Lessons Learned From Early Implementations of Adaptive Courseware

April 2016

SRI Education™
A DIVISION OF SRI INTERNATIONAL

Contract #24133, Work Order #27; Contract #28981, Work Order # 29
SRI Project Nos. 21987 and 22997

Prepared for:

Yvonne Belanger
Senior Program Officer
Measurement, Learning and Evaluation
Education, Postsecondary Success, U.S. Programs
Bill & Melinda Gates Foundation
P.O. Box 23350
Seattle, WA 98102
yvonne.belanger@gatesfoundation.org

Submitted by:

Louise Yarnall, Barbara Means, and Tallie Wetzel
SRI Education
333 Ravenswood Avenue
Menlo Park, CA 94025
Phone: 650-859-3181

Contents

Executive Summary

To address the urgent need to improve student outcomes in developmental and general education courses, higher education institutions are turning to new learning technologies. Prominent among these is adaptive learning courseware that uses computer algorithms to parse learning analytic data collected as students interact with online learning environments. These adaptive algorithms then determine the student's next learning activity and can be used to generate personalized feedback, study reminders, content recommendations, and real-time progress dashboards that both students and instructors may review.

The Bill & Melinda Gates Foundation initiated the Adaptive Learning Market Acceleration Program (ALMAP) to advance evidence-based understanding of how adaptive learning technologies such as adaptive courseware could improve opportunities for low-income adults to learn and to complete postsecondary credentials. Over three academic terms, from summer 2013 through winter 2015, the ALMAP grant program provided 14 higher education institutions with seed funding to incorporate nine adaptive learning products into 23 courses and to conduct quasi-experiments (side-by-side comparisons with comparison groups) to measure their effects on student outcomes, and to gather data on cost impacts and instructor and student satisfaction.

The foundation asked grantees to use the adaptive technologies to address two common hurdles to college completion: success and mastery in gateway general education courses and developmental (remedial) courses. These courses were chosen as targets for improvement because student success in gateway general education courses and developmental courses paves the way to persistence and success in the first two years of college, a time when many students fall off track for completing an associate's or bachelor's degree. The foundation asked grantees to evaluate the impact of their adaptive courseware implementations on student learning, course grades, and probability of course completion. Additionally, in view of trends showing college costs outpacing general inflation since 1979,[1] the foundation directed grantees to explore both the costs and potential savings associated with adaptive courseware implementation.

To strengthen the consistency and credibility of the evidence grantees gathered, the foundation also contracted with SRI International to aggregate and analyze the ALMAP data. SRI's evaluation looked at the learning and cost impacts of adaptive courseware implementations both separately and together. We did not expect to find one answer for the many forms of adaptive instruction and diverse institutions of higher education in the United States, but the ALMAP study attempts to provide information that will help local decision makers identify approaches that might work for them.

SRI assembled learning impact, cost, and satisfaction findings across the portfolio of ALMAP grantee product evaluations. This synthesis of product evaluations encompassed data collected from over 19,500 unique students in classes taught by more than 280 unique instructors. The resulting ALMAP evaluation report provides a glimpse into the state of the art of adaptive learning courseware implementations across the range of U.S. institutions of higher education—from research universities to colleges focused on undergraduate education, and from public community colleges to private online colleges.

[1] *Bureau of Labor Statistics. (2010, September). Back to college: Spotlight on statistics. Washington, DC: Bureau of Labor Statistics. http://www.bls.gov/spotlight/2010/college/*

Portfolio Description

The ALMAP grantees implementing one or more of the nine adaptive learning products were 10 bachelor's degree programs and 4 associate's degree programs. ALMAP grantees used adaptive courseware in 15 gateway general education courses and 7 developmental education courses. The gateway course subjects in which adaptive courseware was used included psychology, biology, business, marketing, and economics. The developmental education courses focused on the mathematics or English language arts proficiencies needed to succeed in college-level work.

ALMAP grantees used the adaptive courseware to make different kinds of changes in course delivery. SRI organized these changes into three use case categories:

- Blended Adaptive vs. Lecture— adaptive courseware was used as part of a shift from traditional lecture to blended instruction;
- Online Adaptive vs. Online— adaptive courseware was used as an enhancement to existing online courses; and
- Blended Adaptive vs. Blended—adaptive courseware was swapped into face-to-face courses already using blends of classroom-based and online approaches to support learning.

Key Findings

In reviewing the learning, cost, and satisfaction outcomes for the ALMAP portfolio, it is important to keep in mind the relative immaturity of the field of adaptive learning technology. Both technology capacity and ways to support instruction and learning with technology are evolving rapidly, and these results should be viewed as snapshots in time. Nevertheless, the impact estimates included in SRI's synthesis all passed screens for adherence to generally accepted research standards, and the resulting data set is one of the largest of its kind for commercially available adaptive courseware products. Major findings were as follows:

EFFECTS ON STUDENT LEARNING AND COURSE COMPLETION

- Some adaptive courseware implementations (4 of the 15 with a data set adequate for analysis) resulted in slightly higher average ***course grades***, but the majority had no discernible impact on grades.
- Overall, in the 16 grantee-provided data sets appropriate for estimating courseware impacts on ***course completion***, the odds of successfully completing a course were not affected by the use of adaptive courseware.
- Only seven controlled side-by-side comparisons of scores on common ***learning assessments*** were available; the average impact of adaptability for these seven was modest but significantly positive.
- The impacts of adaptive courseware varied by use case:
  - Switching from a lecture format to adaptive blended instruction had a positive impact on student learning as measured by posttests.
  - Moving from nonadaptive to adaptive learning systems in fully online courses had a small positive effect on course grades.
  - There were too few cases contrasting blended adaptive versus blended nonadaptive courses to draw any conclusions about impacts.
  - None of the use case analyses found a significant average impact on course completion rates; only 2 of the 16 side-by-side comparisons of completion rates found a significantly positive impact on the odds of course completion.

- Courseware products with adaptivity at a micro level (individual lesson or learning object) produced stronger student outcomes than those with adaptivity at a more macro level.
- The size of the adaptive courseware effect did not vary significantly for different academic disciplines, but impacts appeared to be larger for mathematics and biology courses, a trend that might have attained statistical significance with a larger sample.
- Impacts for Pell Grant students were equivalent to those for students overall. The ALMAP study provided a limited amount of data on the impact of adaptive courseware on outcomes for Pell Grant (low-income) students, but the data that were available provided no indication that adaptive courseware was either more nor less advantageous for Pell Grant students. This finding contrasts with some previous studies that have suggested that online and blended learning approaches put low-income students at a disadvantage.

Effects on Course Costs

- Comparisons of per-student costs for both adaptive courseware and comparison versions of a course found that in most cases costs went up during the first term of adaptive courseware implementation. Course costs are driven largely by instructor labor, and a number of the adaptive products were platforms into which instructors inserted content.
- The adaptive courseware was associated with lower ongoing costs in 7 of the 10 cases with cost data for second and third implementations of adaptive courseware.
- There were only eight cases for which we had both learning impact estimates and comparative cost data. Five of these eight cases had reduced costs but only one of those five produced improved learning outcomes at the same time. In the other four cases of cost reduction, there was no significant change in learning outcomes.

INSTRUCTOR AND STUDENT PERCEPTIONS OF ADAPTIVE LEARNING EXPERIENCES

- Among ALMAP adaptive courseware instructors who responded to SRI's survey, 74% reported satisfaction with the adaptive courseware they used.
- More developmental course instructors (67%) than gateway general education course instructors (49%) planned to use adaptive courseware in the future.
- In short written responses, instructors endorsed the adaptive courseware's real-time progress dashboards as useful for informing their teaching.
- The major concern expressed by instructors was getting students to use the adaptive courseware frequently enough.
- SRI's analysis of responses to student surveys administered by ALMAP grantees indicated that 2-year college students had more positive views of the adaptive courseware than did students at 4-year colleges and universities.
  - Around half (51%) of 4-year college students, compared to over three-quarters (77%) of 2-year college students, reported that they had made positive learning gains with the adaptive courseware.
  - In addition, 56% of 2-year college students reported satisfaction with their adaptive courseware experience compared with only 33% of bachelor's degree students.
- Developmental course students reported higher rates of engagement (60%) and learning gains (95%) with the courseware than did gateway course students (25% and 35%, respectively).

Implications for Future Work

Future research into blended learning technology implementation and efficacy in higher education is badly needed in a field awash in marketing claims. The ALMAP evaluation showed that adaptive courseware is not a homogeneous category that can be established as "effective" or "ineffective." The diversity of products, use cases, and outcomes for the ALMAP implementations yielded important lessons for postsecondary institutions and faculty, courseware suppliers, researchers, and funders.

FOR POSTSECONDARY INSTITUTIONS

- **Postsecondary institutions planning large-scale adoptions of adaptive courseware should conduct their own internal analyses of student outcomes with that courseware compared to other alternatives.** Even when the identical adaptive courseware product is used in different institutions or academic terms, learning outcomes can differ markedly depending on how it is used. To provide a valid answer to questions about the relative effectiveness of the adaptive courseware, these analyses need to establish the baseline equivalence of students in the courseware and comparison course sections being compared.

- **Baseline equivalence is essential for justifying claims about courseware effects,** but the common practice in higher education institutions is to simply compare course success rates without any data on student characteristics (baseline equivalence). ALMAP analyses found that student characteristics and prior learning often vary markedly from course section to section and across terms within the same institution.

- **Adaptive courseware in unlikely to reduce per-student course costs in its initial implementation.** Even when a college does not need to make infrastructure enhancements to accommodate the courseware, there are costs entailed for instructors learning how to use the courseware and in many cases, for instructor insertion or modification of content. First-term cost reductions were observed in only a minority of cases, most of which involved vendor-developed rather than instructor-generated content.

- **Subsequent implementations of adaptive courseware have stronger prospects for cost reductions.** If the same courseware product is implemented for multiple terms, costs often drop considerably from those incurred in the initial term. Institutional decision making around large-scale courseware adoption, should incorporate multi-year projections of both costs and impacts (such as reduced need for repeating developmental courses or reductions in attrition).

FOR POSTSECONDARY INSTRUCTORS

- **Instructors in 2-year colleges and those teaching developmental education courses would do well to consider adaptive courseware options.** The ALMAP evaluation found only minor enhancement of course grades associated with adaptive courseware, but results for the relatively few cases with direct measures of student learning (posttests), were encouraging. Moreover, instructors and students in 2-year colleges and developmental courses reported high levels of satisfaction with their adaptive courseware experiences.

- **Adoptions of adaptive courseware in 4-year institutions should include planning to make sure the courseware's benefits are apparent to students.** Only a third of the students in 4-year colleges responding to surveys about their experiences with adaptive courseware expressed overall satisfaction. It is not clear from the available data whether qualities of the courseware they were using, the way in which the courseware was implemented, or a sense of loss of instructor personal attention was most responsible for their lack of enthusiasm. Instructors of 4-year college gateway courses are encouraged to attend to all of these issues when transitioning to use of adaptive courseware.

- **Instructors can make valuable contributions to student success by sharing their insights about adaptive courseware** with each other and with the field more broadly, including with vendors. Course instructors are in a good position to understand more about sources of college students' satisfaction and dissatisfaction with adaptive courseware as well as ways in which it could better support learning. It is quite possible that courseware designers or instructors can improve courseware effectiveness and quality by changing courseware settings or implementing courseware somewhat differently.

FOR SUPPLIERS OF ADAPTIVE COURSEWARE

- **Courseware providers can leverage partnerships with their higher education clients to obtain better data for use in product improvement.** Courseware developers already analyze the use data available from their learning systems, but typically do not have other important kinds of data regarding how their courseware is being used, student characteristics, and student outcomes. Courseware vendors should seek opportunities to partner more deeply with their field sites to better understand the aspects of course implementation that influence the learning and cost impacts of their products. Data-driven insights can inform continuous improvement cycles so that courseware quality and user experience are enhanced over time.

- **Courseware providers can work with their institutional partners and researchers to articulate and validate implementation guidelines.** As effectiveness and user satisfaction data are accumulated across multiple implementations in different institutional contexts, those data can be analyzed to derive empirically based recommendations for how the courseware should be used, including the total amount of time and spacing of use.

- **Courseware software developers can make sure that it is easy to pull user data revealing the key interactions students have with the courseware.** Courseware log file data can reveal areas where students appear to get stuck, whether all of the major components of the courseware are being used, and indications that students are trying to bypass learning opportunities in order to move through the courseware as quickly as possible. Student "click stream" data need to be aggregated to a level that is interpretable by instructional designers, researchers, and faculty. It should be possible to query and export data for all users within a given class within a given period of time.

FOR RESEARCHERS AND EVALUATORS

- **Analyses of adaptive courseware effectiveness should take into account the specifics of the way in which the courseware was used in a particular implementation.** Researchers should help courseware users understand that learning efficacy is not a trait of a product per se or simply a matter of matching the right product to the right subject matter. Rather, multiple factors affect learning outcomes and to make sense of student outcomes, analyses need to incorporate student characteristics, specifics of how the adaptive courseware is used, aspects of the course beyond the courseware product, and the way learning is measured to make sense of student outcomes.[2]

---

[2] *Means, B., Bakia, M., & Murphy, R. (2014). Learning online: What research tells us about whether, when and how. New York: Routledge.*

- **Course grades and course completion rates are less than ideal as outcome measures for efficacy studies.** These measures reflect more than what students have learned or can do; in many cases, they also reflect student attendance, timely completion of assignments, and class participation. More precise measures of learning, such as common tests, and other more detailed behaviors, such as student assignment completion, should be tracked as outcomes in evaluations of the impact of introducing new kinds of courseware. Especially when comparing outcomes for course sections taught by multiple instructors with different or unknown grading policies, using these measures as student outcomes can bias impact estimates. However, grades and course credits are critical for students and to inform institutional policies so they should be captured in evaluation studies along with more direct measures of student learning.

- **Satisfaction surveys, whether of instructors or students, are insufficient as the only outcome in a courseware evaluation.** Instructor and student perceptions as expressed on surveys can be useful information, but positive survey responses and impacts on learning do not always go hand-in-hand. Moreover, these surveys are dogged by low response rates and selection bias. Use of more frequent but very brief surveys embedded in the courseware could help raise response rates and lower concerns about sampling bias, but a certain amount of judicious observation and interviewing is recommended to provide insights to complement learning data and reports from survey samples.

- **More research is needed to develop cost effective ways to capture the changes in instructional practice associated with implementations of adaptive courseware.** The ALMAP evaluation focused on time devoted to lecture and presentations, but future work should examine (1) how adaptive courseware affects the relative balance between low-level and high-level content interactions between instructors and students and (2) how the automated dashboards in adaptive courseware affect instructors' sensitivity to individual and whole-class learning needs.

FOR FUNDERS

- **Funders seeking to contribute to the knowledge base around courseware effectiveness should foster the use of controlled studies of courseware impacts.** Only 16 of the 64 implementations of courseware as part of ALMAP generated any adaptive courseware and comparison course section data incorporating measures to control for any pre-existing differences between the student groups being compared. The ALMAP evaluation revealed that even higher education institutions with a commitment to innovation and adoption of adaptive learning software are, by and large, unaccustomed to performing controlled studies of course outcomes. Further, adoption of common analytic approaches and common learning outcome measures for developmental and gateway courses would support aggregation of findings across institutions.

- **Postsecondary institutions can support student success by sharing anonymized data sets from side-by-side comparisons of adaptive courseware and other instructional approaches.** As more and more institutions turn to adaptive courseware, sharing data sets linking information about implementation, student administrative data, courseware system data, and student outcomes can build the empirical base needed to answer many of the questions left unanswered by the ALMAP data sets. Linked data sets need to be screened to insure they do not contain personally identifying information, but FERPA (Family Educational Rights and Privacy Act) compliant processes are readily available.

- **Grantees should be encouraged to develop unit-cost formulas that permit fair comparisons between adaptive courseware and comparison course sections.** Per-student costs depend greatly on the size of each course section, and major differences in class size (such as the transition from large lecture classes to blended instruction) can overwhelm any efficiencies or inefficiencies related to courseware per se. Cost analyses should not narrowly focus on whether costs increase or decrease, but rather should take into account the tradeoffs between changes in per-unit cost associated with a unit of learning gain.

- **Funders should encourage modeling of cost effectiveness over the longer term, not just for an individual course.** It is necessary to take a 3-5 year perspective on student outcomes to capture the monetary savings associated with lower odds of needing to retake a developmental course or higher persistence rates. It is also useful to separate the cost ingredients of up-front development, preparation, and infrastructure for initial implementation from those of ongoing instructional delivery.

In conclusion, although ALMAP grantees varied in their capacity for institutional research, the ALMAP grantees made good-faith efforts to address the challenge of measuring learning impacts, costs, and student and instructor perceptions. We acknowledge them not only for their leadership in innovating with new courseware but also for their willingness to share their experiences and findings with the field at large. We view this sharing of results as a first step toward cultivating a critical community of inquiry around the usefulness of new courseware in supporting student learning and course completion in institutions of higher education.

Introduction

U.S. institutions of higher education are facing unprecedented pressure to produce more college graduates in order to increase job attainment and reduce income inequality.[3] At the same time, media headlines routinely report soaring tuition costs, particularly at public institutions for which public tax subsidies have declined substantially over the past decade.[4] As institutions shift more program costs to students and their families, the public perception of higher education has shifted as well. Higher education is no longer considered primarily as an exercise in self-exploration for a privileged few. Instead, it is viewed as a strategic investment in obtaining a credential that will lead to gainful employment in an ever more competitive global job market. Policymakers argue that production of basic college credentials—workforce certificates, associate's degrees, and bachelor's degrees—needs to expand because a greater proportion of middle class jobs will require the specialized knowledge gained primarily through higher education.[5] This higher education cost-benefit discussion plays out against historic rates of debt among college graduates.[6]

As one strategy for addressing these demands, institutions of higher education are turning to emerging technologies. Adaptive courseware systems use computer algorithms that parse learning analytic data collected from online student learning environments and then use these data to generate automated feedback for quizzes, study reminders, content recommendations, and real-time progress dashboards that both students and instructors may review.[7] Adaptive courseware moves students through course content at a pace tailored to each individual.

As higher education leaders seek ways to improve student outcomes, they also are considering return on investment (ROI). Higher education consultants supporting innovation[8] frame the financial discussion as first understanding the types of investments needed for an innovation to improve an institution's student success rate and second as understanding the levers that impact the cost per educational unit (e.g., course, student, or student credit hour) and the resulting ROI.

Over three academic terms, from summer 2013 through winter 2015, the Adaptive Learning Market Acceleration Program (ALMAP) sponsored by the Bill & Melinda Gates Foundation provided 14 early-adopting higher education institutions with seed funding to use nine adaptive learning

---

[3] *Dynarski, S. (2008). Building the stock of college-educated labor. Journal of Human Resources, 43(3), 576-610.*

[4] *Desrochers, D. M., & Kirshstein, R. J. (2012). College spending in a turbulent decade: Findings from the Delta Cost Project. Washington, DC: American Institutes for Research. Johnstone, D. B. (2004). The economics and politics of cost sharing in higher education: Comparative perspectives. Economics of Education Review, 23(4), 403-410.*

[5] *Carnevale, A., Smith, N., & Strohl, J. (2010). Help wanted: Projections of jobs and education requirements through 2018. Washington, DC: Center on Education and the Workforce. Retrieved from http://cew.georgetown.edu/JOBS2018*

[6] *Baum, S., Cardenas Elliott, D., & Ma, J. (2014). Trends in student aid 2014. Princeton, NJ: The College Board.*

[7] *Education Growth Advisors. (2013). Learning to adapt: Understanding the adaptive learning supplier landscape. Boston, MA: Tyton Partners.*

[8] *Staisloff, R. (2013, March/April). How to review your business model: Best practices. Trusteeship Magazine. Retrieved from http://agb.org/trusteeship/2013/3/how-review-your-business-model-some-best-practices*

products in 23 courses and to conduct quasi-experiments to measure the impacts on student outcomes, such as learning, course grade, completion, and enrollment persistence. The foundation intended for ALMAP to advance evidence-based understanding of how adaptive technologies can improve opportunities for low-income adults to learn and complete postsecondary credentials. The foundation asked grantees to focus their implementation of adaptive technologies on two common hurdles to college completion: gateway general education courses and developmental (remedial) courses. Additionally, in view of trends showing college costs outpacing general inflation since 1979,[9] the foundation asked grantees to explore the costs and potential savings associated with adaptive courseware implementation. Finally, the foundation also asked the grantees to describe how instructors and students liked using the courseware.

To bolster the strength of the evidence the grantees gathered, the foundation contracted with SRI International to study ALMAP. To frame the evaluation, we hypothesized that if adding adaptivity improves student learning, particularly for those at risk of failure, this would be a strong benefit, especially if it could be done without increasing costs. If adding adaptivity results in equivalent learning outcomes but saves money, the use of adaptive courseware would also be beneficial. Finally, if adding adaptivity results in equivalent learning outcomes and costs, but instructors and students find it satisfying, there would still be some benefit. We did not expect to find one answer for the many forms of adaptive instruction and diverse institutions of higher education in the United States. Rather, in the ALMAP study SRI attempted to provide information that will help local decision makers find approaches that work for them.

SRI assembled the findings across the portfolio of ALMAP grantee product evaluations. This synthesis encompassed data collected from more than 19,500 unique students in classes taught by more than 300 unique instructors. In assembling the data, SRI followed criteria for rigorous evaluation, such as requiring grantees to have a comparison group and interpreting effects only when grantees had sample sizes of 10 or greater and when students in the course sections being compared were similar on baseline tests.

The ALMAP study answered three core research questions:

1. In what institutional and disciplinary contexts are adaptive learning technologies showing positive impacts on student outcomes?

2. How is the use of adaptive courseware affecting institutional costs?

3. How are instructors and students perceiving and using adaptive learning technologies?

We acknowledge the ALMAP grantee institutions for their pioneering efforts in bringing adaptive technology to higher education and for collaborating with us to collect rigorous evidence that can inform the field. This report provides a glimpse into the state of the art of adaptive learning courseware across the range of U.S. institutions of higher education—from research universities to colleges focused on undergraduate education, and from public community colleges to private online colleges. Many lessons were learned from this research that will be useful in future work.

---

[9] *Bureau of Labor Statistics. (2010, September). Back to college. Spotlight on statistics. Washington, DC: Author. http://www.bls.gov/spotlight/2010/college/*

Portfolio Description

The ALMAP grantees experimenting with adaptivity were 10 bachelor's degree programs and 4 associate's programs. The grantees added adaptivity in one of three scenarios: (1) as part of a shift from traditional to blended instruction; (2) as an enhancement to existing online courses; or (3) as a refinement to an existing learning technology used in blended instruction. SRI organized product evaluation findings by these three use cases:

- Blended Adaptive vs. Lecture—Four institutions (three bachelor's programs and one associate's program) added adaptive courseware as a study tool to support seven face-to-face lecture classes.
- Online Adaptive vs. Online—Seven institutions (five bachelor's programs and two associate's programs) added adaptive courseware to 11 preexisting fully online courses.
- Blended Adaptive vs. Blended—Three institutions (two bachelor's programs and one associate's program) swapped adaptive courseware into four face-to-face courses already using blended technology to support learning.

The ALMAP grantees used adaptive courseware in five different subject disciplines across 23 different course implementations. They implemented the adaptive learning courseware predominantly in foundational or remedial courses in mathematics and English, but other subjects were social sciences (psychology, economics), business, and a lab science (biology). The content areas for the ALMAP courseware implementations are shown in Table 1.

The ALMAP grantees were implementing other changes in instructional delivery at the same time that they added the adaptive courseware. For example, two ALMAP grantees (both associate's degree institutions) added adaptivity as part of a switch from a lecture to an emporium class model. The emporium delivery approach features a lab in which the instructor provides guidance as needed to students learning individually using courseware. Two other ALMAP grantees (both research universities) explored adding adaptivity to gateway major courses at the same time that they were reforming their methods of classroom instruction to focus more on interactive discussion and less on lecture. One ALMAP grantee (an associate's degree institution) studied how adding adaptivity might enhance an instructional delivery shift from a fully online class to a blend of online and occasional face-to-face meetings.

Table 1. Disciplinary Content Areas for ALMAP Courseware Implementations

Discipline	Mathematics*		English Language Arts*		Business		Social Science		Lab Science	Total
Subdiscipline	Basic Math	Algebra	English	Reading & Writing	Business	Marketing	Economics	Psychology	Biology	9
Number courseware implementations	5	1	4	3	1	1	1	3	4	23

* Includes both college-level and remedial courses

Overview of Grantees' Reasons for Using Adaptive Products

Most ALMAP institutions chose to integrate adaptive courseware into gateway and remedial courses with high attrition and failure rates. As noted above, they experimented with adaptivity in a wide range of common college disciplines and subdisciplines. The examples below illustrate the kinds of situations motivating the ALMAP grantees to try out adaptive courseware. (A full list of grantees and their ALMAP adaptive courses appears in Table A1 of Appendix A.)

Remedial Mathematics: In Florida, the state passed legislation curtailing colleges' ability to require students to take remedial courses, so St. Petersburg College redesigned its introductory college mathematics and English classes around an adaptive learning product to better support the diverse learning levels of entering students. Essex County College in urban Newark, New Jersey, faced such high failure rates in its remedial math classes that it created a new required course focused on improving students' independent study skills and redesigned its main mathematics course to be a coached emporium model built around an adaptive learning product.

Remedial English: In Arizona, Rio Salado Community College added an adaptive element to its online English (reading and writing) class because the product provided automated dashboards on students' reading and homework progress, thereby reducing the burden on faculty of tracking the performance of online students at both the class and individual levels. As part of its response to legislation, St. Petersburg College also redesigned its introductory English community college classes around a separate adaptive learning product.

Another approach to supporting adult learners who have been out of school for some time is to embed remedial English support in college-level courses not only in English but in other subjects as well. In online college-level English and economics courses, Shoreline Community College in Washington State used an adaptive product to support this approach. The product offered course activities and student progress dashboards that helped an instructor team composed of one subject matter specialist and one remedial specialist monitor student performance on the distinct course aspects of college-level content and the remedial content. Similarly, faculty members at the University of North Georgia, Gainesville, were already using some online content to support students in English as a second language and remedial courses, and they chose to try out an adaptive product to provide more real-time feedback to instructors so they could tailor instruction to their students' needs.

Introductory Biology and Introductory College Mathematics: Science educators at the University of California, Davis, added adaptive courseware to an introductory biology course with the goals of helping students improve their independent study skills and supporting instructors' use of "interactive" teaching methods. The hope was that a combined interactive and adaptive approach could better highlight what's interesting and fun about studying biology and thus stem the flow of students dropping out of the major after taking a course long known by undergraduates as a "weeder," a term college students use to describe a lower-division course intentionally designed to screen out low-performing students to keep them from advancing into the upper-division courses of the major. A more "back to basics" approach was taken by three colleges that teamed to tackle introductory biology and mathematics course attrition—online Excelsior College, Metropolitan State University in Colorado, and the University of Missouri, St. Louis. These colleges joined forces to test an adaptive product designed to help students memorize the copious terminology and concepts of introductory biology and to become fluent in essential mathematics procedures through regular applied practice.

Business/Marketing/Psychology/College Algebra: ALMAP grantees St. Leo University, Capella University, and SUNY Empire State wanted to address student attrition challenges in their online courses specifically designed for the adult learning market. They turned to adaptive products that offer authoring tools that enable instructors to create more interactive online activities and that give adult learners immediate feedback on their independent study performance. Traditional 4-year schools, North Carolina State University and Arizona State University had been experimenting with offering some of their general education courses online to provide undergraduates with some scheduling flexibility. Both decided to add adaptivity in these courses to support better progress monitoring by both instructors and students.

Product Descriptions

ALMAP higher education institutions tested nine adaptive learning products. These products use learning model algorithms to track learner progress and recommend next steps in a learning path. Some of the products also offer tools for content authoring and curation. The courseware's algorithms use data from quizzes and step-by-step tasks to infer where students are in the learning progression. Algorithms can focus on either the macro or micro levels of student learning.[10] [11] A macro-level student learning model tracks student progress through an entire course. A micro-level algorithm focuses on the domain knowledge required to complete a collection of tasks (lesson/unit) or a single task (learning object). Both types of algorithms characterize student performance according to an underlying idealized domain learning sequence and then recommend next tasks intended to close knowledge gaps or build fluency. A micro-level learning model may focus closely on estimating the accuracy of each step a student takes on a multi-step task, and then provide real-time feedback, encouragement, and hints to keep students on track.

While detailed analysis of the learning model and algorithms for each ALMAP product was beyond the scope of this evaluation, we can provide a sense of each product's adaptive approach by examining (1) the points at which the product uses data to modify the student's learning path (unit of adaptivity), and (2) the frequency of assessment within the product, which can range from an assessment

[10] *Shute, V. J. (1993). A macroadaptive approach to tutoring. Journal of Artificial Intelligence in Education, 4(1), 61-93.*

[11] *Van Lehn, K. (2006). The behavior of tutoring systems. International Journal Artificial Intelligence, 16(3), 227–265. Retrieved from http://dl.acm.org/citation.cfm?id=1435351.1435353*

at the close of each set of units (benchmark) to assessments that alert students whenever they make errors (formative) to assessments that update estimates of student competency after every system interaction (continuous). This approach was used previously by Educational Growth Advisors (2013) to categorize most of the ALMAP products.[12] For those ALMAP products that were not categorized by Education Growth Advisors, SRI researchers did so on the basis of interviews with grantees and courseware vendors and product descriptions. Table 2 characterizes the units of adaptivity and the frequency of assessment for the products used by ALMAP grantees. Nine grantees used products that focused on smaller, micro units of adaptivity and more frequent assessment; four grantees used products that focused on larger, macro units of adaptivity and moderately frequent assessment; and, one grantee used two products, one with micro and one with macro adaptivity.

Products also vary in other ways, including who authors the learning content—the instructor and/or the product vendor. This feature is also shown in Table 2. With respect to authoring, seven grantees used products that promoted content authoring by instructors using vendor tools or algorithms, four used products based on vendor-generated content, and three used products that permitted instructors to curate vendor content. It is important to note that the instructor authoring using vendor tools provides instructors with more influence on instructional design than instructor authoring modalities built on vendor algorithms.

[12] *Education Growth Advisors. (2013). Learning to adapt: Understanding the adaptive learning supplier landscape. Boston, MA: Tyton Partners. Retrieved from http://tytonpartners.com/library/understanding-the-adaptive-learning-supplier-landscape/*

Table 2. ALMAP Products Sorted by Unit of Adaptivity, Assessment Frequency, and Content Authoring Sources

Product	Unit of Adaptivity	Assessment Frequency	Content Authoring
ALEKS	Lesson/unit	Benchmark	Vendor
Open Learning Initiative	Lesson/unit	Formative Advisors	Vendor with some instructor curation
Pearson MyFoundationsLab with Knewton	Lesson/unit	Formative	Vendor with some instructor curation
Pearson MyMathLab with Knewton	Lesson/unit	Formative	Vendor with some instructor curation
Adapt Courseware	Learning object	Formative	Vendor
LearnSmart/Connect	Learning object	Formative	Vendor
Cerego	Learning object	Formative	Instructors using vendor algorithm
Smart Sparrow	Learning object	Formative	Instructors using vendor tools
CogBooks	Learning object	Continuous	Instructors using vendor tools

Source: Market analysis by Education Growth Advisors (2013), interviews with grantees, and vendor marketing documents.

Study Methodology

SRI researchers applied widely accepted evidence standards to consolidate evaluation results reported by the 14 ALMAP institutions. Grantees submitted data over three academic terms in a rolling manner according to their term schedules from August 2013 through August 2015.[13] All grantees were required to implement a quasi-experimental evaluation design comparing student outcomes and cost impacts between course sections using adaptive learning courseware (treatment) and sections not using it (comparison).

SRI analysts reviewed the quality of each grantee's original evaluation design and requested additional design improvements as needed, such as administration of pretests, using test instruments that were not overaligned with the treatment, blinding test scorers to condition, and requiring a minimum of 10 students per condition. The grantees varied in their ability to meet these standards. For the portfolio overall, the basic goal was to permit estimation of baseline equivalence between students in treatment and comparison course sections through the use of pretests. All but two grantees met this standard for at least some of their adaptive courseware implementations (for qualifying student samples see Table A2 in Appendix A). Of the 23 different college courses implementing ALMAP products for up to three terms, 15 quasi-experiments met our criteria for generating impact estimates for course completion and 16 quasi-experiments met our criteria for generating impact estimates for grade outcomes.[14] Only seven quasi-experiments using direct assessments of student learning (i.e., a posttest taken by both the adaptive courseware and the comparison students) met our study inclusion criteria. Stronger designs compared data between treatment and comparison course sections offered concurrently to control for variations in the types of students who enroll at different times of the school year. Ten of the 14 institutions gathered concurrent comparison group data during at least two of their terms.

Student Outcome Data Collection and Analysis

The number of students in ALMAP courses was relatively consistent across terms, as shown in Table 3. The sample demographics varied by term. We observed the following ranges per term among the students in courses that met the inclusion criteria: 10-16% African American, 13-19% Asian, 15-20% Hispanic, 43-50% White, and the rest Other. Pearson's chi-square tests revealed that the racial compositions of the treatment and comparison groups were significantly different ($p < .001$). Across all three terms, a higher proportion of African Americans was assigned to the nonadaptive condition, and a higher proportion of Asians was assigned to the adaptive condition. Although we controlled statistically for differences in pretest scores, this significant difference in the demographics of students in the course sections being compared means that differences in outcomes might be attributable to characteristics of the students rather than to the adaptive nature of the courseware. Twelve of the 14 institutions reported the Pell Grant status of their students. The percentage of Pell Grant students in the data set ranged between 39% and 41% across terms. Underrepresented minorities accounted for between 56% and 63% of Pell Grant students. (For full results on student participation and instructor/course participation, see Tables A2 and A3 in Appendix A.)

---

[13] Data were collected from grantees on the following schedule: first term, from August 2013 through June 2014; second term, from January 2014 through February 2015; and third term, from May 2014 through August 2015.

[14] Some courseware and case combinations were implemented for two or three terms; for these an impact was estimated for each term where the data met our analytic criteria and the weighted average was used as an overall impact estimate for the case.

Table 3. Student and Faculty Participation in ALMAP Studies, by Academic Term

Term	Total Participants		Total Adaptive Participants	
	Students[a]	Faculty	Students	Faculty
1	7,688	238	3,063	100
2	7,458	260	3,150	107
3	8,459	249	3,733	93
Total	23,605[b]	747[c]	9,946[d]	298[b]

[a] Approximately 12.4% (2,933) of students indicated they were repeating the course. However, only 118 of those indicating they were repeaters appeared in more than one term of the ALMAP study data. For the rest of the repeating students, there was no way to distinguish between those who had taken non-ALMAP versions of the course previously and those who were assigned new student IDs when they took an ALMAP course for the second time.

[b] Five institutions used a retrospective comparison approach (e.g., repeating the use of the same comparison group for the adaptive condition in multiple terms). We calculated the number of unique students as 19,697.

[c] Many instructors taught more than one section or in more than one term. The total number of unique instructors is 281 across the entire sample. The total number of unique instructors in the adaptive courseware conditions is 170, of whom 120 also taught control classes.

[d] There were 34 students who appeared in the adaptive condition dataset for more than one term (course repeaters).

Grantees submitted student-level data to SRI for each of three academic terms. The data included student income level (Pell Grant recipient), enrollment status (course repeater, part time), and outcomes (course grade, course completion, persistent college enrollment in the next term, and, when available, final posttest grade). Reporting of direct measures of learning (such as pretests and posttests) was relatively infrequent; most grantees reported course grades, which are obtained more easily but reflect not only learning, but also factors such as class participation, attendance, and homework assignment compliance. In some cases, we observed that the students in the adaptive learning conditions differed considerably (> 0.25 standard deviation) from comparison group students on the measure of prior learning. In those cases, we could not calculate unbiased estimates of adaptive learning.[15] This stipulation removed roughly 70% of the submitted data from the final student outcomes analysis. In addition, five grantees did not provide any data on a prior learning measure and their data were removed from the analysis as well since we had no way of determining the equivalence of the students in the two conditions of the quasi-experiment. Of the remaining nine grantees, all but one had some data removed nearly every term because pretests showed a lack of equivalence between treatment and comparison course sections for one or more of the student subgroups. Only one grantee showed baseline equivalence for every term's data.

[15] What Works Clearinghouse. (2014). Procedures and standards handbook (Version 3). Washington, DC: U.S. Department of Education.

With the remaining usable data, SRI estimated the impact of adaptive courseware by computing an effect size for each adaptive treatment versus comparison course section contrast for course grade and learning assessment.[16] We then used all of the effect estimates in a meta-analysis computing the average effect size on each of these student outcomes (course grade and test scores). For the binary course completion outcome (credit, no credit), we computed an odds ratio, which is the ratio of the odds of a student in the treatment condition completing the course successfully divided by the odds of a student in the control condition doing so. An odds ratio that is not statistically different from 1 indicates that the adaptive courseware had no impact on the odds of successful course completion; an odds ratio of less than 1 indicates that the control group had higher odds of course completion, and an odds ratio of greater than 1 indicates an advantage for the treatment group. These odds ratios were subsequently converted into the effect size metric to allow for easy comparisons of impacts across different student outcomes. (See Appendix B for the odds ratio and corresponding effect size calculation.)

In instances for which we had a sufficient number of cases, we also ran separate analyses for Pell Grant recipients. When considering the following analysis and results, it is important to recall that as the number of analyses conducted on the same sample set increases, so does the likelihood of finding a statistically significant outcome by chance.

In all meta-analyses weighted averages were used. For meta-analyses of effect sizes over terms for individual courses, effect sizes were weighted by the inverse variance (which takes into account both sample size and variability), using a fixed effects model (which assumes the distribution of estimated effect sizes share a common true effect size, and that any deviation away from that true effect size is due simply to sampling error in a given term). When averaging the effect size estimates for use case comparisons (the three scenarios), we weighted the effect size estimates by the inverse of the sum of the within-study and between-study variance, using a random-effects model (which assumes the deviation of effect size estimates away from a common mean is not simply due to sampling error, but also incorporates a random effect). It is important to keep in mind that both fixed-effects and random-effects models make certain assumptions about the data. We discuss these assumptions and different weighting strategies in Appendix B.

After completing the analyses for the entire sample and the three use cases, we also tested several courseware features (unit of adaptivity, content author, and subject domain) to see whether these aspects of the courseware influenced the magnitude of its effect on student outcomes. For all moderator variable tests, we used a Cochran's Q test to test the null hypothesis that the heterogeneity among the mean effect sizes of the various sub-groups of the moderators was no more than would be expected by chance. The input into a moderator analysis is the averaged effect-size estimate and corresponding variance computed from the meta-analysis.

Fifteen of the 23 courseware implementations in the data set met our criteria for inclusion in the course grade meta-analyses, and 16 had data meeting our criteria for inclusion in the course completion meta-analyses. For test scores, only seven of the implementations met the inclusion criteria.

---

[16] An effect size is a standardized measure of impact computed by subtracting the comparison group mean from the treatment group mean and dividing by the standard deviation. When the two conditions have equal outcomes, the effect size is zero; when the treatment group has a better outcome the effect size is positive; when the comparison group has a better outcome the effect size is negative.

Cost Impact Data Collection and Analysis

The ALMAP study also included examination of the costs of using adaptive learning products in relation to learning impacts. This approach represented a departure from past studies, which examined cost factors apart from learning outcomes.[17] To distinguish the ways that adaptive courseware influences cost drivers in higher education, we used an "ingredients approach"[18] to deconstruct the cost-related activities associated with implementing the adaptive learning courseware. Cost data were collected from grantees through interviews and using an Excel spreadsheet to organize costs by a standard set of categories (see Table 4).

The ALMAP study grantees gathered data on two types of costs—up-front costs that include any needed infrastructure upgrades, instructor training, and instructor development of content (Term 1) and ongoing costs (Terms 2 and 3). We compared treatment and comparison conditions in terms of

- Types of faculty assigned (full time, adjunct, teaching assistants)
- Cost-per-student.

Table 4. Cost Ingredients Collected from ALMAP Grantees

Cost Drivers	Relevant Quantities	Quantities to Compute per Student Rates
Training and development costs (up front and ongoing) School Leaders	• Course content customization • Instructor training • Infrastructure upgrades	• Number of instructors • Hourly salary rates per instructor by type (full time, adjunct, teaching assistant) • Number of training and development hours • Number of instructional delivery labor hours and costs • Student course enrollment
Delivery costs	• Delivery of instruction • Textbook costs • Online access fees • Annual technology support costs	

[17] For example, Maher, M. W., Sommer, B., Acredolo, C., & Matthews, H. R. (2002). What are the relevant costs of online education. Davis, CA: University of California, Davis; and Tucker, J., & Neely, P. (2010). Unbundling faculty roles in online distance education programs. The International Review of Research in Open and Distributed Learning, 11(2), 20-32.

[18] Levin, H. M., & McEwan, P. J. (2003). Cost-effectiveness analysis as an evaluation tool. In J. B. Cousins, T. Kellaghan, & D. L. Stufflebeam (Eds.), International handbook of educational evaluation (pp. 125-152). Netherlands: Springer.

Although cost per student (costs for development, training, and instructional delivery divided by number of students) is a common metric used in higher education, it posed challenges in the ALMAP analysis because institutions sometimes had drastically different numbers of students in treatment and comparison course sections. In some cases, grantees preferred to start with a small number of students in the adaptive courseware sections; in others, grantees had decided by Term 3 to significantly reduce the number of students in the comparison course sections. These local implementation decisions led to different denominators that distorted the per-student cost comparisons. Cases where the difference in sample size exceeded 20% required additional modeling to interpret, which was beyond the scope of the current study. For this reason, such cases were removed from the ALMAP costs data set. (For details on the cost analysis, see Appendix B.)

Instructor and Student Experience and Satisfaction Data Collection and Analysis

To understand the experience and satisfaction of instructors and students using adaptive courseware, the ALMAP evaluation gathered survey data each term from both instructors and students. Additionally, analysts incorporated data gathered from the cost data collection process, in which instructional teams in both treatment and comparison conditions estimated the number of hours they devoted per week to three types of instructional practices (lecture/presentation, grading/progress monitoring, and other interactions in and out of class time).

Instructional practice substudy. Because a key goal of using adaptive technology, particularly for grantees switching to blended courseware delivery use (Blended Adaptive vs. Lecture), was to reduce time devoted to lecture-based instruction, one analysis focused on estimating and representing the differences in estimated duration of lecture activity between treatment and comparison conditions. Teams of faculty leaders and instructors who were involved in teaching adaptive and nonadaptive course sections were asked to estimate the average number of hours per week that full-time instructors, part-time instructors, and graduate student assistants spent engaging in five aspects of instructional activity (diagnosis of skill/knowledge, presentation, other interaction such as nonlecture classroom discussion, office hours, email, and labs, lecture preparation and presentation, and monitoring progress of students) and two aspects of student evaluation activity (test proctoring and grading of tests, assignments, and homework). Analysts combined these estimates into three larger categories of instructional practice: lecture/presentation, progress monitoring/grading, and other interactions for treatment and comparison conditions. Then they computed the difference in weekly hours of lecture in adaptive and nonadaptive course sections for each courseware implementation.

Instructor surveys. Instructor surveys administered by SRI included 23 questions on instructors' satisfaction and their perceptions of student satisfaction and learning with the courseware, the adaptive courseware features they used, and challenges and concerns with the courseware. The response rates varied by institution and term. Additionally, one grantee (St. Petersburg) administered its own survey to instructors using the adaptive products; SRI integrated these responses by focusing on questions that were parallel to those on its own survey. (For details, see Table A4 in Appendix A.) In some cases ALMAP grantees did not provide email addresses for all participating instructors in Terms 1 and 2, and so the number of invited instructors in the adaptive condition was considerably smaller in those terms than in the final term of the study (See Table 5).

Table 5. Instructor Survey Participation by Academic Term

Term	Total Participating Faculty in Adaptive Condition		
	Participating	Invited to Survey[a]	No. Responding (Rate, %)
1	100	67	49 (73%)[b]
2	107	93	52 (56%)
3	93	81	40 (49%)
Total	300	241	141 (59%)

[a] Includes St. Petersburg, which conducted a separate instructor survey.

[b] Overall response rate; individual item response rates varied.

Analysts reviewed descriptive statistics for close-ended survey items and conducted qualitative thematic analyses of responses to short-answer questions. Instructor survey responses were aggregated across the three terms and disaggregated by the three use case conditions. Many faculty members took the instructor survey in more than one term; 12 faculty members participated in the survey in all three terms.

Student surveys. The SRI evaluation team asked each grantee to integrate nine questions into surveys that they administered to students in their adaptive learning courses each term. These questions were intended to gather data on student usage and perceptions of adaptive courseware, including helpfulness, satisfaction, interest, enjoyment, engagement, and learning. Some grantees made changes to the survey item wording and item scales, complicating the aggregation of survey responses across grantees. For the aggregate analysis, SRI focused on the proportions of students responding positively to three questions worded similarly across surveys: positive engagement, improved confidence, and perceived progress in learning. (For details, see Table A5 in Appendix A.)

Findings

Findings are presented here for the ALMAP portfolio as a whole and for each of the three use cases. Student outcome findings are presented for all students unless otherwise specified. This section begins with both the overall results from a meta-analysis of impacts of each adaptive course on three student outcomes (course grade, posttest, and course completion) and the results evaluated by each of the three use cases (Blended vs. Lecture, Online Adaptive vs. Online, and Blended Adaptive vs. Blended).[19] Then it describes tests of three moderator variables on course grade outcomes only. The moderator variables were the courseware products' unit of adaptivity type and content authoring modality (the courseware features described in Table 2), and subject domain.

We also present qualitative overviews of the three use cases. They describe the implementation approaches and courseware features, the impacts on learning and completion for all students and for Pell Grant students by grantee, the implementation experiences and perspectives of instructors and students, and the highlights of trends from the cost analysis. The collective findings of the grantees in each use case are summarized in tables for an at-a-glance summary of the direction of findings for student outcomes, costs, instructor satisfaction, and student satisfaction. These tables summarize multiple distinct data sources as follows: significantly positive or a majority of positive outcomes or opinions (denoted by +), significantly negative or a majority of negative outcomes or opinions (denoted by a -), no change or no majority opinion (denoted by ~), or no available data (NA).

Meta-analyses and Moderator Variable Analyses

Across all 15 impact estimates for course grades, there was a very small overall detected effect of 0.079 that nevertheless was statistically significant ($p < .05$). Of the individual courseware impact estimates, 10 of 15 were essentially zero, indicating that **grades in the adaptive and comparison versions of the course usually were equivalent.** Of the remaining five cases. four were significantly positive, and one was significantly negative.

Overall impact estimates for course completion were insignificant. The overall average effect size for course completion was 0.019 ($p = .694$), and the odds ratio was 1.02, indicating that **overall the odds of completing an ALMAP course for a student in an adaptive section were essentially the same as those of completing the course if participating in a nonadaptive section.** Just 2 of 16 individual contrasts were significantly positive and none was significantly negative.

The average effect size for posttest scores among the seven comparisons providing usable data on this outcome was modest but positive (effect size = +0.184), and statistically significant ($p < .05$). Three of the seven quasi-experiments using learning assessments (posttests) had significant positive mean effect sizes, ranging from +.25 to +.77.

---

[19] *For continuous variables such as course grade or assessment score, the effect is estimated as the adaptive minus the comparison group mean divided by the pooled standard deviation. For the binary variable of course completion, the effect estimate is computed based on the rate of incidence for the two groups being compared and their sample sizes.*

Next we ran moderator variable analyses for the course grade outcome, testing three features that vary across different adaptive courseware implementations (adaptivity type, courseware content author, subject domain) to see whether the use case influenced the impacts of adaptive courseware on the three student outcomes.[20] We first ran a Cochran's Q test [21] test to examine the consistency of results among the ALMAP studies in each subgroup defined by these features; in those cases where the subgroups varied significantly in terms of the impact estimates for adaptivity, we examined the subgroup significance patterns.

Adaptivity type. We found significant differences in estimated impact of adaptivity on course grade depending on whether adaptivity occurred at the level of the lesson/unit ("macro adaptivity") or at the level of the learning object (micro adaptivity), Q = 5.62, $p < .05$. Based on a random-effects model, the 10 comparisons involving courseware using micro adaptivity had a small but significant positive effect size on course grade on average (effect size = +0.15, $p < .01$). The 18 comparisons involving macro adaptivity did not significantly improve course grades (effect size = -0.018, $p = .709$).

Courseware content author. The courseware's impact on course grades tended to vary, though not quite significantly, depending on whether it involved content coming from the vendor or from the instructor (Q = 6.53, $p = .089$). Only the subgroup of comparisons involving instructors supplying some content for use with a vendor-provided adaptive algorithm had a significant positive impact on course grades based on a random-effects model (effect size = +0.23, $p =< .01$). Average impacts were not significantly different from zero for adaptive courseware using vendor-provided content or extensive amounts of content input by the instructor.

Subject domain. When the courseware implementations were grouped by subject domain (math, English language arts, biology, social science), the estimated magnitude of the impact of adaptivity did not vary significantly across subjects (Q = 3.02, $p = .388$). With a larger number of cases, the trend toward more positive impacts in mathematics and biology courses than in other subjects might have attained statistical significance.

Because the nature of the comparison course sections differed for the three use cases, we next examined impact estimates for the three subsets of quasi-experiments. These findings are presented below along with information about the implementation approaches, courseware features, and costs for each use case.

Blended Adaptive vs. Lecture

Case overview. In this use case, grantees were not only implementing adaptive approaches but also shifting from face-to-face to blended instruction. One associate's degree-granting institution, St. Petersburg Community College, sought to help remedial students by shifting from lecture classes to blended adaptive courseware in an emporium lab, and three bachelor's degree-granting institutions—Metropolitan State University, University of Missouri-St. Louis (UMSL), and University of California, Davis (UC Davis)—integrated adaptive courseware as study aids in traditional lecture courses in gateway major and general education courses.

---

[20] *We chose course grade as the outcome for these analyses because it had more cases with usable data than the posttest outcome and had greater variability than the related outcome of course completion.*

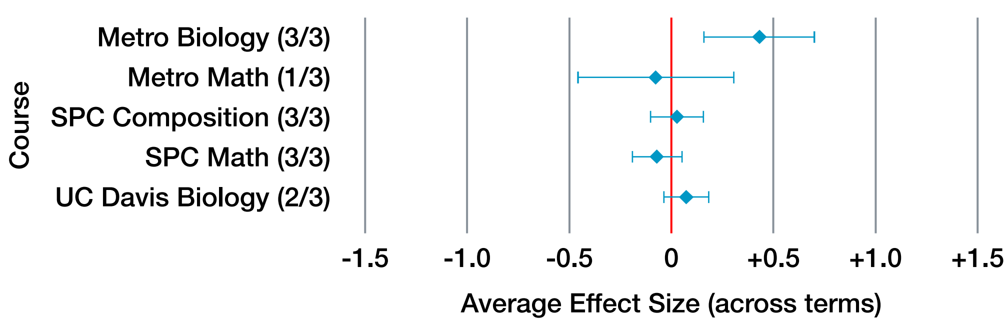
[21] *A Cochran's Q test of between-group differences was calculated using the inverse variance and a fixed-effects model. This is because the purpose of a moderator analysis is to test for systematic differences. A random-effects model inherently assumes any differences are non-systematic, hence the name random-effects.*

Implementation approach and courseware features. At St. Petersburg, students taking intermediate algebra used the adaptive learning products in large computer centers called emporium classrooms with instructors available to coach students. Students used ALEKS courseware, a product that covers course content and provides adaptive feedback as students complete vendor-created units. ALEKS starts learners at different points based on their performance on a pretest, delivers content through online presentations, offers regular assessment of student learning coupled with feedback and progress dashboards, and makes referrals to additional remediation as needed. St. Petersburg also used LearnSmart for English Composition 1. LearnSmart reinforces commonly forgotten concepts by encouraging periodic review, sets up a personalized learning plan for students based on an initial pretest, and tracks student progress, sending alerts if a student falls behind the expected pace. It offers adaptive feedback as students complete each vendor-created learning object.

An Online Learning Initiative (OLI) implementation at UC Davis offered basic biology content in an online multimedia presentation coupled with quizzes to check knowledge, recommendations for remedial or enrichment content, and progress dashboards. The OLI Biology product covers course content and provides adaptive quiz feedback after each vendor-created unit. UC Davis not only implemented the adaptive courseware, but also engaged teaching assistants in teaching in a more interactive way. Metropolitan State and UMSL, working with gateway major and general education students in mathematics and biology, sought to improve basic study skills and used Cerego courseware that homed in on improving students' recall of new terminology and concepts. Cerego is a targeted study aid composed of learning objects customized by instructors.

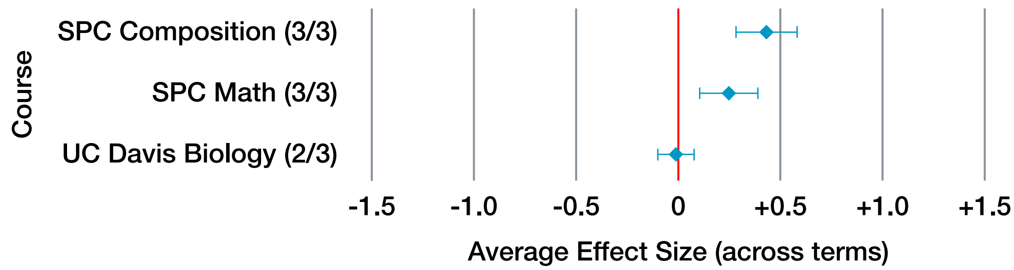
Grantee-by-grantee learning outcomes. Of the five cases with course grade outcomes, one (Metropolitan biology) experienced a small but significantly positive effect from moving to an adaptive blended model. For the three cases with posttest scores, two produced positive effects by moving to an adaptive blended model (St. Petersburg College Composition, St. Petersburg College Math). For the five cases under this scenario with course completion data, only students in the Metropolitan biology course had significantly better odds of completing the course if experiencing an adaptive blended model. (See Figures 1-3.)

Figure 1. Blended Adaptive vs. Lecture Course Grades



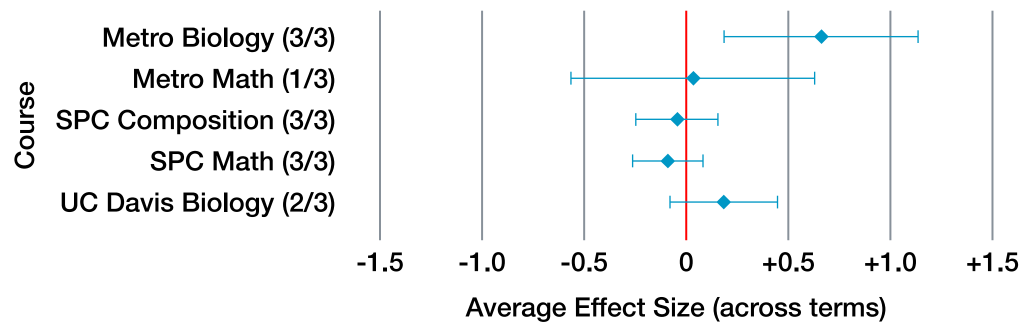
Note: Numbers in parentheses reflect number of terms for which course data met study criteria for inclusion in the analysis.

Figure 2. Blended Adaptive vs. Lecture Posttest Scores



Note: Numbers in parentheses reflect number of terms for which course data met study criteria for inclusion in the analysis.

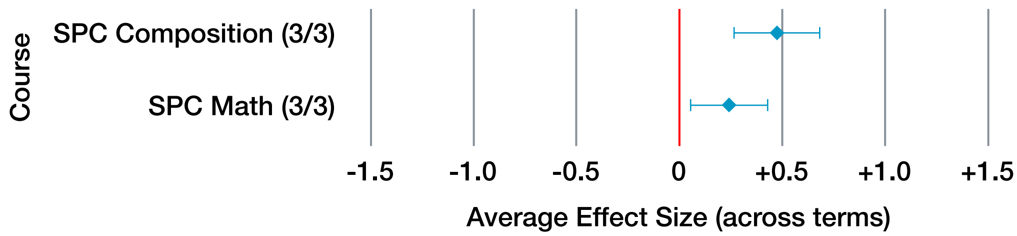
Figure 3. Blended Adaptive vs. Lecture Course Completion



Note: Numbers in parentheses reflect number of terms for which course data met study criteria for inclusion in the analysis.

There were only two instances (St. Petersburg College composition, St. Petersburg College math) in this use case group with posttest outcome data disaggregated by Pell Grant status. Importantly, although small, these adaptive courseware effects on posttest scores of Pell Grant students were significantly positive, as shown in Figure 4.

Figure 4. Blended Adaptive vs. Lecture Posttest Scores for Pell Grant Students



Note: Numbers in parentheses reflect number of terms for which course data met study criteria for inclusion in analysis.

Cost factors. The grantees provided cost data sets meeting our criteria for analysis for five of the Term 1 implementations (estimating start-up costs) and two of the Terms 2 and 3 implementations (ongoing costs). For Term 1 start-up costs, we found reductions in overall per-student costs in three cases and higher costs in two cases. The cases with relatively low start-up costs at St. Petersburg ($15 more per student than comparison in math; $8 more per student in English) and UC Davis ($6 more per student than comparison in biology) employed vendor-authored courseware. Faculty development time in these implementations was spent mostly on selecting topics from the courseware offerings. The start-up costs for courseware built around faculty content and delivered through a vendor memorization algorithm were somewhat higher for UMSL biology ($30 more per student than comparison) and math ($27 more per student). For Terms 2 and 3 ongoing costs, we found cost reductions in both of the cases that met our criteria for inclusion in the analysis. These cost reductions (14% at St. Petersburg math; 50% at UMSL math) were related to instructors devoting fewer hours to instruction rather than to changes in the types of faculty (e.g., tenure track vs. adjunct) assigned to the course sections being compared.

Instructor and student experiences. Instructional practice data indicated only a modest reduction in the amount of time devoted to lecture/presentation in the blended adaptive course sections (M = 6.9 hours per week) compared with face-to-face sections (M = 7.3 hours). These findings suggest that the transformation of instructional practices may not have been as strong as some grantees had hoped.

At St. Petersburg, 92% of mathematics instructors using ALEKS reported making significant changes to their usual instructional practice so as to use the vendor-created content and dashboards to guide instruction; 40% of instructors reported usability challenges.[22] In contrast, only 15% of St. Petersburg's English teachers who were using LearnSmart/Connect reported making changes to their usual instructional practices and 60% reported usability challenges connected to the courseware. Most instructors in both courses reported satisfaction that the courseware permitted students to learn effectively (see Table 6). St. Petersburg students in both courses gave both courseware products high ratings for being engaging, satisfying, and making information location easy, but relatively few reported that the products increased their confidence in their mastery of the subject matter.

[22] *More descriptive details were not provided through this college-designed and administered survey, which was different from the one completed by instructors at other ALMAP grantee institutions.*

At the bachelor's degree-granting institutions, reducing lecture time devoted to basics was a goal for most of the grantees, and there was some evidence this occurred in the adaptive courseware sections about half the time. Some instructors reported kicking off class meetings with discussions of hard-to-learn concepts or reducing emphasis on basic grammar or terminology definitions based on their reviews of reports from real-time courseware dashboards.

Instructors generally reported more satisfaction with the adaptive courseware than students did, and when instructors raised concerns, they were about accuracy or the depth of the content or about the alignment between the courseware quizzes and the course's final exam.

Students gave mixed reports of how engaging they found the adaptive courseware, but generally they reported positive impacts on learning. Students at Metropolitan reported using Cerego just "a few times or more," raising questions about whether the level of usage provided a fair test of the product's potential. Instructors occasionally reported challenges in getting students to use the courseware frequently enough to achieve benefits. The most consistently positive instructor responses in this group were about Cerego; faculty saw the courseware as addressing a particular problem for their students: memorizing basic content. No student data were provided by UC Davis.

Results for the contrasts between Blended Adaptive and Lecture course sections are summarized in Table 6.

Table 6. Summary of Results for Blended Adaptive vs. Lecture Course Sections

Grantee	ALMAP Product	Better student outcomes?	Lowered ongoing cost?	Instructor satisfaction?	Student satisfaction?
St. Petersburg	ALEKS	~	+	+	+
Metropolitan	Cerego	~	NA	+	+
St. Petersburg	LearnSmart	~	NA	+	+
UC Davis	OLI	~	NA	+	NA
UMSL	Cerego	NA	+	+	-

+ Majority significantly positive, - Majority significantly negative, ~ Mixed or nonsignificant effects
NA = No data meeting criteria for inclusion in analysis

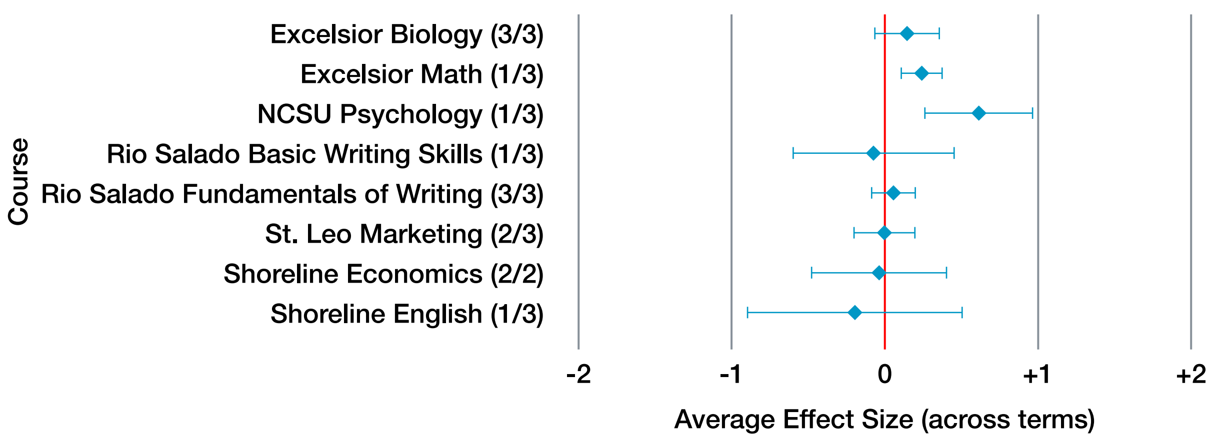
Online Adaptive vs. Online

Case overview. Two online colleges (Capella and Excelsior), three 4-year college online courses (SUNY Empire State, Saint Leo University, and North Carolina State University), and two 2-year college online programs (Rio Salado College and Shoreline Community College) compared online adaptive approaches with pre-existing nonadaptive online instruction. The content areas targeted in these implementations tended to be job relevant and geared toward working adults (e.g., business, economics, marketing, psychology, writing), but a few involved general education content in English, mathematics, and biology.

Implementation approach and courseware features. Some of these grantees used adaptive learning products to add livelier interactive content to their online courses, perhaps as a way to address weak student retention. Some sought to help online learners improve their study skills and reduce the burden on the largely adjunct instructor pool teaching these online courses. Five different types of adaptive courseware products were involved in the comparisons between adaptive online and other online courses, and they ranged from those requiring high instructor involvement in content customization (CogBooks, Smart Sparrow) to those obtaining content from faculty leaders (Cerego) to those that were used mainly to deliver basic content in vendor-created units with support for student progress tracking by instructors (MyFoundationsLab with Knewton and Adapt Courseware).

Grantee-by-grantee learning outcomes. Of the eight cases with course grade data meeting our inclusion criteria in this use case group, two (Excelsior math and North Carolina State University psychology) found a significant positive impact for adding adaptivity to an online course. (See Figure 5.)

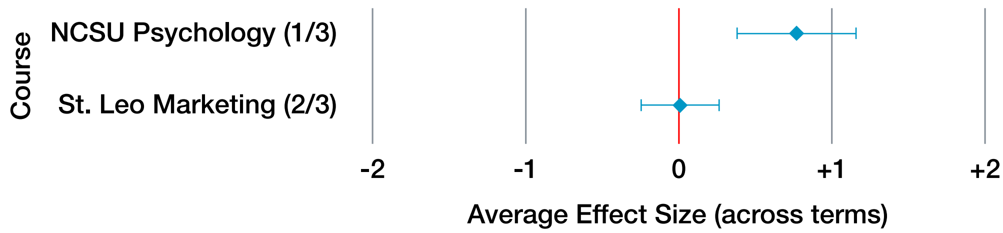
Figure 5. Online Adaptive vs. Online Course Grades



Note: Numbers in parentheses reflect number of terms for which course data met study criteria for inclusion in analysis.

Only two courses in this use case had estimates for the impact of adaptivity on posttest scores. The same North Carolina State University Psychology course that had a positive impact on grades had a positive impact on posttest scores as well. (See Figure 6.)

Figure 6. Online Adaptive vs. Online Posttest Scores

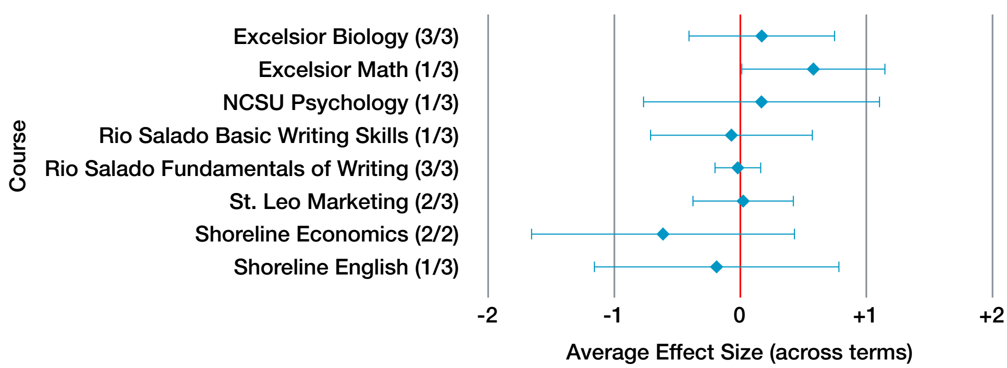


Note: Numbers in parentheses reflect number of terms for which course data met study criteria for inclusion in analysis.

Of the eight courses under this scenario with course completion data, only students in the Excelsior math course had significantly better odds of completing the course in the adaptive condition (see Figure 7).

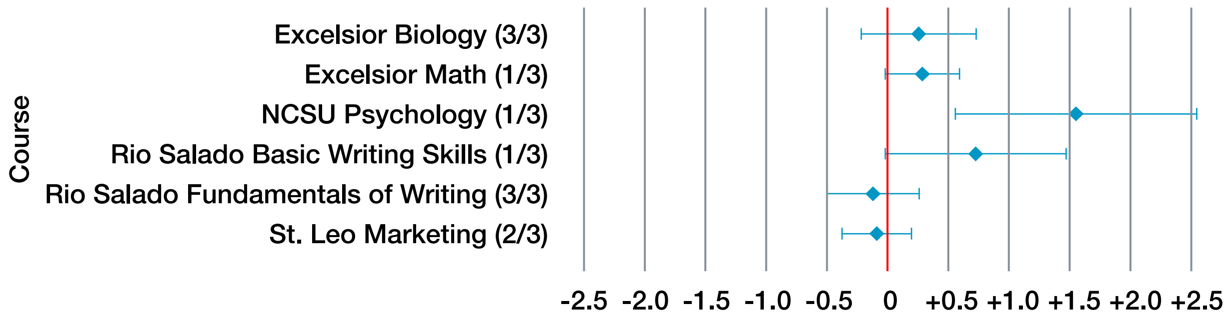
Fewer implementations met our criteria for inclusion in the meta-analysis for Pell Grant recipients. As shown in Figure 8, of the six implementations with course grade outcomes for Pell Grant students, the only one with a positive impact on grades was the sameNCSU Psychology course that had a positive impact on grades for students overall (Figure 8). Only a single implementation had a qualifying data set for posttest scores for Pell Grant students (St. Leo Marketing) and this implementation found no adaptivity effect for these students (nor had it found one for students overall, as shown in Figure 6).

Figure 7. Online Adaptive vs. Online Course Completion



Note: Numbers in parentheses reflect number of terms for which course data met study criteria for inclusion in analysis.

Figure 8. Online Adaptive vs. Online Course Grades for Pell Grant Students



Note: Numbers in parentheses reflect number of terms for which course data met study criteria for inclusion in analysis.

Cost factors. The grantees in this use case provided cost data meeting our criteria for analysis for just four Term 1 implementations (estimating start-up costs) and two Terms 2 and 3 implementations (ongoing costs). Term 1 start-up costs were higher for the adaptive course sections in all four cases, ranging from an additional $487 more per student for Shoreline's adaptive economics course to $1,319 more per student for Shoreline's adaptive English course. In all four cases the higher costs were associated with the engagement of instructors in creating online content using vendor tools. For Terms 2 and 3 ongoing costs, we found cost savings of 21% in the adaptive version of one of Rio Salado's English courses and a small cost increase of 6% in St. Leo University's marketing course. The cost shifts appeared to be related to associated changes in instructor labor.

Instructor and student experiences. The introduction of adaptive courseware into online courses was accompanied by a modest drop in the number of hours of lecture/presentation (M = 2.9 hours per week for the adaptive condition and M = 3.5 hours per week for the comparison condition).

Online instructors positively endorsed adaptive courseware features that encouraged students to pose targeted questions (CogBooks), but they were cautious about adaptive materials that seemed to increase burdens on adult learners. They mentioned concerns about adaptive courseware that threatened to "overwhelm" adult learners with content (Cerego), presented technical barriers to accessing multimedia lessons (CogBooks), or introduced learning pathways that failed to align with course or college schedules (Knewton). Instructors attributed course attrition and disaffection in the adaptive version of their online courses to these types of factors. Some said it was too soon to tell how the courseware was working (Smart Sparrow, Adapt Courseware). Pearson's MyFoundationsLab with Knewton at Rio Salado drew positive survey ratings from both instructors and students, but instructor responses to open-ended survey questions indicated various growing pains, such as ensuring smooth transfers of student grades from the courseware to the college system and making sure students were not simply rushing through the material. Results for the contrasts between Online Adaptive and Online course sections are summarized in Table 7.

Table 7. Summary of Results for Online Adaptive vs. Online Course Sections

Grantee	ALMAP Product	Better student outcomes?	Lowered ongoing cost?	Instructor satisfaction?	Student satisfaction?
Shoreline	CogBooks	~	NA	+	+
SUNY	CogBooks	NA	NA	+	+
Capella	CogBooks	NA	NA	+	+
St. Leo	Smart Sparrow	~	-	~	+
Excelsior	Cerego	~	NA	+	~
NC State	Adapt Courseware	+	NA	-	+
Rio Salado	Pearson w/ Knewton	~	+	+	+

+ Majority significantly positive, - Majority significantly negative, ~ Mixed or nonsignificant effects
NA = No data meeting criteria for inclusion in analysis

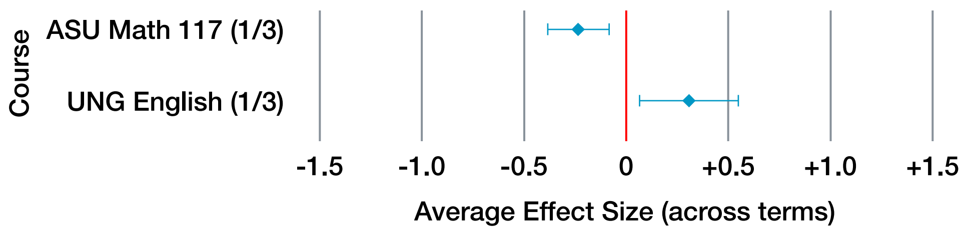
Blended Adaptive vs. Blended

Case overview. Essex County College used a blended adaptive model along with a shift to emporium instruction to help remedial mathematics students. Arizona State University (ASU) and the University of North Georgia-Gainesville (UNG) switched to adaptive courseware from other types of learning technology. ASU focused on mathematics students and UNG focused its ALMAP implementation on remedial English students.

Implementation approach and courseware features. Essex used ALEKS for mathematics in an emporium class and compared this model with using another blended learning product in conjunction with regular lecture classes. Both bachelor's-degree-granting institutions were seeking to reduce the amount of lecture and increase active student learning in the classroom, so they used the adaptive courseware to support self-study of basic material usually presented in class lectures. In both cases, the courseware—Pearson's MyMathLab with Knewton at ASU and LearnSmart at UNG for remedial English students—sets up a personalized learning plan based on an initial pretest and provides regular alerts to keep students on track. Both products provide students with adaptive feedback and supplemental instruction after they engage with vendor-created learning objects or units.

Grantee-by-grantee learning outcomes. Of the two cases in this use case group with course grade data meeting our inclusion criteria (ASU Math and UNG English), one (UNG English) achieved small but significantly positive results in the adaptive condition and one achieved small but significantly negative results (ASU math). (See Figure 9.)

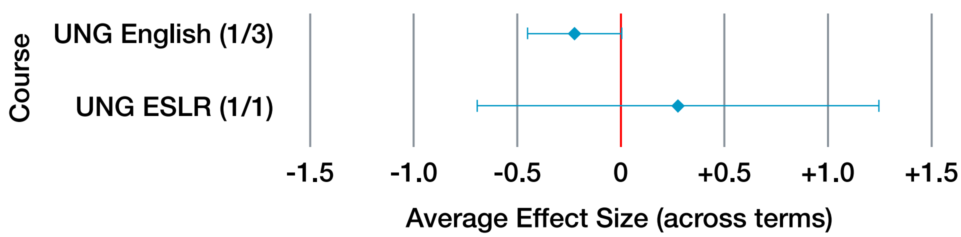
Figure 9. Blended Adaptive vs. Blended Course Grades



Note: Numbers in parentheses reflect number of terms for which course data met study criteria for inclusion in analysis.

Of the two cases in this use case group with posttest data meeting our inclusion criteria (UNG English and ESLR), neither saw any significant impact of adding adaptivity to a blended course delivery mode. (See Figure 10.)

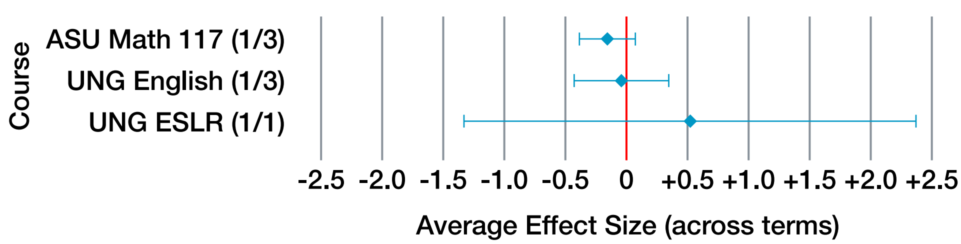
Figure 10. Blended Adaptive vs. Blended Posttest Scores



Note: Numbers in parentheses reflect number of terms for which course data met study criteria for inclusion in analysis.

Of the three courses with course completion data meeting our inclusion criteria, none showed students having better odds of completing a course when adaptivity was added to a blended course. (See Figure 11.)

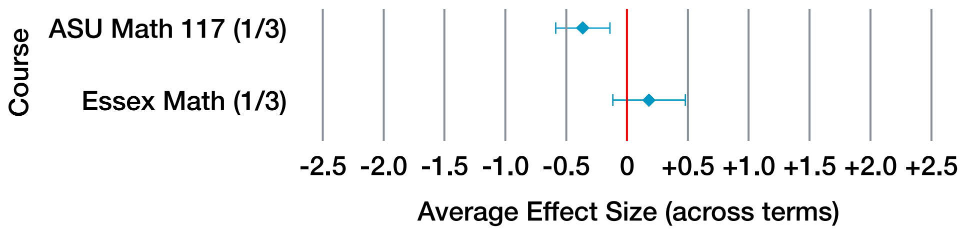
Figure 11. Blended Adaptive vs. Blended Course Completion



Note: Numbers in parentheses reflect number of terms for which course data met study criteria for inclusion in analysis.

There were just two side-by-side contrasts of impacts for Pell Grant students meeting our analysis criteria. As shown in Figure 12, the ASU Math 117 course section using a blended approach featuring adaptivity produced poorer outcomes for Pell Grant students than did the contrasting nonadaptive blended course; there was no effect of adding adaptivity to the blended Essex math course.

Figure 12. Blended Adaptive vs. Blended Course Grades for Pell Grant Students



Note: Numbers in parentheses reflect number of terms for which course data met study criteria for inclusion in analysis.

Cost factors. None of the grantees in this use case provided cost data sets meeting our criteria for analysis of either Term 1 or Terms 2 and 3 implementations.

Instructor and student experiences. As might be expected, introducing an adaptive product into a course that was already using blended learning did not have a material effect on the total number of hours spent in lecture/presentation each week (*M* = 3.5 hours for the adaptive blended condition compared and *M* = 3.6 hours for the blended condition). Nevertheless, other kinds of changes in instructional approach were triggered by the adaptive products.

At Essex, most mathematics instructors using ALEKS reported having to make significant changes to their usual instructional practice in order to use the courseware's vendor-created content and dashboards to guide instruction. Slightly less than half of Essex students reported enjoying ALEKS, but most reported that they learned more using it in an emporium class than they learned in a traditional lecture class. Instructors at ASU were split between those who were satisfied and those who were dissatisfied; 44% of instructors at UNG were satisfied. Short-answer responses from ASU instructors indicated that they felt the courseware's adaptive recommendations were changing too quickly to provide coherent guidance to students and that they found few benefits other than the dashboard tracking student progress. At UNG, short-answer responses indicated that students experienced technical problems signing on and had some difficulty finding the spaces where they were supposed to answer short-answer quiz questions; perhaps for this reason, not all the students completed the assignments. At both these universities, most students were dissatisfied with their adaptive courseware experience. Results for the Blended Adaptive and Blended course sections are summarized in Table 8.

Table 8. Summary of Results for Blended Adaptive vs. Blended Course Sections

Grantee	ALMAP Product	Better student outcomes?	Lowered ongoing cost?	Instructor satisfaction?	Student satisfaction?
ASU	Pearson w/ Knewton	~	NA	~	-
Essex	ALEKS	NA	NA	+	+
UNG	LearnSmart	~	NA	-	-

+ Majority significantly positive, - Majority significantly negative, ~ Mixed or nonsignificant effects
NA = No data meeting criteria for inclusion in analysis

Conclusions and Discussion

We summarize the findings for student learning, cost analysis, and satisfaction by instructors and students as follows.

- Overall, across the ALMAP trials, adding adaptivity to developmental and gateway courses had no effect on course completion rates after controlling for students' initial achievement levels under any of the three possible use cases. Only a couple of the 16 individual courseware implementations increased course success rates significantly; none had a significant negative impact.

- For course grades, we found a statistically significant but small positive impact on average for implementations adding adaptive courseware to a fully online course. However, the statistically significant average impact for adding adaptive courseware to an existing fully online course could be attributed to a large positive impact for one adaptive online psychology course at a single institution during one term. Adding adaptive courseware to an existing blended course, or switching from a traditional lecture course to a blended model using adaptive courseware, on the other hand, had no effect on grades.

- On average, **the adaptive courseware effect size for the seven comparisons providing usable posttest score data on this outcome was a small but statistically significant +0.13.** There were too few implementations with posttest data to support drawing conclusions based on separate effect size estimates for the three use cases.

- The ALMAP study provided a limited amount of data on the impact of adaptive courseware on outcomes for Pell Grant students, but the data that were available suggest that **by and large impacts were similar for Pell Grant students as for students overall.**

- Despite the unevenness in impacts on student outcomes across the studies, **most instructors, particularly those in 2-year colleges and teaching developmental education courses, reported satisfaction with their adaptive courseware experiences.**

- **Even so, fewer than half of the instructors planned to continue using the adaptive courseware.** At 4-year colleges, this may have been in response to the relatively low satisfaction levels among students.

- The ALMAP evaluation findings suggest that adaptive courseware be considered as an option for developmental and 2-year degree program courses, given **high levels of 2-year instructor and student satisfaction,** even though the positive impacts on learning and attainment outcomes were modest.
- The ALMAP data indicate **adaptive courseware is unlikely to reduce per-student course costs during the first term of implementation but may do so in subsequent terms.** Five of the eight comparisons of costs for the second and third implementations of adaptive courseware compared to earlier versions of the course found cost savings.
- **The ultimate goal—better student outcomes at lower cost—remains elusive.** There was only a single case in the ALMAP portfolio for which we could substantiate both cost savings and improved learning outcomes.

WE OFFER THE FOLLOWING RECOMMENDATIONS FOR FUTURE PORTFOLIO EVALUATIONS:

- **Baseline equivalence is essential for justifying claims about courseware effects,** but the common practice in higher education institutions is to simply compare course success rates without any data on student characteristics (baseline equivalence). ALMAP analyses found that student characteristics and prior learning often vary markedly from course section to section and across terms within the same institution.
- **Institutional analyses of adaptive courseware effectiveness should take into account the specifics of the way in which the courseware was used in a particular implementation.** The same piece of courseware can be used in many different ways, especially in blended learning implementations, and these differences can affect learning outcomes.
- **Postsecondary institutions can support student success by sharing anonymized data sets from side-by-side comparisons of adaptive courseware and other instructional approaches.** As more and more institutions turn to adaptive courseware, sharing data sets linking information about implementation, student administrative data, courseware system data, and student outcomes can build the empirical base needed to answer many of the questions left unanswered by the ALMAP data sets. Linked data sets need to be screened to insure they do not contain personally identifying information, but FERPA (Family Educational Rights and Privacy Act) compliant processes are readily available.
- **Researchers should help courseware users understand that learning efficacy is not a trait of a product per se or simply a matter of matching the right product to the right subject matter.** Rather, multiple factors affect learning outcomes and to make sense of student outcomes, analyses need to incorporate student characteristics, specifics of how the adaptive courseware is used, aspects of the course beyond the courseware product, and the way learning is measured to make sense of student outcomes.[23]

---

[23] *Means, B., Bakia, M., & Murphy, R. (2014). Learning online: What research tells us about whether, when and how. New York: Routledge.*

- **Course grades and course completion rates are less than ideal as outcome measures for efficacy studies.** These measures reflect more than what students have learned or can do; in many cases, they also reflect student attendance, timely completion of assignments, and class participation. More precise measures of learning, such as common tests, and other more detailed behaviors, such as student assignment completion, should be tracked as outcomes in evaluations of the impact of introducing new kinds of courseware. Especially when comparing outcomes for course sections taught by multiple instructors with different or unknown grading policies, using these measures as student outcomes can bias impact estimates. However, grades and course credits are critical for students and to inform institutional policies so they should be captured in evaluation studies along with more direct measures of student learning.

- **Satisfaction surveys, whether of instructors or students, are insufficient as the only outcome in a courseware evaluation.** Instructor and student perceptions as expressed on surveys can be useful information, but positive survey responses and impacts on learning do not always go hand-in-hand. Moreover, these surveys are dogged by low response rates and selection bias. Use of more frequent but very brief surveys embedded in the courseware could help raise response rates and lower concerns about sampling bias, but a certain amount of judicious observation and interviewing is recommended to provide insights to complement learning data and reports from survey samples.

- **More research is needed to develop cost effective ways to capture the changes in instructional practice associated with implementations of adaptive courseware.** The ALMAP evaluation focused on time devoted to lecture and presentations, but future work should examine (1) how adaptive courseware affects the relative balance between low-level and high-level content interactions between instructors and students and (2) how the automated dashboards in adaptive courseware affect instructors' sensitivity to individual and whole-class learning needs.

- **Funders should encourage modeling of cost effectiveness over the longer term, not just for an individual course.** It is necessary to take a 3-5 year perspective on student outcomes to capture the monetary savings associated with lower odds of needing to retake a developmental course or higher persistence rates. It is also useful to separate the cost ingredients of up-front development, preparation, and infrastructure for initial implementation from those of ongoing instructional delivery.

In conclusion, although ALMAP grantees varied in their capacity for institutional research, the ALMAP grantees made good-faith efforts to address the challenge of measuring learning impacts, costs, and student and instructor perceptions. We acknowledge them not only for their leadership in innovating with new courseware but also for their willingness to share their experiences and findings with the field at large. We view this sharing of results as a first step toward cultivating a critical community of inquiry around the usefulness of new courseware in supporting student learning and course completion in institutions of higher education.

Appendix A

Summary of Study Participation

- The ALMAP evaluation findings suggest that adaptive courseware be considered as an option for developmental and 2-year degree program courses, given **high levels of 2-year instructor and student satisfaction,** even though the positive impacts on learning and attainment outcomes were modest.

- The ALMAP data indicate **adaptive courseware is unlikely to reduce per-student course costs during the first term of implementation but may do so in subsequent terms.** Five of the eight comparisons of costs for the second and third implementations of adaptive courseware compared to earlier versions of the course found cost savings.

- **The ultimate goal—better student outcomes at lower cost—remains elusive.** There was only a single case in the ALMAP portfolio for which we could substantiate both cost savings and improved learning outcomes.

WE OFFER THE FOLLOWING RECOMMENDATIONS FOR FUTURE PORTFOLIO EVALUATIONS:

- **Baseline equivalence is essential for justifying claims about courseware effects,** but the common practice in higher education institutions is to simply compare course success rates without any data on student characteristics (baseline equivalence). ALMAP analyses found that student characteristics and prior learning often vary markedly from course section to section and across terms within the same institution.

- **Institutional analyses of adaptive courseware effectiveness should take into account the specifics of the way in which the courseware was used in a particular implementation.** The same piece of courseware can be used in many different ways, especially in blended learning implementations, and these differences can affect learning outcomes.

- **Postsecondary institutions can support student success by sharing anonymized data sets from side-by-size comparisons of adaptive courseware and other instructional approaches.** As more and more institutions turn to adaptive courseware, sharing data sets linking information about implementation, student administrative data, courseware system data, and student outcomes can build the empirical base needed to answer many of the questions left unanswered by the ALMAP data sets. Linked data sets need to be screened to insure they do not contain personally identifying information, but FERPA (Family Educational Rights and Privacy Act) compliant processes are readily available.

- **Researchers should help courseware users understand that learning efficacy is not a trait of a product per se or simply a matter of matching the right product to the right subject matter.** Rather, multiple factors affect learning outcomes and to make sense of student outcomes, analyses need to incorporate student characteristics, specifics of how the adaptive courseware is used, aspects of the course beyond the courseware product, and the way learning is measured to make sense of student outcomes.[23]

---

[23] *Means, B., Bakia, M., & Murphy, R. (2014). Learning online: What research tells us about whether, when and how. New York: Routledge.*

Table A1. ALMAP Grantees' Adaptive Products, Comparison Modes of Content Delivery, Subject Areas, Time Frames, Institution Types, and Experimental Designs

Institution & Adaptive Product	Adaptive Course Mode	Comparison Course Mode	Subject & Learning Measure	Time (wks)	Institution Type	Design & Baseline Equivalence (BE) Quality
Arizona State University Pearson/ Knewton	Blended independent use of Pearson online text plus Knewton with inquiry classes	Blended independent use of Pearson online text with inquiry classes	Math Course grade	17	4-year research university	QE* design, retrospective comparison, teachers same, same curriculum, can compute baseline equivalence (BE) (ALEKS test first term; Pearson test second and third term)
Capella University CogBooks	Fully online course with CogBooks with instructor content	Fully online course	Business Psychology Course grade	10	Online university	QE design, concurrent comparison, teachers different, different curriculum, demographic matching, no pretest for BE
Essex County College ALEKS	Blended lab with coach on call and ALEKS for homework plus study skills classes	Blended lecture class with choice of independent online homework tools	Math Algebra Course grade	15	2-year community college	QE design, concurrent comparison among different sections of both courses, teachers different, different curriculum, can compute BE (placement test)
Excelsior College (main campus only) Cerego	Fully online course with Cerego	Fully online course	Math Biology Course grade	8 15	Online university	QE design, concurrent comparison using student self-selection, teachers same, different curriculum, can compute BE (GPA)
Metropolitan College Cerego (Excelsior partner)	Blended course with Cerego homework	Lecture class	Math Biology Course grade		4-year university	QE design, retrospective comparison using student self-selection, teachers same, different curriculum, can compute BE (pretest)
NC State University Adapt Courseware	Fully online course with Adapt	Blended lecture course with either Cengage Coursemate or Textbook	Psychology Posttest	18	4-year land grant university	Experimental design, concurrent comparison, teachers different, can compute BE (SAT)

* QE = quasi-experimental

Table A1. ALMAP Grantees' Adaptive Products, Comparison Modes of Content Delivery, Subject Areas, Time Frames, Institution Types, and Experimental Design (Concluded)

Institution & Adaptive Product	Adaptive Course Mode	Comparison Course Mode	Subject & Learning Measure	Time (wks)	Institution Type	Design & Baseline Equivalence (BE) Quality
Rio Salado College Pearson/ Knewton	Fully online course plus Pearson online text with Knewton for homework	Fully online course	Writing Posttest	13	2-year community college	Experimental design, concurrent comparison, teachers same, different curriculum, can compute BE (placement test)
Saint Leo University Smart Sparrow	Blended course with Smart Sparrow	Fully online course	Marketing Posttest	8	Adult learning program within 4-year college	QE design, concurrent comparison, teachers same, different curriculum, can compute BE (GPA)
Shoreline Community College (Northeastern)	Fully online course with CogBooks with instructor content	Face-to-face lecture classes	English Economics Course grade	11	2-year community college	QE design, retrospective comparison, teachers same, different curriculum, can compute BE (COMPASS)
St. Petersburg College ALEKS LearnSmart	Blended labs with coach on call while using ALEKS (Math) or LearnSmart (English) for homework	FFace-to-face lecture classes	Math English Posttest	16	2-year community college	QE design, concurrent comparison, teachers same, different curriculum, can compute BE (courseware and faculty tests)
SUNY Empire State College CogBooks	Fully online course with CogBooks with instructor content	Fully online course with textbook	Math English Posttest	14	4-year university	QE design, concurrent comparison, teachers same, different curriculum, no data for BE
University of California, Davis Online Learning Initiative (OLI)	Blended lecture plus inquiry-based discussion sections with OLI	Face-to-face lecture class with discussion sections	Biology Posttest	12	4-year research university	QE design, concurrent comparison Terms 1 and 2; retrospective Term 3; teachers different; different curriculum; can compute BE (SAT)
University of Missouri-St. Louis Cerego (Excelsior partner)	Blended course with Cerego homework with instructor content	Lecture class	Math Biology Course grade	16	4-year university	QE design, retrospective comparison using student self-selection, teachers same, different curriculum, could compute BE for Term 1 but not Term 2 (pretests vary)
University of North Georgia, Gainesville LearnSmart Connect	Blended course with LearnSmart Connect	Blended course with MyReading Lab	ESL Reading English Reading Posttest	16	4-year university	QE design, mixed comparison (ESL retrospective; others concurrent), teachers mixed (same only for Reading), different curriculum, can compute BE (COMPASS)

Table A2. Student Outcomes Evaluation Participation by Institution

Institution	Term 1		Term 2		Term 3	
	Treatment	Comparison	Treatment	Comparison	Treatment	Comparison
Arizona State University	738	1,131	584	1,131	1,153	1,131
Capella University*	50	54	110	123	155	139
Essex County College	408	1,540	158	957	179	1,283
Excelsior College*	177	388	208	372	251	374
Metropolitan State University*	88	134	113	77	120	137
North Carolina State University	45	53	55	61	42	52
Rio Salado College§	112	120	197	215	163	150
Saint Leo University	70	84	110	79	81	95
Shoreline Community College*	47	48	50	48	11	21
St. Petersburg College*	297	309	501	528	165	152
SUNY - Empire State	47	45	41	42	25	25
University of California, Davis	548	467	441	251	987	718**
University of Missouri-St. Louis*	401	199	353	199	236	199
University of North Georgia - Gainesville	41	26	249	218	129	218
Total	**3,063**	**4,625**	**3,150**	**4,308**	**3,733**	**4,726**
% Reported course-repeating students	**9.4**	**13.6**	**13.7**	**17.2**	**8.5**	**11.2**

* Institution had two courses. ** Combined instructors from Terms 1 and 2. §Reflects data collected in Terms 2, 3, and 4.
Note: Approximately 12.4% (2,933) of students indicated they were repeating the course. However, only 118 of those indicating they were repeaters appeared in more than one term included in the ALMAP study. The rest of the repeating students were either assigned new student IDs or had taken the course previous to ALMAP.

Table A3. Faculty Participation by Institution

Institution	No. Same Instructors Term to Term	Term 1		Term 2		Term 3	
	Term 1–Term 3	Treatment	Comparison	Treatment	Comparison	Treatment	Comparison
Arizona State University*	1	11	13	10	13	11	13
Capella University	0	2	2	3	13	3	15
Essex County College	2	16	49	8	34	7	47
Excelsior College*	0	26	30	27	31	27	31
Metropolitan State University*	0	3	2	3	2	2	3
North Carolina State University	0	2	4	1	7	1	6
Shoreline Community College*	1	2	2	2	2	1	1
Rio Salado College*§	2	3	5	7	11	8	8
Saint Leo University*	0	3	4	5	3	4	5
St. Petersburg College*	0	11	11	19	19	7	7
SUNY - Empire State*	1	3	3	1	1	3	3
University of California, Davis	0	9	7	9	7	10	7
University of Missouri-St. Louis*	2	5	2	4	2	3	2
University of North Georgia – Gainesville*	3	4	4	8	8	6	8
Total	**12**	**100**	**138**	**107**	**153**	**93**	**156**

* Included same instructors between treatment and comparison (serving as controls) (13% of total instructors in the study). §Reflects data collected in Terms 2, 3, and 4.

Table A4. ALMAP Instructor Survey Response Rates for Terms 1, 2, and 3

Institution	Term 1			Term 2			Term 3		
	N Invited*	*N* Response	Resp Rate (%)	*N* Invited*	*N* Response	Resp Rate (%)	*N* Invited	*N* Response	Resp Rate (%)
Arizona State University	3	3	100	2	2	100	11	5	45
Capella University	2	2	100	3	1	33	3	1	33
Essex County College	22	11	50	8	2	25	7	5	71
Excelsior College	NA	NA	NA	15	5	33	17	5	29
Metropolitan State University	3	2	67	3	2	67	2	0	0
North Carolina State University	1	1	100	1	1	100	1	0	0
Shoreline Community College	2	2	100	1	1	100	1	1	100
Rio Salado College§	3	3	100	7	7	100	8	7	88
Saint Leo University	3	2	67	3	1	33	4	3	75
St. Petersburg College (administered own survey)	11	8	82	19	16	84	7	4	57
SUNY - Empire State	3	3	100	4	4	100	2	1	50
University of California, Davis	8	6	75	19	5	26	10	1	10
University of Missouri-St. Louis	2	2	100	4	3	75	3	2	75
University of North Georgia - Gainesville	5	5	100	8	6	75	6	5	83
Total*/response rate	**67**	**49**	**73**	**93**	**52**	**56**	**81**	**40**	**49**

NA = No data collected.
* Total invited is a subset of total instructor participation; not all instructors were invited to take the survey.
§Reflects data collected in Terms 2, 3, and 4.

Table A5. ALMAP Student Survey Response Rates for Terms 1, 2, and 3

Institution	Term 1			Term 2			Term 3		
	Total	N Response	Resp Rate (%)	Total	N Response	Resp Rate (%)	Total	N Response	Resp Rate (%)
Arizona State University	121	63	52	584	47	8	1153	116	10
Capella University	50	15	30	110	32	29	155	39	25
Essex County College	408	NA	NA	158	20	13	179	139	78
Excelsior College	177	80	45	218	66	30	264	88	33
Metropolitan State University	88	54	61	113	87	77	120	91	76
North Carolina State University	45	45	100	55	39	71	42	32	76
Shoreline Community College	47	32	68	17	10	59	11	3	27
Rio Salado College§	112	62	55	197	143	73	NA	NA	NA
Saint Leo University	70	NA*	NA*	110	57	52	81	54	67
St. Petersburg College	297	NA**	NA**	501	229	46	165	92	56
SUNY - Empire State	47	38	81	41	41	100	25	21	84
University of California, Davis	545	NA**	NA**	439	NA**	NA**	NA**	NA**	NA**
University of Missouri-St. Louis	401	NA**	NA**	353	113	32	236	NA**	NA**
University of North Georgia - Gainesville	41	7	17	264	34	13	129	9	7
Total*/ % response rate**	**728**	**396**	**54**	**2721**	**918**	**34**	**1171**	**684**	**58**

NA = No survey data submitted

* Data were not in an analyzable format (aggregated)

** Grantees used their own items that did not align with evaluator's suggested items.

*** Total reflects only those grantees with survey data.

§ Reflects data collected in Terms 2, 3, and 4.

Appendix B

Methodology and Analysis Details

ALMAP Analysis of Student Outcome Results

SRI examined student outcomes using data provided by the ALMAP grantees that met generally accepted criteria for rigorous evaluation. To be included in SRI's analysis, data for a side-by-side comparison had to represent at least 10 students in each condition being compared, include pretest measures related to the course subject, and if posttest data were used, the same measure had to be available for comparison and courseware students and the test could not be over-aligned to the courseware.

For those data sets meeting these criteria, an SRI analyst tested the impacts of the adaptive learning courseware on the core student outcomes—course grade, completion, and, when available, posttest score—using a hierarchical linear model analysis.

DESCRIPTION OF GRADE, POSTTEST, AND COMPLETION DATA

We reviewed the mean, standard deviation, and count for each continuous outcome variable (course grade, posttest) and the proportion and count for each categorical outcome variable (completing "1" or not completing "0") that the grantee gave us.

Course letter grades were translated to numeric values using the following assignments:

Letter Grade	Grade Value
A+	4.33
A	4
A-	3.67
B+	3.33
B	3
B-	2.67
C+	2.33
C	2
C-	1.67
D+	1.33
D	1
D-	0.67
E	0
F	0

In the case of dichotomous completion outcomes, the percentage of students labeled as "1" (the condition is true, that is, the student completed the course successfully with a grade of C- or better) was reported, and the standard deviation was omitted.

DETERMINING DATA TO INCLUDE IN THE SYNTHESIS

Next, we reviewed each comparison for potential bias. Bias occurs when the adaptive learning students and students in the comparison group differ substantially on measures of prior achievement. In effect, the two groups were not starting at the same place educationally. This makes comparisons of outcomes difficult to interpret. While we can statistically correct for minor differences in prior achievement, when those differences are large (one federal statistical standard defines "large" as greater than .25 standard deviation), no degree of statistical modeling can eliminate the bias in the impact estimates. In these cases, we did not include the case in our synthesis.

In many cases, grantees did not report data for prior achievement. Lacking the ability to either confirm or disconfirm equivalence between the adaptive learning and comparison groups, we reported the unadjusted outcome differences in reports to grantees, but we excluded these data sets from the synthesis described in this report.

These data set quality restrictions removed 413 comparisons (60%) of the data out of 687 possible comparisons provided by grantees.

MODEL FOR SYNTHESIS ANALYSIS OF GRADE, POSTTEST, AND COMPLETION DATA

We modeled 74 comparisons for course grade and 36 comparisons for posttest via a hierarchical linear model.

We used raw scores and computed standardized effect sizes taking pretest scores into account. We used pretests or pretest proxies when pretests were not available. The pretest proxies were college entrance examination scores (e.g., SAT, Accuplacer, COMPASS).

We computed adjusted mean outcome scores by setting the pretest score to the grand mean and using the hierarchical linear model shown below to compute mean outcomes for the adaptive learning and comparison groups. Standard errors for these adjusted means were also computed and used to form 95% confidence intervals around the means. For the covariates (pretest proxy, pretest) and outcomes involving a grade or continuous score (course grade, posttest), the raw effect was in the actual units of the measure. That is, the raw effect was the difference in mean grade points (course grade), while the posttest effect was in terms of posttest points.

For course grade and posttest, the general model used for those ALMAP data sets passing the screen for potential bias was a hierarchical linear model of the form:

$$\text{Outcome}=b_0+b_1\,\text{Treatment}+b_2\,\text{Pretest}+e$$

where

b_0........... The intercept for the comparison group (not reported)

Treatment.................. Dichotomous variable set equal to 1 for adaptive learning group students and 0 for comparison group students

b_1.................. Coefficient representing the adaptive learning effect (the expected difference in mean posttest scores between the adaptive learning and comparison groups after adjusting for the pretest)

b_2.................. Expected change in outcome for each unit change in pretest (not reported)

e.................. Residual error with a complex variance structure, taking into account clustering of students within classrooms and institutions.

For the case of course completion, which is a dichotomous outcome (taking the value of 1 or 0), an equivalent multilevel logistic regression model was estimated. We entered 73 cases for completion into a logistic regression model specified identically to the linear model above except for the dichotomous nature of the outcome and a logit link function. For this dichotomous variable, the raw effect was an odds ratio. This is a multiplier telling us what the increase in odds is for a yes outcome under the adaptive learning condition. For example, an odds ratio of 1.4 for completion says the odds of a treatment student completing are 1.4 times greater than those for a comparison student.

To put all these on a comparable scale, an effect size is provided. For continuous outcomes (course grade, posttest score), this is the difference between adaptive learning and comparison outcomes expressed in standard deviation units. For the dichotomous outcome of course completion, a translation to an equivalent continuous effect size was computed (technically, the effect size is computed as the log odds ratio divided by 1.65, a method attributed to Cox).

One way to interpret the effect size is to ask what rank a student who scored at the median (50th percentile) of the comparison group would be expected to score under the treatment (i.e., in the adaptive version of the course). This translation is tabulated below. An effect size of 0.2 would indicate that an average comparison group student would score at the 58th percentile if in the adaptive version of the course.

Effect size	0	0.2	0.4	0.6	0.8	1.0
Equivalent percentile	50	58	66	73	79	84

Outcome data were analyzed for all students and for Pell Grant students. For individual grantee course implementations with adequate data, analyses also were run for part- time students and for students repeating the course, but the latter analyses were not aggregated across grantees for this report because of the small number of qualifying data sets.

WEIGHTING STRATEGIES

When averaging effect sizes, analysts used different models and weights for different aggregation purposes: (1) combining effect size estimates across terms for a single course, (2) estimating an overall effect across students who took different courses, and (3) estimating the effect of moderators.

When averaging the effect size estimates over a term for individual courses taught within a given institution, analysts weighted the effect size estimates by the inverse variance, assuming a fixed-effects model. A fixed-effects model using inverse variance weighting was selected because it is reasonable to assume that every term of data for an individual course is measuring the same effect size for that course. In other words, we assumed that all terms for a given course were functionally equivalent to each other and that any deviation away from the true mean was due to sampling error. In this case, we were not generalizing beyond the population of coursetakers at the specific institution.

When averaging effect size estimates from different institutions by use case (blended adaptive versus lecture, online adaptive versus online, and blended adaptive versus blended), subject domain (math, English, social science, and biology), content authoring (four categories), and adaptivity type (learning object-macro versus lesson/unit-micro), the effect size estimates were weighted by the inverse of the total variance, assuming a random-effects model. The total variance is defined as the sum of the within-study variance and the between-study variance (commonly referred to as tau squared or $\tau 2$). This is because it was not reasonable to assume that the different studies being averaged were functionally identical or shared the same true effect size.

The assumptions underlying the random-effects model should be kept in mind when reviewing the ALMAP meta-analyses. First, a random-effects model allows for different studies to have their own true effect size, but assumes that the collection of true effect sizes among the studies included in the analysis have a normal distribution around a central value. We do not have enough studies in the various sub-groups of the moderators to fully know whether or not this assumption holds. Secondly, the random-effects model relies on the computation of the between-study variance (commonly referred to as tau-squared or $\tau 2$). Our estimation of tau-squared is likely imprecise due to the relatively small number of studies.

An alternative possible approach would be to weight simply by sample size, without imposing a model (random or fixed). The advantage of this approach is that the assumption of normality does not need to be met. This approach allows flexibility in the sense that it does not assume that the effect size estimates of the studies being aggregated are functionally equivalent or having a meaningful relationship underlying them.

Table B-1 shows how the estimated means and confidence intervals changed according to the different weighting strategies and modeling selections. It should be noted that while the confidence intervals change with the different analytic approaches, the pattern of significant and insignificant average impacts does not.

When conducting a moderator analysis of one of the three possible moderators tested—subject domain, content authoring, and adaptivity type—we returned to using inverse variance weighting using a fixed-effects model. A moderator analysis by its very nature assumes that a meaningful relationship exists among the effect size estimates being aggregated, and tests whether this relationship is altered by the moderating variable in a systematic way. For this reason, we chose a fixed-effects model for the moderator variable analyses. (Recall a random-effects model assumes effect size estimates vary in non-systematic ways, and tries to adjust for that.)

Table B-1. Course Grade Outcome Analysis, by Weighting Procedure

Comparison	Number of Effect Estimates	Weighted Mean	Lower Confidence Interval	Upper Confidence Interval
Weighting by Sample Size				
Blended Adaptive vs. Lecture	12	0.039	-0.028	0.106
Online Adaptive vs. Adaptive	14	0.127	0.050	0.204
Blended Adaptive vs. Blended	2	-0.177	-0.314	-0.041
Fixed Effects, Weighting by Inverse Within-Study Variance				
Blended Adaptive vs. Lecture	12	0.036	-0.029	0.102
Online Adaptive vs. Adaptive	14	0.138	0.062	0.214
Blended Adaptive vs. Blended	2	-0.085	-0.213	0.042
Random Effects, Weighting by Inverse Total Variance				
Blended Adaptive vs. Lecture	12	0.057	-0.053	0.166
Online Adaptive vs. Adaptive	14	0.125	0.030	0.220
Blended Adaptive vs. Blended	2	0.028	-0.504	0.559

Cost Analysis Overview

Data were collected from grantees using an Excel spread sheet to capture information on a variety of costs related to course delivery for both the ALMAP courses using adaptive courseware and comparison courses. These data were the following:

- Course enrollment
- Technology implementation costs and annual technology support costs
- Per-student costs for materials, such as textbooks and access codes
- Number of hours and associated salary and wage costs for faculty, adjuncts, teaching assistants (TAs), and other personnel spent on initial course development
- Number of instructors (faculty, adjuncts, TAs, others) involved in course delivery and their hourly cost
- Number of hours spent by instructors on all course delivery activities (diagnosis of skill, presentation, other interaction, course preparation, progress monitoring, test proctoring, and evaluating tests and assignments) and the associated cost.

For each course, multiple iterations of the cost capture tool were reviewed by SRI analysts, and grantees were involved in discussions with SRI staff to clarify and refine the data. Once finalized, cost data were analyzed to calculate metrics for

- Total costs
- Total development costs (Note that development costs were not amortized over the expected life of the course. This choice was made to reflect the actual timing of financial outlays to help policy makers and administrators with budgetary considerations.)
- Total instructional labor costs (the sum of the costs of all instructional activities by all involved instructors)
- Materials costs, which included the costs to students for textbooks and access codes, as well as the costs paid by institutions or with grant funds to cover these items
- Per-student costs (total costs divided by student enrollment to derive a cost per student for both adaptive course sections and comparison course sections for each institution). Cases where the difference in sample size between treatment and comparison conditions exceeded 20% required additional modeling and results would have been difficult to interpret. For this reason, such cases were removed from the ALMAP cost data set.
- Using those comparisons meeting our inclusion criteria, we computed the average adaptive and comparison group costs for each use case group (Blended Adaptive vs. Lecture, Online Adaptive vs. Online, Blended Adaptive vs. Blended). Initial costs were examined using Term 1 data. Ongoing costs were examined by aggregating across Terms 2 and 3. The direction of average cost change between the adaptive condition and the comparison condition was classified as an increase, no change, or decrease in costs. Ratios of cases showing an increase, decrease, or no change were tallied for each use case group and described qualitatively.

SRI Education™

SRI Education, a division of SRI International, is tackling the most complex issues in education to help students succeed. We work with federal and state agencies, school districts, major foundations, nonprofit organizations, and international and commercial clients to address risk factors that impede learning, assess learning gains, and use technology for educational innovation. SRI International, a research center headquartered in Menlo Park, California, creates world-changing solutions making people safer, healthier, and more productive.

Silicon Valley
(SRI International headquarters)
333 Ravenswood Avenue
Menlo Park, CA 94025
+1.650.859.2000
education@sri.com

Washington, D.C.
1100 Wilson Boulevard, Suite 2800
Arlington, VA 22209
+1.703.524.2053

www.sri.com/education

STAY CONNECTED