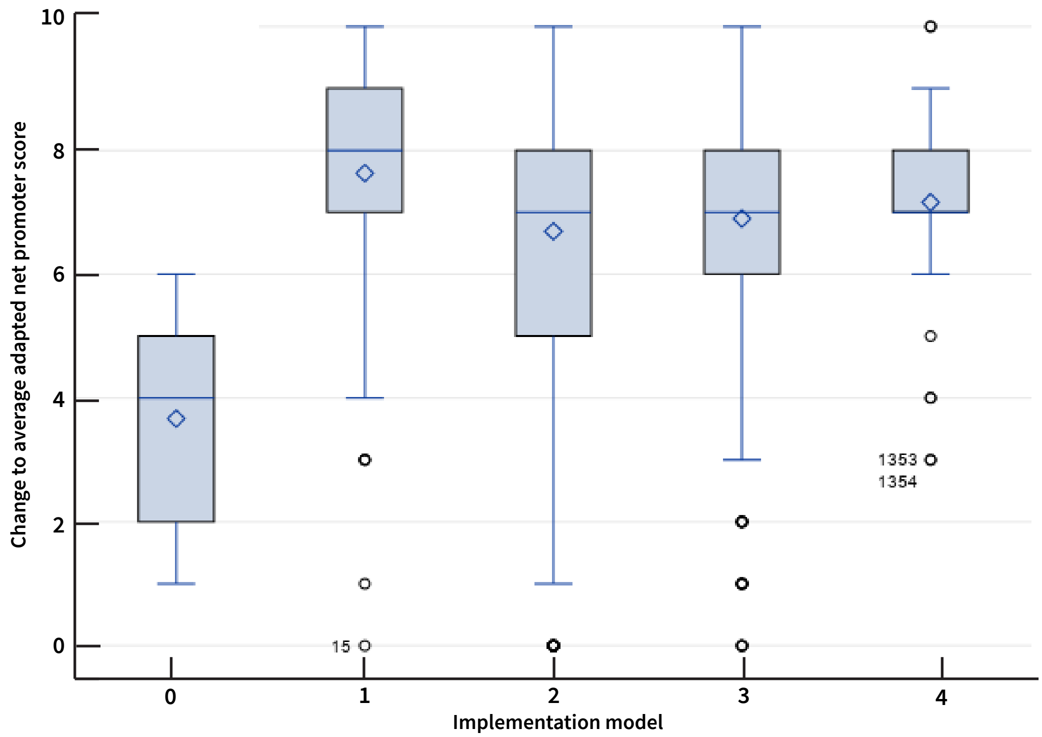
When examined by discipline, average ratings were similar but the average rating from students using Achieve in Economics courses was slightly higher (6.87) and students using Achieve in Calculus courses was slightly lower. There was no meaningful difference between the ratings of students who were taking the course as a major requirement and those taking it as an elective. The average rating of students codified as college ready was significantly higher than those not college ready, as was true for students more motivated to succeed in the course.

Average rating of likelihood to recommend was also examined by institutional context and the instructor's chosen implementation pattern. Investigating whether there were differences in student acceptance by these subgroups could support instructor adoption and implementation decisions. The results can be foundin Table 16.

An ANOVA was calculated to understand if the difference in ratings was statistically significant among implementation groups. Results of the model suggest that there is a statistically significant difference among groups $F(4)=7.01$, $p<.0001$. The box and whisker plot in Graph 3 illustrates the differences. Interestingly, students of instructors who assigned only homework gave the highest average rating and then there was an incremental improvement in average ratings as the implementation pattern became more robust.

An analysis of qualitative data suggests that this difference may have emerged for two reasons, or a combination of both: (1) students who were assigned only homework tended to report that the course load was lighter than students of instructors who implemented more robust implementation patterns, and that they appreciated not having as much out-of-class work to do. For example, *"I liked it — I got less work than in other courses" and "it was two assignments a week so it was really quick, I loved it"* and (2) students in courses where they had to navigate through the system more (i.e from one activity to another) met with more early stage navigation frustrations that would be expected in a first semester beta than did students who used it for only one activity type. When asked to explain their rating, students in more robust implementation patterns tended to comment more often on navigation and usability issues associated with a beta test.

Graph 3. Difference in likelihood to recommend a course to a friend if Achieve was being used, by instructor implementation model



Note: Distribution of NPS
F 7.01; Prob>F <.0001

However, students of the two instructors in the most robust implementation category had the second highest rating, on average (7.6). When asked to explain their rationale for the rating they also noted the navigation complexities but were the group most likely to comment that Achieve supported content mastery and helped them perform better on assessments.

One question on the post-survey asked "please rate the extent to which you agree that Achieve was easy to use" (scale: 1 = "strongly disagree," 2 = "disagree," 3 = "agree," 4 = "strongly agree"). In general, students agreed (M=3.06, SD=0.71) that Achieve was easy to use. This finding was interesting especially since it was in the first of three semesters of beta testing, we expected that the performance and usability areas still being optimized would result in a much lower average rating of ease of use.

Like average rating of recommendation, it is important to measure ease of use among subgroups of students. With the goal of making Achieve easy to use for all students regardless of their background or level of experience, understanding whether there are systematic differences could help focus ongoing user experience research and development efforts. Average responses to the question, rate the extent to which you agree that Achieve was easy to use (Scale 1 = strongly disagree through 4 = strongly agree) can be found in Table 17.

Table 17. Average student ease of use rating overall and by subgroup

	Ease of use rating
Overall (n=1,241)	3.06
Discipline	
Biology (n=203)	2.99
Calculus (n=203)	3.02
Chemistry (n=678)	3.12
Economics (n=764)	3.04
First generation college student	
Yes (n=353)	3.21
No (n=1,208)	3.02
Comfort level with digital	
More (n=1,176)	3.05
Less (n=44)	2.92
Experience with digital	
Used before (n=532)	3.04
Did not use before (n=697)	3.07

Note: Scale is 0-10

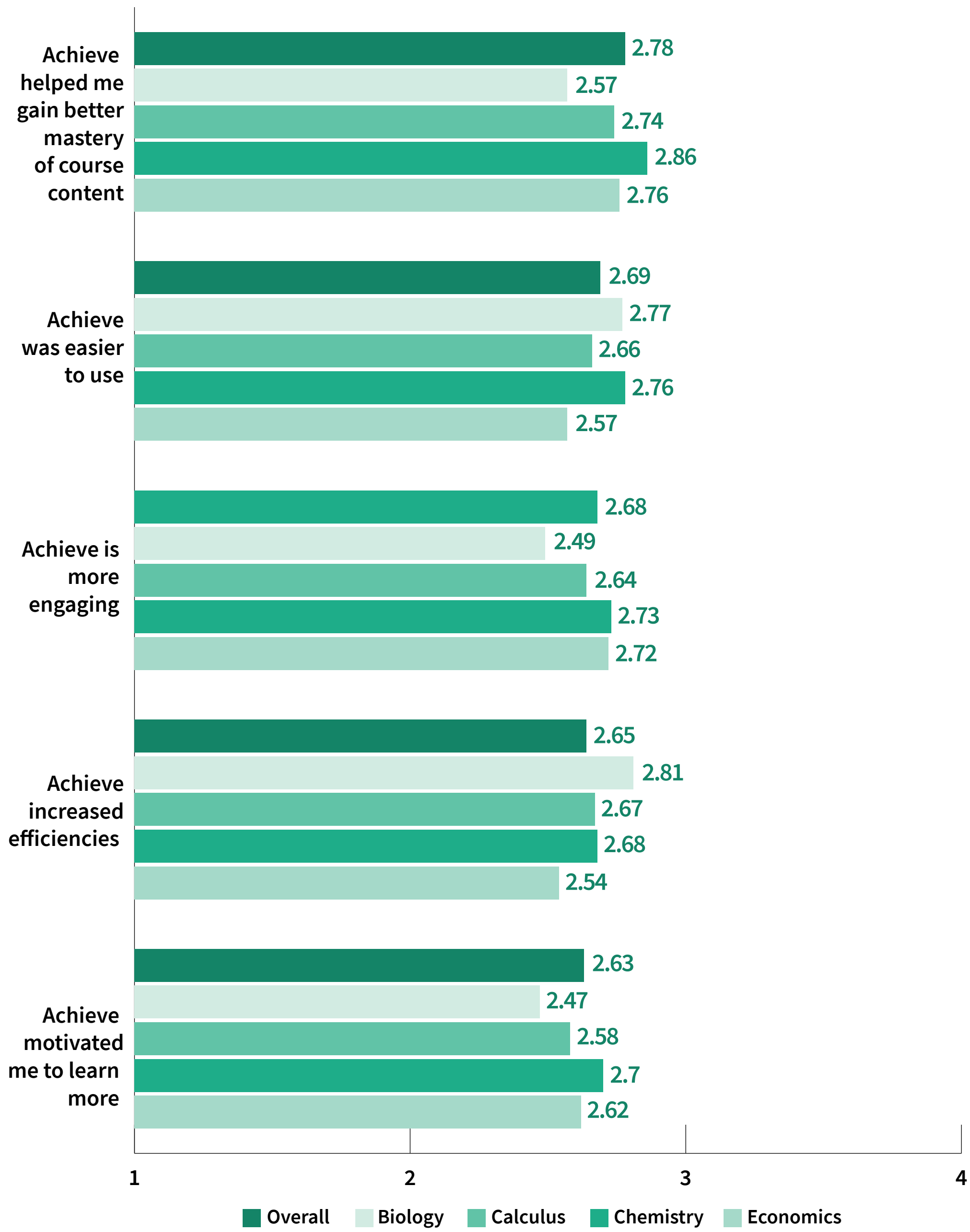
> "I love how easy Achieve is to use — I just set all of my assignments in the beginning of the semester and it basically ran itself."

When examined among disciples there were descriptive differences among average ratings of ease of use but they were not statistically significantly different $F(1,241) = 1.89$; $p = 0.1300$. There also was no meaningful difference in the perception of ease of use when disaggregated by whether the student was the first in their family to attend college, their comfort level with technology, and whether they had used a publisher provided digital tool in a course before. This finding was interesting because we hypothesized that there would be overall lower ratings of ease of use in this first semester of beta and that students less comfortable with technology, those who hadn't used it before, or first generation students would have statistically significantly lower ratings of ease of use.

Comparison to previous digital tools used

Students who had indicated on the pre-survey that they had used a digital learning tool in the past were asked to consider that digital learning tool (or those tools) and rate their level of agreement with five statements on a scale of 1 = "strongly disagree," 2 = "disagree," 3 = "agree," 4 = "strongly agree" on the post-survey. On average, students had moderately more positive perceptions of Achieve than the previous tool(s). When asked to rate their level of agreement that Achieve helped them

Graph 4. Student level of agreement with statements asking them to compare Achieve with previous digital learning tool(s) used, by discipline



gain better mastery of the course content the average rating was 2.78. Students in Economics courses gave a higher than average rating on this item and students in Biology courses gave a lower than average rating. An average rating of 2.69 was found when students were asked their level of agreement with *"Achieve was easier to use,"* again with Economic students more positively agreeing that Achieve was easier to use. Similar trends emerged for ratings of *"Achieve is more engaging"* and *"Achieve motivated me more to learn."* All findings can be found in Graph 4.

Table 18. Differences in student difficulty ratings before Achieve and the semester with Achieve

How difficult is:	Without Achieve	With Achieve	Difference	Significance
Actively engage in classroom discussion	2.76	2.83	+0.065*	t(845)=2.13, p=0.034
Come to class having completed assignments that were due	3.00	3.04	+0.044	t(846)=1.45, p=0.147
Comprehend material	2.67	2.76	+0.083*	t(845)=2.74, p=0.006
Recall concepts that I had memorized	2.78	2.78	+0.005	t(844)=0.16, p=0.871
Come to class prepared to participate	3.07	3.03	-0.045	t(845)=-1.58, p=0.115
Interact with the instructor	2.99	2.91	-0.077*	t(845)=-2.60, p=0.009

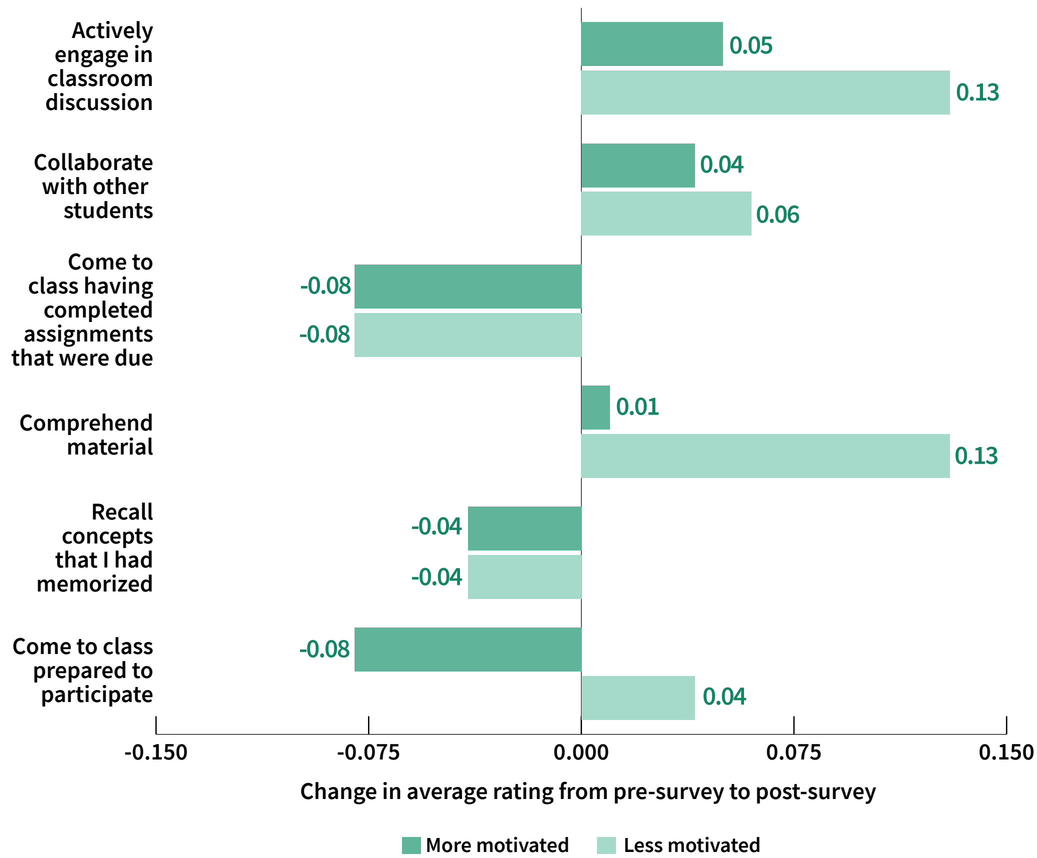
Note:
1 = strongly disagree
2 = disagree
3 = agree
4 = strongly agree

Note: *represents a statistically significant difference at the .05 level.

In an effort to also examine whether Achieve was helping to curb student challenges, on the pre-survey students were asked to rate how difficult a set of six activities typically are for them in college courses. The rating scale was 1 = "very difficult," 2 = "difficult," 3 = "not difficult," 4 = "not difficult at all." They were then asked to rate the same set of items on the post-survey when considering the semester that they used Achieve.

A statistically significant difference was found among one set of paired items, *"Actively engage in classroom discussion."* Students reported that actively engaging in classroom discussion was significantly less difficult during the semester that Achieve was used than it typically is for them (t(845)=2.13, p=0.034). Differences among all six matched pairs can be found in Table 18.

Graph 5. Change in student challenges by level of motivation



Instructors often report that it's difficult to support positive classroom behaviors among students who are less motivated to succeed in the course. Therefore, we wanted to examine whether there were any differences in rating of difficulty among that subgroup of students. Graph 5 suggests that there are. For example, students less motivated to succeed reported that it was easier to come to class prepared to participate when using Achieve while those more prepared did not tend to think so. And, while both groups of students reported that it was easier to actively engage in classroom discussion and comprehend materials, students less motivated to succeed report a substantially greater change in perception. All differences between students more and less motivated to succeed can be found in Graph 5.

RESEARCH QUESTION 3

Is use of Achieve related to academic performance in the course?

We hypothesized that more use of Achieve would positively influence assessment scores. Based on previous research however, we suggested that a student's level of academic preparedness when entering college would also contribute to this relationship. We also suggested students being nested within instructor in this sample would influence the relationship. So, we proposed that high school grade point average (HSGPA) should be controlled and a hierarchical linear model employed.

First, the dependent variable, final exam score, was examined descriptively. Valid final exam score data were available for 1,703 students and scores ranged from 0.00 to 104.40 with an average score of 74.92 (SD = 20.12). The skewness of the distribution was -1.88 and kurtosis was 3.90. The distribution is presented in Graph 6.

Graph 6. Distribution of final exam score

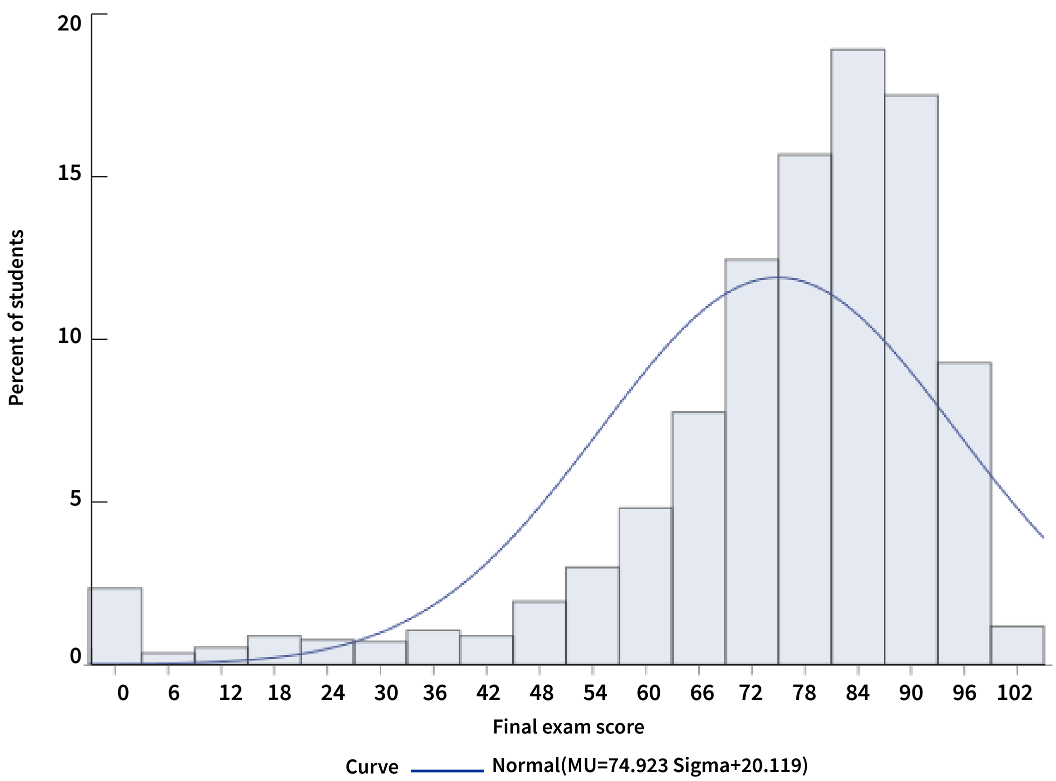
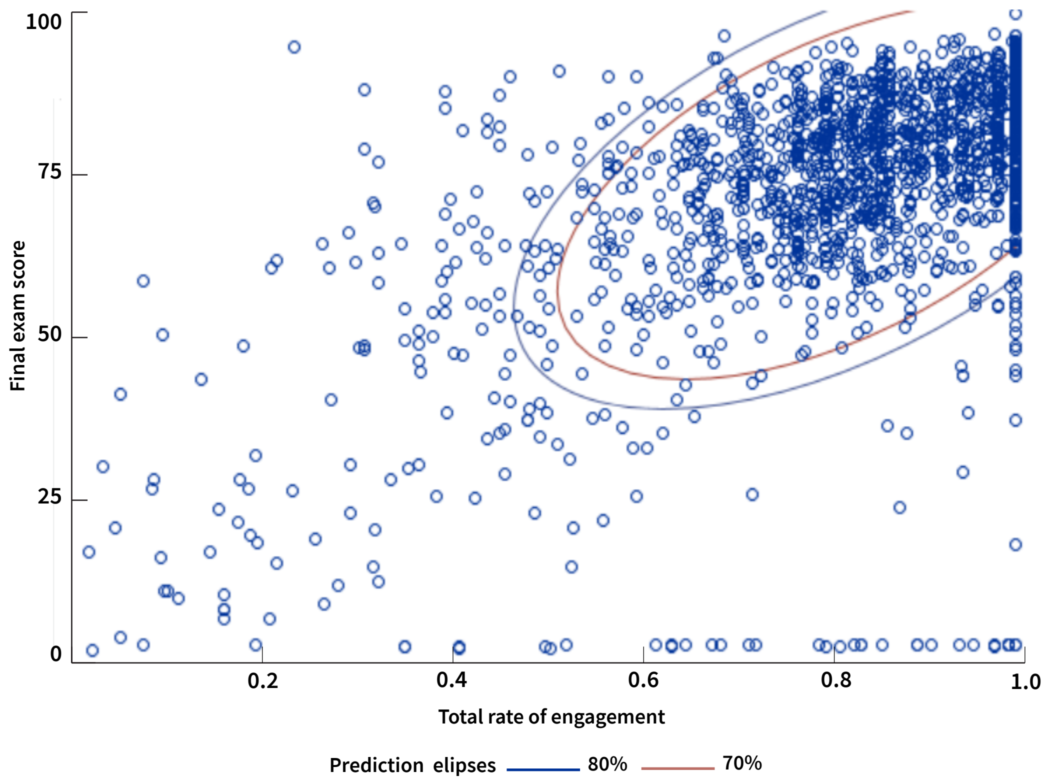


Figure 1. Correlation between engagement in Achieve and final exam score

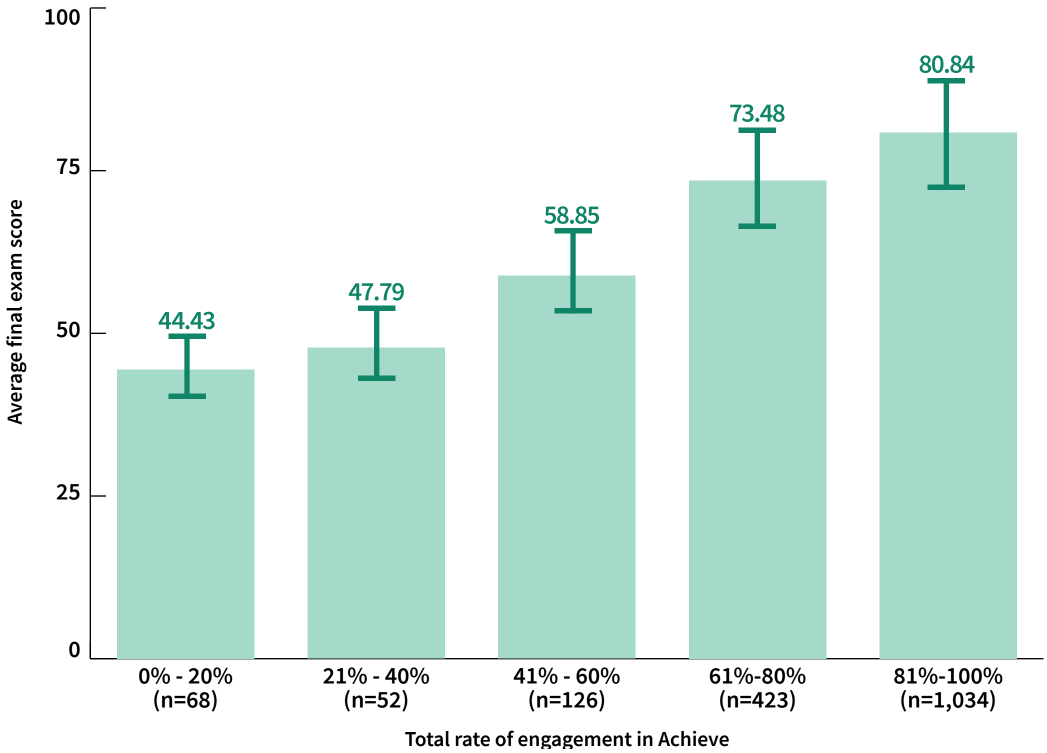


Note: Observations: 1664; Correlation 0.538; p-Value <.0001

For ease of interpretation “rate of engagement” bands were developed, and the average final exam score for the students that fell in each band was calculated. Students who engaged in 0%-20% of assigned activities (n=68) earned an average final exam score of 44.43. Those who engaged in 21%-40% (n=52) earned an average of 47.79. Those who engaged in 41%-60% (n=126) earned an average of 58.85. Those who engaged in 61%-80% (n=423) earned an average of 73.48 and those who engaged in 81% to 100% of assigned activities (n=1,034) earned an average of 80.84.

Because we predicted that a student’s level of academic performance when entering college would contribute to this relationship as well, the correlation between HSGPA and final exam score was calculated and a significant correlation was found .24 (p <.0001), confirming that we should control for this variable in our model. And, since students were grouped within instructors who likely had systematic differences in both characteristics and pedagogical approaches we had to account for that nesting. Consequently, we implemented a hierarchical linear model and controlled for HSGPA to investigate the net relationship between use of Achieve and final exam scores when controlling for related covariates.

Graph 7. Average student final exam scores by rate of engagement in Achieve



> “I’m so impressed with the performance of students class. The average in every test I have given in this class was higher than the average in test for the other two classes I am teaching without Achieve. I believe Achieve played an important role in students success and retention in my class.”

A hierarchical linear model was calculated using PROC MIXED in SAS. The first model investigated was the unconditional model with no predictors to assess between instructor variation in final exam scores. The intraclass correlation coefficient was calculated as ICC = (223.18)/(223.18+243.49) = 0.4782. This finding suggests that about 48% of the variability in final exam scores was attributed to the instructor that the student had, and 52% of the variability was attributable to the student. This finding aligned with our hypothesis and confirmed that we should employ a hierarchical linear model. The findings are displayed in Graph 7.

Table 19. Hierarchical linear model building to examine the influence of engagement in Achieve on final exam scores

	Model 1	Model 2	Model 3	Model 4
Fixed effects				
Intercept	74.93*(28.51)	44.46*(11.29)	45.42*(11.44)	12.14* (3.40)
HSGPA		8.38*(10.38)	8.11*(9.11)	4.31*(6.08)
Engagement				57.36*(31.64)
Error variance				
Level-1	243.49*(8.43)	222.84*(7.87)	222.35*(28.23)	128.38*(4.66)
Level-2 Intercept	223.18*(56.13)	221.78*(55.78)	196.37*(69.40)	186.90*(94.56)
HSGPA			2.11(3.61)	1.59(6.09)
Model fit				
AIC	14310.7	13619.4	13620.9	12469.7
BIC	14315.3	13625.5	13628.5	12478.9

Note: *denotes statistical significance, $p<.05$; ICC = .48, error variance reported in Table 19
Values based on SAS PROC MIXED. Entries show parameter estimates with standard errors in parentheses
Estimation Method = ML; Satterthwaite degrees of freedom

We continued the model building process by including HSGPA as a fixed effect, then including it as a random effect. Results of the second model demonstrated that the overall model was significant as was HSGPA ($p<0.0001$). We then added HSGPA as a random effect to determine if the influence on final exam score varied among instructors. The model remained significant as did HSGPA ($p<0.0001$). Results from this model and summary results of this model building process are presented in Table 19.

We evaluated the change in AIC and BIC and concluded that Model four was the best fitting model. And, given that the inclusion of student level of engagement in Achieve emerged as significant we have evidence to conclude that use of Achieve is predictive of academic performance in the student’s course. More specifically, the more assigned activities that a student engages in, the higher they can expect their final exam score to be in the course. To put it in language that illustrates the practical significance of this finding, for every ten percent increase in a student’s engagement in assigned activities, they can expect a 5.7 percentage point increase on their final exam score.

Table 19. Hierarchical linear model building to examine the influence of engagement in Achieve on final exam scores among Biology students

	Model 1 (n=179)	Model 2 (n=178)	Model 3 (n=178)	Model 4 (n=178)
Fixed effects				
Intercept	78.27*(23.72)	44.00*(5.32)	44.00*(5.32)	39.66*(7.10)
HSGPA		9.28*(4.52)	9.28*(4.52)	
Engagement				45.03*(8.50)
Error variance				
Level-1	230.60*(9.29)	204.72*(9.27)	204.72*(9.27)	163.16*(9.26)
Level-2 Intercept	65.22*(1.68)	68.90*(1.73)	68.86*(1.73)	65.85*(1.76)
HSGPA			0	
Model fit				
AIC	1502	1475.5	1475.5	1436.3
BIC	1501.8	1475.3	1475.3	1436.1

Note: *denotes statistical significance, p<.05; ICC = .22, error variance reported in Table 20
Values based on SAS PROC MIXED. Entries show parameter estimates with standard errors in parentheses
Estimation Method = ML; Satterthwaite degrees of freedom

RESULTS BY DISCIPLINE

Biology

To measure the influence of engagement in Achieve on the dependent variable, final exam scores, a hierarchical linear model was calculated using PROC MIXED in SAS. The intraclass correlation coefficient (0.25) suggests a quarter of the variation in final exam scores could be attributed to the instructor that was teaching the course. The results of the model building process to account for instructor variability and controlling for HSGPA among Biology students is presented in Table 19.

Table 21. Hierarchical linear model building to examine the influence of engagement in Achieve on final exam scores among Calculus students

	Model 1 (n=203)	Model 2 (n=197)	Model 3 (n=197)	Model 4 (n=194)
Fixed effects				
Intercept	74.52*(31.56)	58.15*(7.51)	58.21*(8.05)	22.48*(2.74)
HSGPA		4.51*(2.28)	4.62*(2.18)	4.13*(2.21)
Engagement				45.41*(7.57)
Error variance				
Level-1	242.65*(9.85)	229.29*(9.68)	227.47*(9.69)	167.64*(9.56)
Level-2 Intercept	36.01(1.54)	64.42(1.69)	0	34.71(0.37)
HSGPA			4.63(1.67)	3.66(0.49)
Model fit				
AIC	1709.1	1654	1652.2	1577
BIC	1709.6	1654.8	1653	1578.2

Note: *denotes statistical significance, p<.05; ICC = .22, error variance reported in Table 21
Values based on SAS PROC MIXED. Entries show parameter estimates with standard errors in parentheses
Estimation Method = ML; Satterthwaite degrees of freedom

Calculus

To measure the influence of engagement in Achieve on the dependent variable, final exam scores, a hierarchical linear model was calculated using PROC MIXED in SAS. The intraclass correlation coefficient (0.22) suggests 22% of the variation in final exam scores could be attributed to the instructor that was teaching the course. The results of the model building process to account for instructor variability and controlling for HSGPA among Calculus students is presented in Table 21.

After examining AIC and BIC differences, Model four was determined to be best fitting. The results suggest that for every ten percent increase in Calculus student's engagement in Achieve they can expect a 4.5 percentage point increase in their final exam scores.

Table 22. Hierarchical linear model building to examine the influence of engagement in Achieve on final exam scores among Chemistry students

	Model 1 (n=678)	Model 2 (n=638)	Model 3 (n=638)	Model 4 (n=617)
Fixed effects				
Intercept	74.47*(28.94)	36.56*(7.01)	34.70*(7.21)	10.34*(2.15)
HSGPA		10.67*(8.10)	11.17*(7.76)	4.85*(4.56)
Engagement				62.21*(19.56)
Error variance				
Level-1	279.88*(19.29)	245.66*(17.73)	245.97*(17.74)	143.40*(17.43)
Level-2 Intercept	37.45(1.32)	25.52(1.24)	0	52.70(1.53)
HSGPA			1.67(1.24)	0
Model fit				
AIC	5757.7	5342	5324	4844
BIC	5763.9	5342.3	5342.3	4844.4

Note: *denotes statistical significance, p<.05; ICC = .12, error variance reported in Table 22
Values based on SAS PROC MIXED. Entries show parameter estimates with standard errors in parentheses
Estimation Method = ML; Satterthwaite degrees of freedom

Chemistry

To measure the influence of engagement in Achieve on the dependent variable, final exam scores, a hierarchical linear model was calculated using PROC MIXED in SAS. The intraclass correlation coefficient (0.12) suggests 12% of the variation in final exam scores could be attributed to the instructor that was teaching the course. The results of the model building process to account for instructor variability and controlling for HSGPA among Chemistry students is presented in Table 22.

The estimated G matrix of Model three was not positive definite and the estimate, standard error, z-value and p value were not produced for the intercept, suggesting that the constant was a negative. However, this was not overly problematic, so we continued the model building process. The estimated G matrix of Model four was not positive definite and the estimate, standard error, z-value and p value were not produced for HSGPA, suggesting it should be considered for removal. It was removed and a fifth model was run. However, after examining AIC and BIC differences, the fourth model including HSGPA was determined to be best fitting. The results suggest that for every ten percent increase in Chemistry student's engagement in Achieve they can expect a 6.2 percentage point increase in their final exam scores.

Table 22. Hierarchical linear model building to examine the influence of engagement in Achieve on final exam scores among Economics students

	Model 1 (n=643)	Model 2 (n=624)	Model 3 (n=624)	Model 4 (n=614)
Fixed effects				
Intercept	73.58*(9.38)	49.52*(5.30)	49.67*(5.35)	12.14*(3.40)
HSGPA		6.50*(4.50)	6.45*(4.33)	4.31*(6.08)
Engagement				57.36*(31.64)
Error variance				
Level-1	209.05*(17.79)	200.37*(17.52)	200.20*(17.43)	128.38*(4.66)
Level-2 Intercept	611.12*(2.22)	582.29*(2.22)	566.65*(2.00)	186.90*(94.56)
HSGPA			1.10(0.14)	1.59(6.09)
Model fit				
AIC	5317.1	5136.9	5138.9	12469.7
BIC	5318	5138.1	5140.4	12478.9

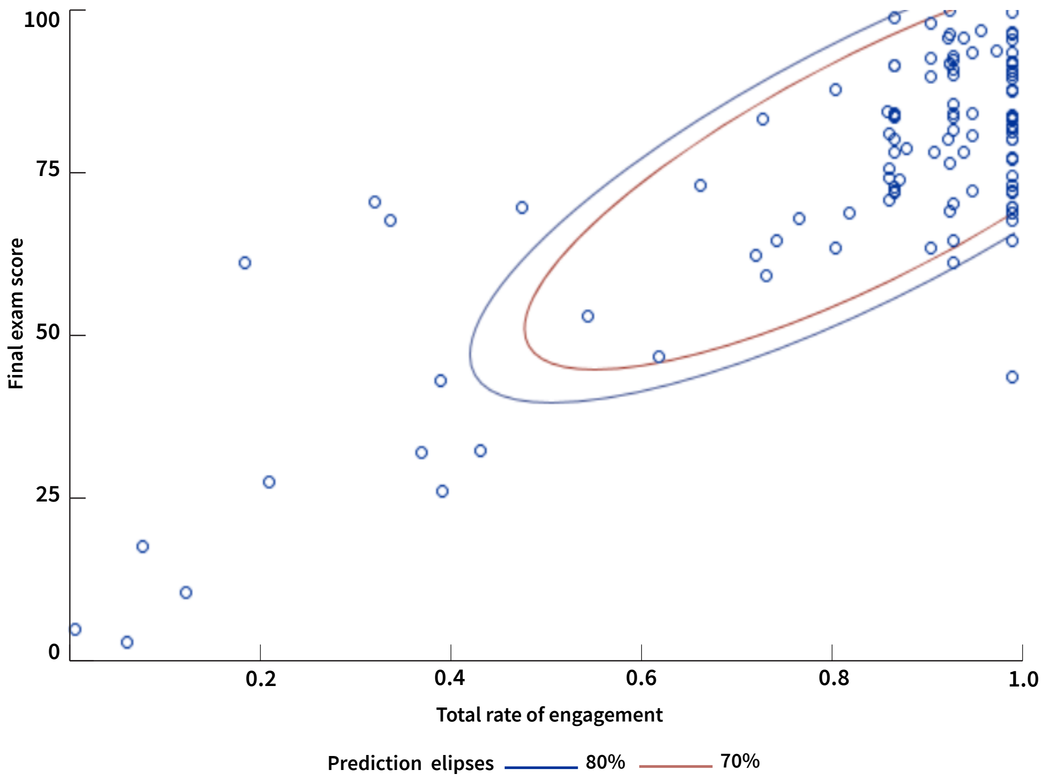
Note: *denotes statistical significance, p<.05; ICC = .25, error variance reported in Table 23
Values based on SAS PROC MIXED. Entries show parameter estimates with standard errors in parentheses
Estimation Method = ML; Satterthwaite degrees of freedom

Economics

To measure the influence of engagement in Achieve on the dependent variable, final exam scores, a hierarchical linear model was calculated using PROC MIXED in SAS. The intraclass correlation coefficient (0.12) suggests 12% of the variation in final exam scores could be attributed to the instructor that was teaching the course. The results of the model building process to account for instructor variability and controlling for HSGPA among Chemistry students is presented in Table 23.

After examining AIC and BIC differences, Model four was determined to be best fitting. The results suggest that for every ten percent increase in Calculus student's engagement in Achieve they can expect a 5.7 percentage point increase in their final exam scores.

Figure 2. Relationship between engagement in Achieve and final exam score among students in "post-class only" implementation model



Note: Observations 114; Correlation 0.7985; p-Value <.0001

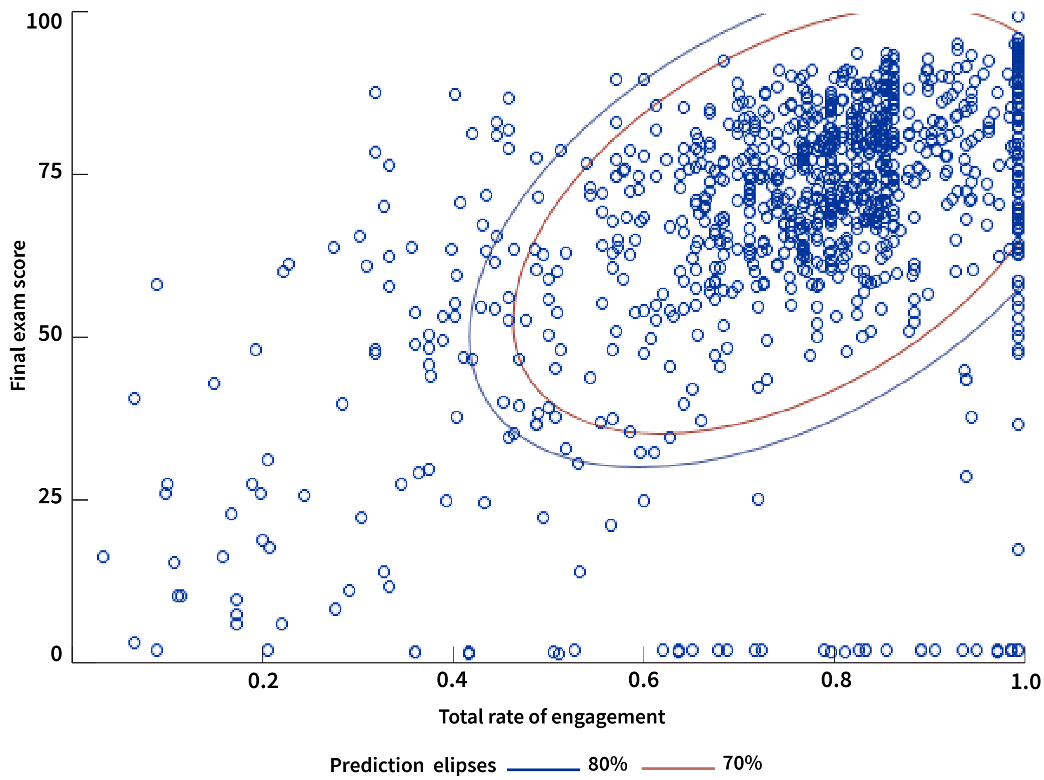
RESULTS BY IMPLEMENTATION MODEL

Post-class summative only.

To understand the relationship between engagement in Achieve and final exam scores among students in courses where a "post-class summative" implementation model was used, the correlation between the total rate of engagement (all assignments engaged in*students/all assignments assigned*students) and final exam score was calculated using PROC CORR in SAS. A significant correlation was found .80 (p<.0001).

To further measure the influence of engagement in Achieve on final exam scores among this group of students, a multiple linear regression was calculated using PROC REG in SAS (a hierarchical linear model was not used because once disaggregated by implementation model sample size was not large enough) to predict student final exam score based on students rate of engagement in post-class summative assessments in Achieve (all summative assessments assigned*student/all summative assessments assigned) when controlling for high school grade point average (note smaller sample size due to missingness on HSGPA). A significant regression equation was found (F(2,18)p=.0124), with an adjusted R2 of 0.31, suggesting that 31% of the variability in students' final exam score among students in this implementation model could be attributed to their prior academic performance and the proportion of assignments that they engaged in. The squared partial correlation (type 2) was 0.31 suggesting that the majority of the variability accounted for in the model could be attributed to engagement in Achieve. The findings are displayed in Figure 2.

Figure 3. Relationship between engagement in Achieve and final exam score among students in "pre-class formative and post-class formative or summative" implementation model



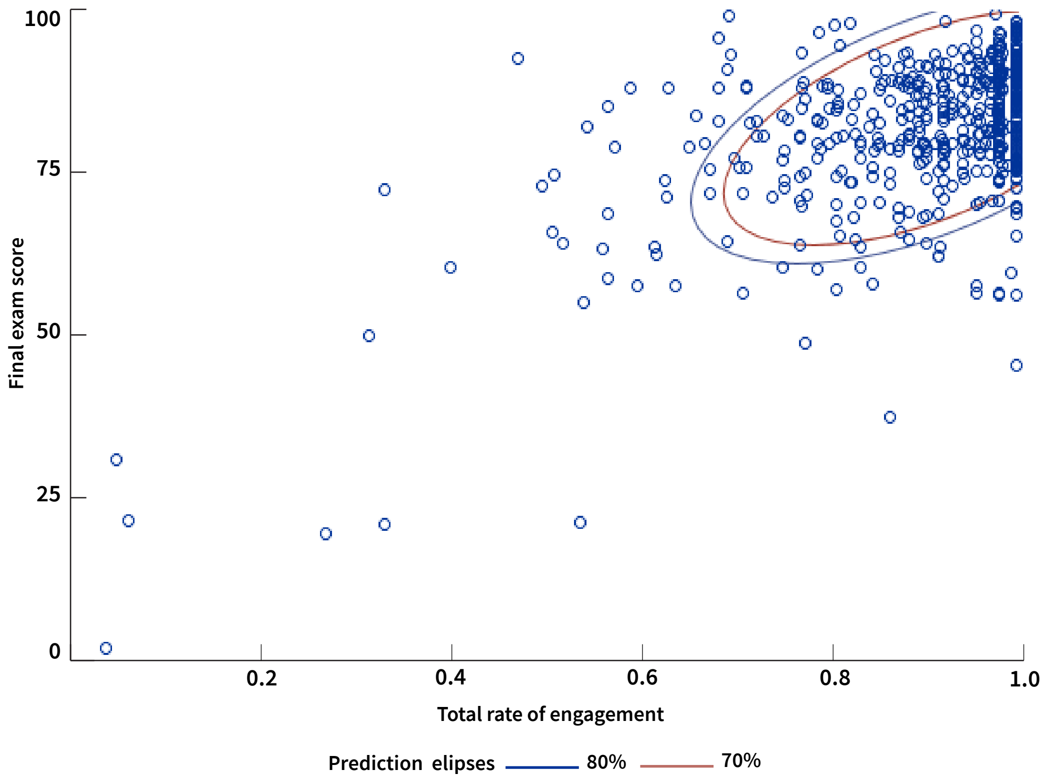
Note: Observations: 962; Correlation 0.4682; p-Value <.0001

Pre-class formative and post-class formative or summative.

To understand the relationship between engagement in Achieve and final exam scores among students in courses where a "pre-class formative and post-class formative or summative" implementation model was used, the correlation between the total rate of engagement (all assignments engaged in*students/all assignments assigned*students) and final exam score was calculated using PROC CORR in SAS. a significant correlation was found .47 ($p<.0001$).

To further measure the influence of engagement in Achieve on final exam scores among this group of students, a multiple linear regression was calculated to predict student final exam score based on students rate of engagement in pre-class formative assessments, and post-class assessments in Achieve when controlling for high school grade point average among students in this implementation model (note smaller sample size due to missingness on HSGPA). A significant regression equation was found ($F(4,95)p<.001$), with an adjusted R^2 of 0.39, suggesting that 39% of the variability in students' final exam score among students in this implementation model could be attributed to their prior academic performance and the proportion of assignments that they engaged in. The squared partial correlations (type 2) for pre-class formative, post-class formative, and post-class summative were 0.04, 0.16, and 0.01. The findings are displayed in Figure 3.

Figure 4. Relationship between engagement in Achieve and final exam score among students in “pre-class formative and post-class formative or summative”



Note: Observations 546; Correlation 0.5488; p-Value <.0001

Pre-class formative and post-class formative and summative.

To understand the relationship between engagement in Achieve and final exam scores among students in courses where a “pre-class formative and post-class formative and summative” implementation model was used, the correlation between the total rate of engagement (all assignments engaged in*students/all assignments assigned*students) and final exam score was calculated using PROC CORR in SAS. a significant correlation was found .55 (p<.0001).

To further measure the influence of engagement in Achieve on final exam scores among this group of students multiple linear regression was calculated to predict student final exam score based on students rate of engagement in pre-class formative assessments, and post-class assessments in Achieve when controlling for high school grade point average among students whose instructors assigned pre-class formative and both types of post-class activities (note smaller sample size due to missingness on HSGPA). A significant regression equation was found (F(4,384)p<.0001), with an adjusted R2 of 0.40, suggesting that 40% of the variability in students’ final exam score among students in this implementation model could be attributed to their prior academic performance and the proportion of assignments that they engaged in. The squared partial correlations (type 2) for pre-class formative, post-class formative, and post-class summative were 0.12, 0.01, and 0.03. The findings are displayed in Figure 4.

Table 24. Summary of relationships between engagement in Achieve and final exam score, by implementation model

Implementation model	Correlation between rate of engagement overall and final exam score	Multiple linear regression: Analysis of variance	Multiple linear regression: Adjusted R^2
Post-class summative	0.80**	5.66*	0.31
Pre-class formative and post-class formative or summative	0.47**	16.53**	0.39
Pre-class formative, post-class formative, and post-class summative	0.55**	63.31**	0.40
Pre-class formative, in-class, post-class formative, and post-class summative	0.23	Insufficient valid data to calculate	

Note: *denotes statistical significance at the p<0.05 level, ** denotes statistical significance at the p <0.001 level

> "The students completing pre-class assignments meant they were coming to class much more prepared than in previous semsters."

Pre-class formative, in-class, post-class formative and summative.

Finally, the same investigation was conducted among students whose instructors assigned pre-class, in-class, post-class formative, and post-class summative activities in Achieve. Neither the correlation between total engagement and final exam scores (.23, p=0.190) nor the relationship that emerged in the linear regression in this implementation model were significant (F(2,31)p=.291).

A summary of the correlations and the variability in a student's final exam score that engagement in Achieve accounts for by implementation model can be found in Table 24. Note that disaggregation results in small sample sizes so it will be important to replicate the analyses when the Fall 2019 replication study is concluded.

Discussion

When building Achieve the goals were to meet the needs of instructors and a changing student population in a higher education landscape that is always evolving. To develop a learning tool to achieve these goals, research-based learning science principles set the foundation and a design and iteration process focused on close partnership with instructors and students followed. The unusual step of beginning efficacy research in development enabled early insights that contributed to data-driven development decisions and offered early insights to instructors.

Building a portfolio of evidence of efficacy over time helped close three of the gaps that currently exist in educational technology efficacy. First, that insights are available to instructors before they have to make adoption and implementation decisions — curbing false starts and frustrations. Second, the design, development, and iteration process is part of the efficacy argument. Evidence derived from these steps offers instructors confidence that the tool was designed to meet their needs and the needs of their students. And third, the evidence of effectiveness provided in this report is examined at the aggregate and within educational context and implementation pattern. Establishing evidence in this way, the findings are relevant to instructors in specific contexts who are planning to use the tool in a specific way. And, evidence of effectiveness when used in a specific way may provide insights that help instructors considering using the tool in a way they otherwise would not have.

This study provided sound evidence that research-based learning engineering acts as the foundation for Achieve. Achieve was conceived based on six learning design principles. First, developing learning motivation which is accomplished through goal setting and positive feedback within assessments. Second, providing personalized and adaptive experiences enabled through adaptive, gamified reading quizzes and interactives. Third, targeting cognitive and memory elements which students realize in the "testing effect" enabled by retrieval practice and retaking assessments. Fourth, Achieve is built on a well-constructed active learning model of interconnected pre-, in-, and post-class learning experiences. Fifth, the creation of interactive and constructive opportunities enabled by interactive assessments and in-class activities that promote engagement and collaborative learning. And finally, metacognition and self-regulation which is enabled throughout the learners experience. For example, self-regulation is promoted by the organization of the student's course planner that helps them track what assignments are coming due, the progress they have made on assignments they have begun, and where they have completion gaps.

Also demonstrated in the evidence provided is that Achieve was developed based on three core learning science foundations. First, content and assessments tagged to effective learning objectives enable instructional alignment across learning experiences and drive assessment task development. Second, Achieve supports impactful assessment practice by improving transparency of learning goals to stakeholders, supporting instructional alignment, and enabling monitoring of learner progress, and timely, targeted interventions. Third, Achieve offers empowering analytics through the dashboard reports that provide actionable insights for instructors. The information provided in the instructor-facing dashboards report against learning objectives to enable monitoring and mastering of concepts, application of skills, and development of attributes.

With generative research established, an evolving cycle of efficacy research was implemented beginning when the product was in its alpha stage with a formative evaluation, conducted outside of and independent from an instructor's live course. Following the formative evaluation was the study presented in this report, an implementation study of Achieve. To evaluate instructor implementation, student engagement, and teaching and learning outcomes, Achieve was evaluated across multiple disciplines, used with multiple textbooks, and used within multiple educational contexts. Three primary research questions were evaluated in this study. First, how are instructors using Achieve and how are students engaging with it. In order to understand efficacy, use should first be systematically investigated, as outcomes will likely vary based on implementation. In this evaluation, instructors used Achieve regularly throughout the semester and four distinct implementation patterns emerged. The four patterns included (1) assignment of post-class summative assessments only (2) assignment of pre-class formative assessments and either post-class formative or post-class summative assessments (3) pre-class formative assessments and post-class formative and summative assessments and (4) pre-class formative, in-class, post-class formative and post-class summative assessments. There were differences in implementation pattern by educational context, for example instructors at two-year institutions were more likely to fall into atern one and less likely to fall into pattern two. Additionally, all instructors in pattern one teach at large institutions. And, instructors assigning the most robust pattern (four) were more likely to be teaching at highly selective institutions.

Student engagement rates in Achieve were relatively high (overall engagement average = 80.3%), which was interesting given the previous literature that students engage in digital learning tools at a relatively low rate. Qualitative data suggest that students enjoyed the game-like features of Achieve assessments, that they perceive the activities to have helped them prepare for class, and that they perceived Achieve to help them prepare for exams — likely explaining the high engagement rates. When considered by activity type, engagement rates varied. The highest rates were observed among pre-class formative assessments (average overall engagement rate in pre-class activities = 85.6%). The high pre-class formative assessment rates were consistent across discipline. Survey and focus group data validate this finding suggesting that they are the activity types students have the highest perception of. The lowest engagement rates (average in-class engagement rate = 45.6%) emerged among in-class activities and this finding was consistent across disciplines as well. However, qualitative data suggest that it may be because they were supplemental materials that were not required however, limitations of the beta version of Achieve did not allow instructors to assign activities for zero points. It may be that in-class activities were not required, explaining the low engagement rates.

Once implementation, engagement, and perception were well-understood, the authors explored whether use of Achieve influenced final-exam scores when student prior academic performance, baseline level of motivation, and who the instructor was controlled for. Also explored was whether there was a difference in the relationship between usage an exam scored by implementation pattern. There was a significant correlation between a student's rate of engagement across all activities and their final exam scores ($r=.054, p<0.001$). It was hypothesized that a student's baseline level of academic preparedness and their level of motivation to succeed in the course were also related to final exam scores. And, since students were nested within instructor it was determined that nesting had to be accounted for. A hierarchical linear model was built. Results suggested that the more assigned activities that a student engages in, the higher they can expect their final exam score to be in the course. That is, for every ten percent increase in a student's engagement in assigned activities, they can expect a 5.7 percentage point increase on their final exam score.

In this sample of instructors, an average of 76.75 activities were assigned for credit over the course of the semester in Achieve. Consider, for example, that a student had engaged in 60 of them and earned a 75 on their final exam in the course. Had the student engaged in seven additional assigned activities, they could have expected an 80.7 on their final exam, bringing them from a grade of a C to a grade of a B — regardless of their level of academic preparedness coming into college. Analyses were replicated within each discipline examined in this study and a significantly positive relationship emerged in each discipline.

Due to a limitation of data, correlations (rather than hierarchical linear models) were calculated to examine the relationship within implementation pattern. Results suggest that the more robust an implementation pattern, the more overall variance in final exam score use of Achieve accounted for. Interestingly, there was the strongest correlation between engagement and final exam scores among students who were assigned only post-class summative assessments. A possible explanation for the strength of the relationships could be because the content on post-class summative assessments in Achieve are much more closely aligned to the content in final exams. However, this is a relationship that will be further examined in the replication study.

Limitations and future research

The results in this study are very promising and contribute sound evidence to the efficacy argument of Achieve, but like all applied research there are important limitations to discuss. First, this study was conducted in the first semester of beta testing Achieve, meaning that the tool was in a formative state and still in development. As such, instructors may have used Achieve in a different way or more robustly if it was fully developed. Implementation patterns that emerged in this study will be compared with those observed in the replication study to understand if different user cases are more robust with a more fully developed tool.

Second, instructors who agree to participate in an early beta test of a digital learning tool are likely to have more positive perceptions of these tools to begin with and be more comfortable with technology. As we discussed in the description of our sample, nearly all instructors were comfortable with technology and have positive perceptions of it. In the replication study we attempted to partner with more instructors who are less comfortable with technology to ensure that the supports that are built into Achieve enable the same outcomes to be realized by less comfortable instructors as we observed with more comfortable instructors this study.

Most important to note is that the design and analyses presented in this study are correlational and therefore causal statements cannot be made based on the results. Although we controlled for student prior academic performance and baseline level of motivation, there are a myriad of other factors that could be contributing to the outcomes measured. A quasi-experimental study that will build on these results and the results from the replication study is planned for Spring 2020. The results from the analyses conducted during that study will enable causal statements of efficacy.

Conclusion

Macmillan Learning took the unusual approach of beginning to collect evidence of efficacy in Achieve's infancy. These results were used for both the evolution and optimization of Achieve and so that instructors could have timely insights that were relevant to them before they made adoption and implementation decisions. The findings from this study are promising and suggest that Achieve can help all students succeed, but should be interpreted as results from an early beta product. The authors look forward to building on this evidence in the studies to come.

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., Bain, P. M., & Chamberlain, R. W. (2012). The value of applied research: Retrieval practice improves classroom learning and recommendations from a teacher, principal, and a scientist. Educational Psychology Review, 24, 437-448.

Akinoğlu, O., & Tandoğan, R. O. (2007). The Effects of Problem-Based Active Learning in Science Education on Students' Academic Achievement, Attitude and Concept Learning. Online Submission, 3(1), 71-81.

Alexander, P.A., Schallert, D.L., & Reynolds, R.E. (2009). What is learning anyway? A topographical perspective considered. Educational Psychologist, 44(3), 176-192.

Bai, Y., & Chang, T. (2016). Effects of class size and attendance policy on university classroom interaction in Taiwan. Innovations in Education and Teaching International, 53(3), 316-328.

Bernardo, A., Esteban, M. Fernández, Cervero, A., Tuero, E., & Solano, P. (2016). Comparison of personal, social and academic variables related to university drop-out and persistence. Frontiers in Psychology, 7(1610), 1-9.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. Annual Review of Psychology, 64, 417-444.

Bjork, R.A., & Bjork, E.L. (1992). A new theory of disuse and of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), From learning processes to cognitive processes: Essays in honor of William K. Estes (Vol. 2, pp. 35-67). Hillsdale, NJ: Erlbaum.

Cho, M., Kim, Y., & Choi, D. (2017). The effect of self-regulated learning on college students' perceptions of community of inquiry and affective outcomes in online learning. The Internet and Higher Education, 34, 10-17.

Conley, A.M. (2012). Patterns of motivation beliefs: Combining achievement goal and expectancy-value perspectives. Journal of Educational Psychology, 104(1), 32-47.

Crouch, C.H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. American Journal of Physics, 69(9), 970-977.

DeLozier, S.J., & Rhodes, M.G. (2017). Flipped Classrooms: A review of key ideas and recommendations for practice. Educational Psychology Review, 29(1), 141-151.

Do, S.L., & Schallert, D.L. (2004). Emotions and classroom talk: Toward a model of the role of affect in students' experiences of classroom discussions. Journal of Educational Psychology, 96(4), 619-634.

Dunlosky, J., & Metcalfe, J. (2009). Metacognition. Thousand Oaks, CA: Sage.

Ehrlinger, J., Mitchum, A.L., & Dweck, C.S. (2016). Understanding overconfidence: Theories of intelligence, preferential attention, and distorted self-assessment. Journal of Experimental Social Psychology, 63, 94-100.

English, M., & Kitsantas, A. (2013). Supporting student self-regulated learning in problem- and project-based learning. Interdisciplinary Journal of Problem-Based Learning 7(2), 128-150.

Feng, M., Heffernan, N.T., & Koedinger, K.R. (2009). Addressing the assessment challenge with an online system that tutors as it assesses. User Modeling and User-Adapted Interaction: The Journal of Personalizational Research, 19(3), 243-266.

Fyfe, E.R., & Rittle-Johnson, B. (2016). Feedback both helps and hinders learning: The causal role of prior knowledge. Journal of Educational Psychology, 108(1), 82-97.

Goedert, J. D., Pawloski, R., Rokooeisadabad, S., & Subramaniam, M. (2013). Project-oriented pedagogical model for construction engineering education using cyberinfrastructure tools. Journal of Professional Issues in Engineering Education and Practice, 139(4), 301-309.

Greene, J.A., & Azevedo, R. (2009). A macro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system. Contemporary Educational Psychology, 34, 18-29.

Hattie, J. & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81-112.

Hirsh-Pasek, K., Zosh, J.M., Golinkoff, R.M., Gray, J.H., Robb, M.B., & Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. Psychological Science in the Public Interest, 16(1), 3-34.

Joët, G. Usher, E.L., & Bressoux, P. (2011). Sources of self-efficacy: An investigation of elementary school students in France. Journal of Educational Psychology, 103(3), 649-663.

Kay, R., & Kletskin, I. (2012). Evaluating the use of problem-based video podcasts to teach mathematics in higher education. Computers & Education, 59(2), 619-627.

Kendeou, P., Walsh, E.K., Smith, E.R., & O'Brien, E.J. (2014). Knowledge revision processes in refutation texts. Discourse Processes, 51(5-6), 374-397.

Kim, M., & Schallert, D.L. (2011). Building caring relationships between a teacher and students in a teacher preparation program word-by-word, moment-by-moment. Teaching and Teacher Education, 27, 1059-1067.

Kizilcec, R.F., Pérez-Sanagustín, M., & Maldonado, J.J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in massive open online courses. Computers & Education, 104, 18-33.

Knox, J. (2017). Data power in education: Exploring critical awareness with the "learning analytics report card." Television & New Media, 18(8), 734-752.

Mayer, R.E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. American Psychologist, 63, 760-769.

Mayer, R.E. (2014). Incorporating motivation into multimedia learning. Learning and Instruction, 29, 171-173.

McDaniel, M.A., & Butler, A.C. (2010). A contextual framework for understanding when difficulties are desirable. In A.S. Benjamin (Ed.), Successful remembering and successful forgetting: A festschrift in honor of Robert A. Bjork (pp. 175-198). New York, NY: Psychology Press.

Muehlenkamp, J., Weiss, N., & Hansen, M. (2015). Problem-based learning for introductory psychology: preliminary supporting evidence. Scholarship of Teaching and Learning in Psychology, 1(2), 125-136.

Park, Y., & Jo, I. (2015). Development of the learning analytics dashboard to support students' learning performance. Journal of Universal Computer Science, 21(1), 110-133.

Pekrun, R., Elliot, A.J., & Maier, M.A. (2009). Achievement goals and achievement emotions: Testing a model of the joint relations with academic performance. Journal of Educational Psychology, 101(1), 115-135.

Richey, J. E., & Nokes-Malach, T. J. (2015). Comparing four instructional techniques for promoting robust knowledge. Educational Psychology Review, 27, 181-218.

Richland, L. E., Stigler, J. W., & Holyoak, K. J. (2012). Teaching the conceptual structure of mathematics. Educational Psychologist, 47(3), 189-203.

Roberts, L.D., Howell, J.A., & Seaman, K. (2017). Give me a customizable dashboard: Personalized learning analytics dashboards in higher education. Technology, Knowledge, and Learning, 22, 317-333.

Santos, O.C., Boticario, J.G., & Pérez-Marín, D. (2014). Extending web-based educational systems with personalised support through user centred designed recommendations along the e-learning life cycle. Science of Computer Programming, 88, 92-109.

Sawyer, K. (Ed.). (2014). The Cambridge handbook of the learning sciences (2nd ed.). New York: Cambridge University Press.

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. Computers in Human Behavior, 78, 397-407.

Simpson, O. (2013). Students retention in distance education: Are we failing our students? Open Learning, 28(2). 105-119.

Sternberg, R.J. (2007). Culture, instruction, and assessment. Comparative Education, 43(1), 5-22.

Tseng, J.C.R., Chu, H.C., Hwang, G.J., & Tsai, C.C. (2008). Development of an adaptive learning system with two sources of personalization information. Computers & Education, 51, 776-786.

Tseng, K. H., Chang, C. C., Lou, S. J., & Chen, W. P. (2013). Attitudes towards science, technology, engineering and mathematics (STEM) in a project-based learning (PjBL) environment. International Journal of Technology and Design Education, 23(1), 87-102.

Volet, S., Summers, M., & Thurman, J. (2009). High-level co-regulation in collaborative learning: How does it emerge and how is it sustained? Learning and Instruction, 19, 128-143.

Wolters, C.A. (2004). Advancing achievement goal theory: using goal structures and goal orientations to predict students' motivation, cognition, and achievement. Journal of Educational Psychology, 96(2), 236-250.

APPENDIX

Table A1. Distribution of Instructor baseline background and pedagogical data

		%
Years teaching		
	First year	0
	1-5 years	11.43
	6-10 years	25.71
	11-15 years	22.86
	More than 15 years	40.00
Comfort with technology		
	Extremely comfortable	40.00
	Comfortable	40.00
	Uncomfortable	17.14
	Extremely uncomfortable	2.9
Level of agreement that technology can enhance their pedagogical framework		
	Strongly agree	37.14
	Agree	60.00
	Disagree	0.00
	Strongly disagree	2.86
Used publisher provided tools before?		
	Yes	65.71
	No	34.29
Primary reason for using publisher provided digital learning tools		
	Increase efficiencies for me and my students	34.78
	Influence more positive student behaviors	21.74
	More available resources	30.43
	More effective than print alone	13.04
Is your classroom an active learning classroom?		
	Yes	79.41
	No	20.59
Types of assessments typically given in this course		
	Formative only	5.88
	Summative only	11.76
	Formative and summative	82.35
Typical primary method for monitoring student performance		
	Performance on homework assignments	14.71
	Performance on in-class activities/labs	8.82
	Performance on in-class assessments	67.65
	Performance in a learning platform	2.94
	Performance on online assessments	5.88

Table A2. Student background, demographic, and experience characteristics.

		%
Dicipline		
	Biology	9.52
	Calculus	12.24
	Chemistry	39.48
	Economics	38.70
Year in college		
	Dual enrolled	<1
	First	61.80
	Second	19.65
	Third	10.69
	Fourth	3.71
	Fifth	<1
	Other	2.41
Eligible for federal financial aid	Eligible	63.80
First generation	Yes	22.64
Gender		
	Male	45.75
	Female	53.81
	Prefer not to say	<1
Taking the course as disciple requirement	Yes	75.53
Traditionally underrepresented		27.94
HSGPA		3.65
College readiness status		
	College ready	71.93
	Not college ready	28.07
Baseline level of motivation to succeed		
	More	75.61
	Less	24.39
Comfort level with technology		
	Extremely comfortable	18.01
	Comfortable	68.44
	Uncomfortable	12.39
	Extremely uncomfortable	1.15
Level of agreement that publisher provided tools enhance learning		
	Strongly agree	18.73
	Agree	70.97
	Disagree	9.51
	Strongly disagree	<1

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.

Dr. Jeff Bergin
Vice President Learning Research & Design

Jeff has led curriculum, instructional, and learner experience design for various educational technology companies for the past 20 years. Dr. Bergin leads a team of learning researchers and experience designers. In previous roles, he has led the design of personalized and mobile products. Dr. Bergin holds a Ph.D. from Arizona State University and has presented and published on topics including learning design, online learning, and technology augmented instruction.

Dr. Adam Black
Chief Learning Science and Strategy Officer

Adam is a recognized pioneer in improving learner outcomes. From identifying promising areas of learning science, to directing the development of market-leading digital products (used by more than 26 million learners), and spearheading novel approaches for assessing impact, Adam has 24 years of experience. Adam holds a BSc in Physics from the University of Edinburgh and a PhD in Astrophysics from the University of Cambridge, has two patents pending in analytics innovation, and has won national and global awards for digital product innovation.

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**

macmillan learning