RESEARCH STUDY NOTE

LearningCurve

Key Findings:

- Instructors who assigned more LearningCurve activities had higher overall course grades than those who assigned less or no LearningCurve activities.
- Course performance increased as students completed more LearningCurve activities.
- Students who were the first in their family to attend college especially benefited academically from completing more LearningCurve activities.

macmillan learning

BACKGROUND

In order to be successful in higher education, students must have an understanding of what they know and what they don't know. This understanding is formed in large part by feedback they receive in various academic settings. Feedback can be given in a variety of formats including through instructors, classmates, digital tools, or the learner themselves through self-evaluation.

Previous research has found students who receive more frequent, low-stakes assessment with quality feedback also have higher motivation, academic engagement, metacognitive skills, adaptive strategy use, academic performance, and likelihood of transferring their knowledge to new contexts (Cogliano et al., 2020; Han & Finkelstein, 2013; Shute, 2008). This type of assessment, frequently referred to as formative assessment, is a specific type of "assessment for learning"—it's not meant to be summative "proof" of learning, but rather support the learning process. Feedback is a critical component for formative assessment to support learning—it helps learners identify gaps in their knowledge/skill sets (Angelo & Cross, 1993; Bennett, 2010; Black & William, 1998).

Furthermore, mixed assessment methods, including low-stakes assessment, have been identified as an opportunity to provide a more equitable learning environment and help remove barriers to underrepresented minority students (Cotner & Ballen, 2017; Malespina & Singh, 2022). Mechanisms underlying this relationship are still being explored. One possibility is that active-learning pedagogy, which includes prelecture preparation and frequent low-risk assessment, provides a more structured environment for students to engage in frequent problem solving and practice. Previous research found this pedagogy was related to an increase in all students' academic performance in an introductory biology course, but was especially beneficial for underrepresented minority students (Haak et al., 2011). Another possibility is that frequent low-stakes practice helps bolster students' self-efficacy, which leads to increased academic performance, which has a disproportionately greater benefit for underrepresented students (Ballen et al., 2018).

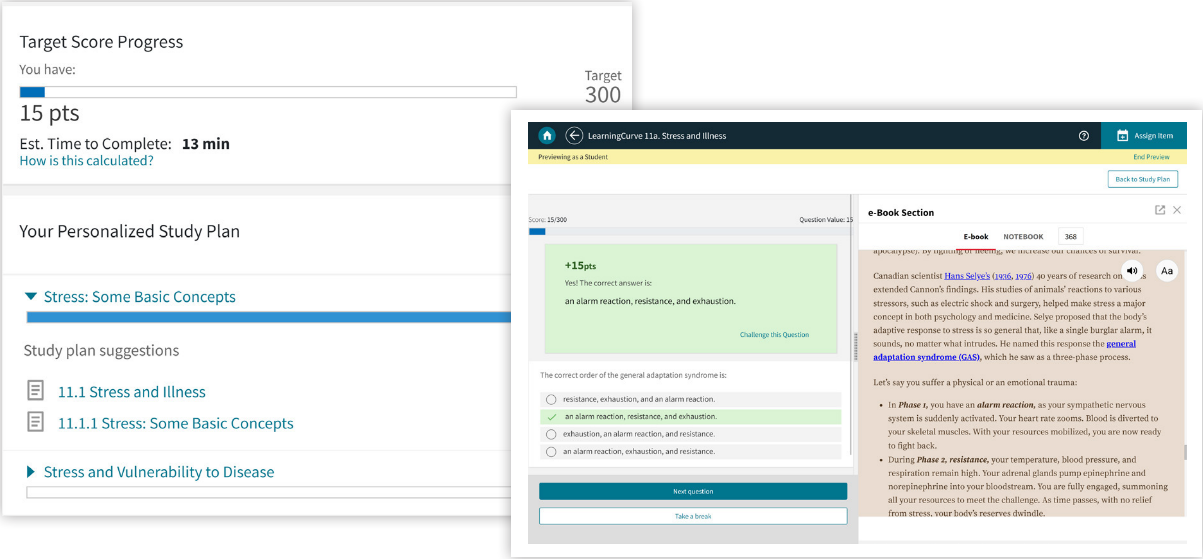
Another possibility is that high-stakes assessments may be more likely to trigger stereotype threat. Stereotype threat occurs when individuals are reminded of being the target of negative stereotypes, which primes a sense of uncertainty in ability (Schmader, 2010). The increased pressure to "demonstrate" ability that's present in high-stakes assessments is more likely to trigger this type of thinking. This can result in expending cognitive effort beyond the task at hand, and thus decreased performance. Three groups of students that have been identified as commonly experiencing stereotype threat in higher education include racial/ethnic minority students, women students (particularly in STEM), and first-generation college students (Dennehy et al., 2018).

Macmillan Learning created LearningCurve adaptive quizzing with the goal of providing students formative assessment with quality feedback that adapts to students' current level of understanding in a low-stakes environment. LearningCurve guides students through a series of quiz questions that adapt to their individual level of understanding of course material, providing feedback whenever students need it, until they reach mastery of a concept/topic.

Macmillan Learning funded a series of research studies, across six semesters (2019-2022) and 161 institutions, to examine the impact of LearningCurve. Participating instructors were given training and implementation recommendations, but use of LearningCurve activities was not required to participate in the study and implementation choices varied by instructor.

PRODUCT OVERVIEW

LearningCurve (LC) was developed in collaboration with active college instructors and based on a wealth of learning science research around three best practice principles: retrieval practice, spacing, and formative assessment with feedback. The act of retrieving information from memory through testing (i.e., retrieval practice) has been shown to lead to better memory of the material than rereading course text or notes (Roediger & Karpicke, 2006). Not only is it important to study and learn by testing yourself, it's also important to space out study sessions/testing. Spreading out study sessions over weeks rather than days has been found to relate to better performance on exams as well as long-term memory retention (Delaney et al., 2010). Formative assessment with feedback was discussed in the background section above.



Based on these research-based principles, LC is an algorithmic quizzing system designed to help students achieve mastery of course material. Students are given multiple-choice and short-answer questions with immediate feedback during and after each question. Students always have access to the relevant ebook section with no loss of points. They can also choose to receive a hint, which lowers the total points earned for that particular question. They can also choose to view the answer with feedback, which doesn't reduce their earned points but also doesn't gain them any points. Students who enter an incorrect answer also get targeted feedback on why that choice was incorrect and are given the opportunity to try again for less earned points. LearningCurve gives a student more questions in an area in which they're experiencing difficulty until the student masters that topic, thus adapting to their current level of understanding.

Mastery does not mean perfection, just that a majority of the questions within a topic are answered correctly. Students must typically earn 150 points in order to show mastery of a topic (displayed as filling in a progress bar). If there are 3 topics in an activity, the target score for that activity is 450 points. Importantly, students can keep trying until they reach mastery, earning full credit for the activity. They also can go back to previously done LC activities at any time during the course to review or use as a study tool.

To help students organize content review, LC offers students a Personalized Study Plan. Students can access this study plan during or after the LC activity. The study plan breaks down their performance on each topic and presents them with suggestions for further practice. Instructors can also assess student performance after they've completed LC activities including time spent on the assignment, number of questions answered, and overall class proficiency by topic.

STUDY DESIGN

Ethics and 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 an accredited, third-party Institutional Review Board organization with no affiliation with Macmillan Learning. Macmillan Learning seeks third-party review to eliminate any bias in the decision of the exemption. The data 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.

Sample

The full study sample included participants spanning across six semesters from fall 2019 through spring 2022. This robust sample included 177 unique instructors teaching 333 courses. Eight different subject areas (chemistry, biochemistry, biology, calculus, precalculus, psychology, economics, and English) are represented in the study. Instructors came from 161 institutions across the United States and Canada. The sample included a range in institution and course sizes as well as course formats (i.e., face-to-face, virtual synchronous, virtual asynchronous).

The variation in participating institutions and instructors enabled a diverse student sample. The full student sample included 40% non-White or Asian, 24% who were first in their families to go to college, 66% who were eligible for financial aid, and 38% who had a high school GPA lower than 3.5 across a total of 9,803 participating students.

Methods

After consenting to be part of the larger research study, participating instructors were given brief training on LC, practical information on where to locate and assign LC to students within the Achieve program and its functionality. The LC training was included as part of a broader training on Achieve given by a curriculum specialist and lasting approximately 45 minutes. If instructors requested follow-up training on any feature within Achieve, additional training was given. Participating instructors were not required to assign LC to their students, enabling a more naturalistic implementation. Use of LC was, however, observed and documented by the research team.

Students who consented to participate in the study granted researchers access to their course performance data as well as their LC usage data. Furthermore, as part of participating in the study, students were asked to complete two additional surveys (beginning and end of semester) to share sociodemographic information, as well as general perceptions of Achieve

DATA ANALYSIS

Statistical modeling (i.e., linear mixed model) was used to isolate the unique impact of completing LearningCurve on student course performance (overall course grade and exam average grade as percentages). In order to partial out the unique impact of LC, several factors were included in the model to control for other variables researchers thought would likely impact academic performance. The variables were:

- subject,
- course mean grade,
- course mean exam average,
- student college readiness (i.e., high school GPA, SAT/ACT scores),
- student financial aid eligibility,
- student gender,
- student race/ethnicity,
- first generation college student status,
- number of LC assignments assigned within a course,
- and number of LC assignments completed by individual students.

Including these variables in the model was an attempt to equate students on background variables, prior academic performance, and current academic setting in order to bolster the argument that the impact of LC is not simply a reflection of “better” students completing more LC activities.

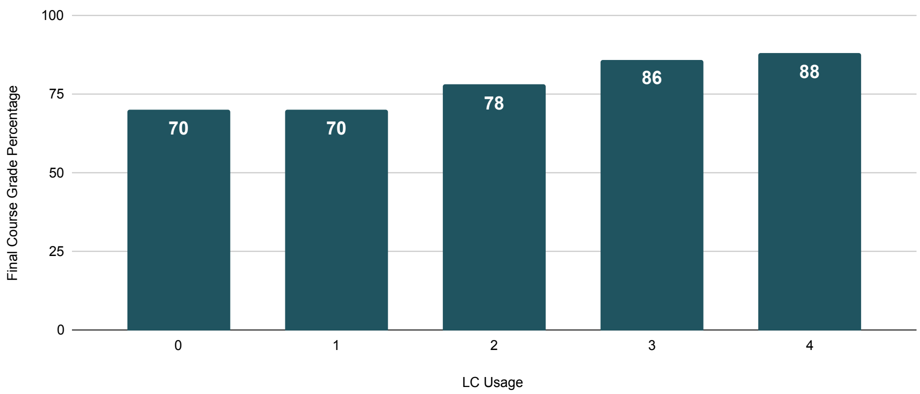
RESULTS

Course-level Analysis

Course-level analyses included all courses participating in the research studies whether they chose to assign LC activities or not. This analysis was focused on implementation patterns—how many LC activities should be assigned to see impacts on student academic outcomes? How do those outcomes compare to courses who didn't assign any LC activities? Because different subjects (e.g., psychology vs. biology) have different numbers of available LC activities, LC assignment was grouped by the number of LC activities assigned as a percentage of the maximum number available within that subject (e.g., Psychology had 40 available LC whereas Precalculus had 12).

Figure 1 displays the results of analyzing the impact of assigning LC on courses' overall average grade (as a percentage). As seen in the figure, the LC assigned category was significantly related to courses' average final grade. Courses who assigned a higher percentage of LC activities also had higher average course grades. This was particularly apparent when comparing group 2 and above (courses who assigned 25% or greater of the LC activities available) to groups 0 and 1 (those courses who didn't assign any LC activities or less than 25%). Courses in group 2 and above had an increase of about 3 percentage grade points, which was statistically significant.

Figure 1. Course Grade Performance by LC Assignment Category



Note
The LC usage categories were defined as follows:

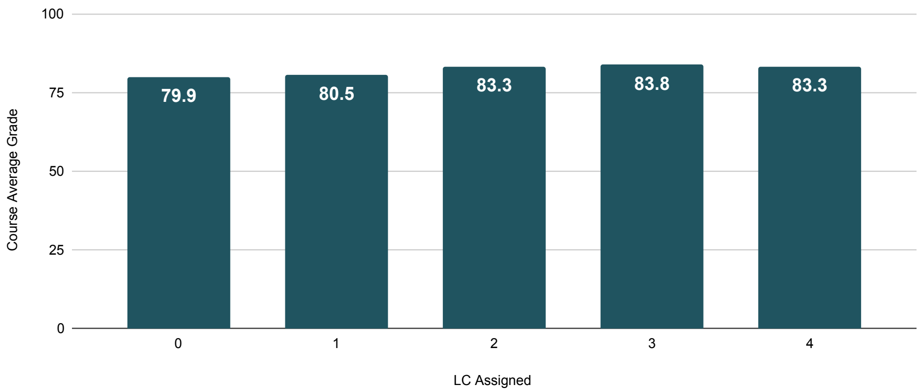
0– No LC completed
1– Less than 50% of the overall median completed,
2– 50%-99% of the overall median completed,
3– 100% of the overall median completed,
4– Greater than the median.

Student-level Analysis

Student-level analyses focused on those courses that assigned at least one LC assignment. This analysis was focused on student completion patterns—how many LC activities need to be completed for students to see the benefits? Figure 2 below displays the results of analyzing the impact of LC usage on final course grade (as a percentage). LC usage was grouped by students' completion of the median number of LC activities completed in their course in order to control for implementation/subject differences. The median was used as an indicator of what was typical or expected within a particular course.

As seen in the figure, LC usage category was significantly related to students' final course grade. Students completing a higher percentage of LC activities also earned higher course grades. Students who completed at the median or greater than the median number of LC in their course (groups 3 and 4) had an increase of approximately 16 grade percentage points compared to students who completed less than 50% of the median (groups 0 and 1).

Figure 2. Student Final Grade Performance by LC Usage Category



Note
The LC assigned categories were defined as follows:

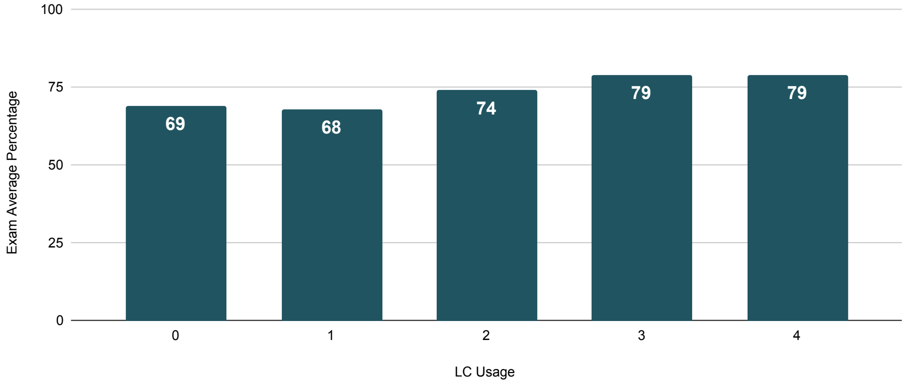
0–No LC assigned
1–Less than 25% of max,
2–25%-49% of max,
3–50-75% of max,
4–Greater than 75% of max.

Many participating study instructors included Achieve activity performance, including LC completion, into students' final grade calculation. Therefore, it was important to validate the effect of LC on student learning by examining another outcome. Researchers were given access to students' course exam average (as a percentage) by instructors for consenting students. Since exams were not directly tied to Achieve nor LC, this would provide evidence that LC impacted student learning outcomes more generally.

Figure 3 below displays the results of analyzing the impact of LC usage on exam average (as a percentage). Similar to the course grade analysis, LC usage was grouped by students' completion of the median number of LC activities completed in their course.

As seen in the figure, the LC usage category was significantly related to students' average exam grade. Students completing a higher percentage of LC activities also earned higher average exam grades. Students who completed at the median or greater than the median number of LC in their course (groups 3 and 4) had an increase of approximately 10 percentage grade points compared to students who completed less than 50% of the median (groups 0 and 1).

Figure 3. Student Exam Performance by LC Usage Category



Note
The LC usage categories were defined as follows:

0– No LC completed
1– Less than 50% of the overall median completed,
2– 50%-99% of the overall median completed,
3– 100% of the overall median completed,
4– Greater than the median.

Interaction with First Generation College Student Status

The researchers were also interested in how use of LC could be particularly beneficial for underrepresented groups and students experiencing barriers to college success. The low-stakes environment of LC may enable students to practice and test their knowledge, increasing their self-efficacy without triggering stereotype threat. To do so, use of LC was examined as an interaction with gender, race/ethnicity, and first generation college student status to see if LC had an even greater benefit for certain students.

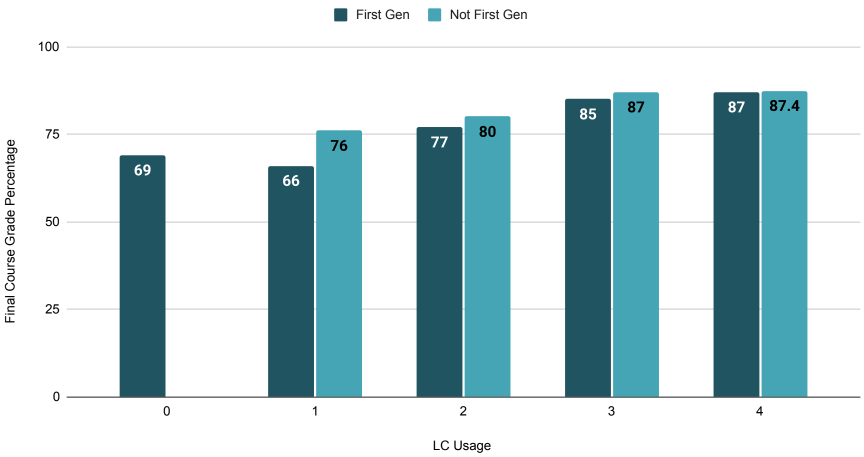
An interaction occurs when the impact of one variable on an outcome depends on another variable. For example, if we were examining the effect of a particular teaching method on student learning that teaches a math concept within the context of baseball, we'd likely see an interaction with students' prior knowledge of baseball. For students who had prior knowledge of baseball, their learning may be disproportionately greater than those students who didn't have prior knowledge of baseball. Therefore the effect of one variable (instruction) on an outcome (math learning) depends on another variable (prior knowledge of baseball).

There was not a significant interaction between LC usage and gender nor race/ethnicity in predicting student outcomes (course grade and exam grade), meaning the impact of using LC was similar across these groups. There was a significant interaction between LC usage and first generation college student status, meaning the impact of using LC was different for these students.

Figure 4 displays the results of analyzing the interaction between LC usage and first generation college student status on final course grade. Increased usage of LC was particularly beneficial for first generation college students' final grade compared to non-first generation college students.

As seen in the figure, as LC usage increases, the differences in final grade between first generation students and non-first generation students is significantly reduced. When first generation students are completing less than 50% of the median LC in their course (group 1), their final grade is about 10 percentage grade points lower than their non-first generation peers who similarly completed less than 50%. However, when they complete 100% of the median or greater (groups 3 and 4), that difference is no longer statistically significant and reduced to 2 percentage grade points or less.

Figure 4. Course Grade Performance by LC Usage Category X First Generation Status



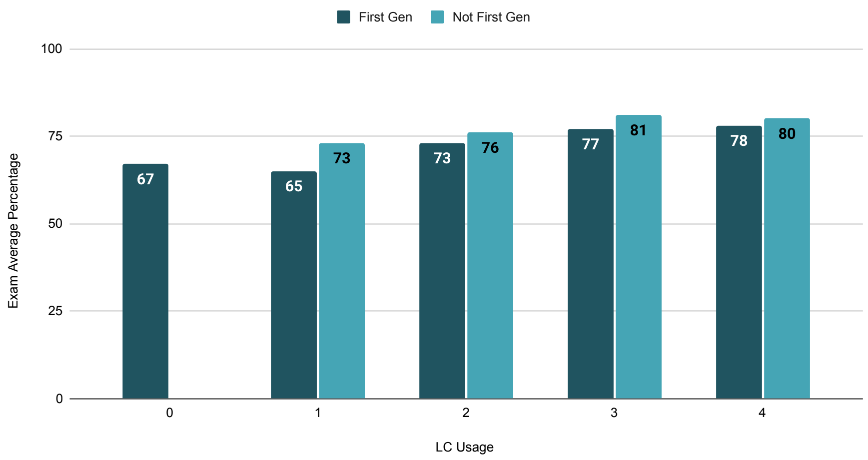
Note
The LC usage categories were defined as follows:

0– No LC completed *(only First Generation students fell into this category)
1– Less than 50% of the overall median completed,
2– 50%-99% of the overall median completed,
3– 100% of the overall median completed,
4– Greater than the median.

This analysis was replicated for average exam grades. Figure 5 displays the results of analyzing the interaction between LC usage and first generation college student status on average exam grade. Similarly, there was a significant interaction in which increased usage of LC was particularly beneficial for first generation college students' average exam grade compared to non-first generation college students. As LC usage increases, the differences in exam average grade is significantly reduced.

When first generation students are completing less than 50% of the median LC in their course (group 1), their average exam grade is about 8 percentage grade points lower than their non-first generation peers who also completed less than 50%. However, when they complete 100% of the median or greater (groups 3 and 4), that difference is no longer statistically significant. When they complete greater than the median LC assigned (group 4), that difference is reduced to just 2 percentage grade points. The reduction in the difference in performance between first generation and non-first generation students is often referred to as bridging the "equity gap", which is critically important so that all learners have similar opportunities to experience success.

Figure 5. Course Exam Performance by LC Usage Category X First Generation Status



Note
The LC usage categories were defined as follows:

0–No LC completed *(only First Generation students fell into this category)
1–Less than 50% of the overall median completed,
2–50%-99% of the overall median completed,
3–100% of the overall median completed,
4–Greater than the median.

IMPLICATIONS FOR INSTRUCTORS

Overall, the research findings suggest a benefit of LearningCurve on student learning and academic outcomes. Instructors who assigned a higher percentage of available LC activities also saw higher average course grades. Furthermore, students who completed more LC activities compared to what was typical in their course (median) had higher course grades and exam grades. Lastly, students who were the first in their family to go to college especially benefited from completing more LC activities. Completing more than what was typical in their course (median) led to an increase in about 10 percentage grade points for both their final grade and exam grade.

Instructors can use LearningCurve to help students gain mastery in their course topics in a low-stakes and supportive environment. This can be used in combination with other active-learning pedagogy techniques to help bolster the potential benefits. For example, another identified active-learning strategy is pre-lecture preparation. Instructors could assign LC activities strategically, assigning content/topics that are being covered in the upcoming class, thus using it as an opportunity for “pre-lecture preparation”.Based on the previous results, instructors should assign at least 25%-50% of available LC activities within their subject area to see gains in course grades. Students should complete at least what is typical in their course (median) or above to see gains in their course performance. Furthermore, while this was not tested directly as part of the current research study, instructors who use the available information on students’ LC performance in order to identify gaps in students’ knowledge, as well as a signal to adapt instruction, provide particular resources, or enact interventions, might see an even greater benefit of LC.

DISCUSSION

The current work examined the effect of LearningCurve adaptive quizzing within Achieve across a diverse sample of institutions, instructors, and students. While arguments of causality cannot be made without further experimental research, the results indicate that assigning and completing LC activities was significantly related to students' performance in their courses. The opportunity to practice and test their knowledge in a low-stakes environment was positively related to course performance. This relationship was particularly powerful for students who were the first in their family to go to college.

The low-stakes environment, quality feedback, and resources that LC provides may have enabled first generation students to learn from their mistakes and identify misconceptions. This may have led to increased self-efficacy in the related course topics. Self-efficacy has been found to relate to academic performance above and beyond prior knowledge and be particularly important for underrepresented groups (Ballen et al., 2018; Elias & MacDonald, 2007). The LC environment may also have been less likely to trigger stereotype threat, allowing students to focus all cognitive effort on the task at hand. Future research should continue to explore potential mechanisms underlying this relationship.

LIMITATIONS AND FUTURE RESEARCH

While the current work represented a large and diverse sample, a convenience sample was used. This was not a true experiment with random assignment. A multitude of variables were used to serve as statistical controls, but the lack of random assignment is a limitation. Individual differences of students not captured by the variables used as controls cannot be ruled out as potential confounding variables.

Instructors' implementation of LC activities was also not controlled. Instructors were free to use LC as much or as little as they deemed necessary. Furthermore, instructors' use of low-stakes assessment or active-learning strategies more generally outside of Achieve and LC was not measured or controlled. Some instructors may incorporate these strategies more or less in their instruction, which may affect the impact of LC.

Future experimental studies could test the impact of LC by randomly assigning students within the same course or instructor to either receive LC activities or not. This design would help strengthen arguments of causality by ruling out both individual differences and instructional differences as potential explanations for group differences. Future research could also incorporate qualitative methods to complement the quantitative analyses. Qualitative methods such as observations of instruction and in-depth interviews with instructors and students could help identify pedagogical patterns including low-stakes assessment/active-learning. A more complete understanding of how instructors are incorporating these strategies into their courses can help clarify how LearningCurve can complement and bolster instruction.

REFERENCES

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017). Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning. CBE—Life Sciences Education, 16(4), ar56.

Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning (CBAL): A preliminary theory of action for summative and formative assessment. Measurement, 8(2-3), 70-91.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: principles, policy & practice, 5(1), 7-74.

Cogliano, M., Bernacki, M. L., & Kardash, C. M. (2021). A metacognitive retrieval practice intervention to improve undergraduates' monitoring and control processes and use of performance feedback for classroom learning. Journal of Educational Psychology, 113(7), 1421.

Cotner, S., & Ballen, C. J. (2017). Can mixed assessment methods make biology classes more equitable?. PLoS One, 12(12), e0189610.

Delaney, P. F., Verkoeijen, P. P., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. Psychology of learning and motivation, 53, 63-147.

Dennehy, T. C., Smith, J. S., Moore, C., & Dasgupta, N. (2018). Stereotype threat and stereotype inoculation for underrepresented students in the first year of college. In R. S. Feldman (Ed.), The first year of college: Research, theory, and practice on improving the student experience and increasing retention (pp. 309–344). Cambridge University Press.

Elias, S. M., & MacDonald, S. (2007). Using past performance, proxy efficacy, and academic self-efficacy to predict college performance. Journal of Applied Social Psychology, 37(11), 2518-2531.

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. Science, 332(6034), 1213-1216.

Han, J. H., & Finkelstein, A. (2013). Understanding the effects of professors' pedagogical development with Clicker Assessment and Feedback technologies and the impact on students' engagement and learning in higher education. Computers & Education, 65, 64-76.

Malespina, A., & Singh, C. (2022). Gender differences in test anxiety and self-efficacy: why instructors should emphasize low-stakes formative assessments in physics courses. European Journal of Physics, 43(3), 035701.

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. Psychological science, 17(3), 249-255.

Schmader, T. (2010). Stereotype threat deconstructed. Current Directions in Psychological Science, 19(1), 14-18.

Shute, V. J. (2008). Focus on formative feedback. Review of educational research, 78(1), 153-189.

Thomas, A., & Cross, P. "Minute paper." Classroom assessment techniques: A handbook for college teachers, 2, 148-153.