

The Use of Adaptive Learning Technology in Mathematics and Biology

Jason Bryer¹ & Bruce Homer²

¹ City University of New York, School of Professional Studies

² City University of New York, The Graduate Center

Author Note

This project was funded by the Bill & Melinda Gates Foundation as part of the Adaptive Learning Acceleration Program (ALMAP). We would also like to acknowledge our colleagues who helped in implementing this project: Scott Dalrymple, Peter Cain, Patti Croop, Bethany de Barros, Margie Dunn, Heidi Langer-Atkinson, Paul Mumma, John Racquet, Andrew Smith-Lewis, and Jane Weyers. Supplemental materials including analysis scripts and data are available at <https://github.com/jbryer/Cerego-Paper>

Correspondence concerning this article should be addressed to Jason Bryer, 119 W 31st St, New York, NY 10001. E-mail: jason.bryer@cuny.edu

Abstract

The use of technology to adapt learning to students' performance is not new (Skinner, 1958). However, with increased availability of computers, tablets, and smart phones, reimagining of Skinner's initial teaching machines has proliferated in the EdTech market. This study examines the use of Cerego, and adaptive learning technology framework, as a supplement to asynchronous online mathematics and biology courses. Using propensity score analysis to adjust for selection bias, results suggest that students who use Cerego as a supplement score between 5 and 13 percentage points higher than their non-using counterparts in most weekly quiz grades. Results suggest there is little or no impact on overall course grade, in large part due to the inconsistent use of Cerego throughout the course.

Keywords: computer adaptive learning, online learning

Word count: X

The Use of Adaptive Learning Technology in Mathematics and Biology

Over the past decade, online learning has become an increasingly important component of higher education. This has accelerated since the COVID-19 pandemic. Data from the National Center for Education Statistics' Integrated Postsecondary Education Data System (IPEDS) indicates that there are over 5.2 million students learning online, and that approximately 32% of higher education students will take at least one online course. In the coming years, the importance of online learning will continue to grow. In a recent survey of academic leaders, over 70% indicated that online learning is "critical to their institutions long-term strategy" (Allen & Seaman, 2013). The number of online offerings has increased every year for the past decade, and is projected to continue to increase (B. Means, Bakia, & Murphy, 2014).

B. Means et al. (2014) suggest that there are at least four major factors driving the push for more online education. First, with the increased power and decreased cost of mobile technology, we are living more and more of our lives online, and students now expect education to be part of this online shift. A second factor is the belief of many educators that online learning can address a number of the more difficult challenges currently facing education. For example, Wise (2010) argues that technology can help solve several of the most pressing crises in education, including the scarcity of skilled teachers. A third factor driving the push towards online education is economic, as online courses are generally less expensive to offer than face-to-face classes. The final driving force identified by Means et al. is the belief that digital technologies have the potential to deliver a more enriched educational experience compared to what students typically experience in traditional classrooms. Regardless of the reasons, it is evident that more and more students will receive some, if not all, of their education online.

The growing importance of online learning calls for even more research on this topic. Much of the early work in this area addressed the issue of whether or not online learning was

superior to (or at least as effective as) traditional face-to-face instruction. Although the findings have been mixed, in general, online learning has been found to be at least as effective as traditional education (e.g. Johnson, Aragon, Shaik, & Palma-Rivas, 2000; Maki, Maki, Patterson, & Whittaker, 2000; Neuhauser, 2002). More detailed examinations have generally found that online (or distance) education can result in improved learning outcomes under certain conditions. For example, in one of the first meta-analyses on this topic, Bernard et al. (2004) found that there were no overall differences between students' learning outcomes in traditional versus online learning. However, when the authors examined different types of online learning, they found that although students in synchronous online classes fared worse than students in traditional classes, students in asynchronous online classes actually had better learning outcomes. Similarly, Barbara Means, Toyama, Murphy, Bakia, and Jones (2009) used meta-analysis to examine studies that directly contrasted face-to-face with online learning, and found that students in the online classes tended to have better learning outcomes compared to students in the face-to-face classes. When the authors took a closer examination of the studies, they found that this effect was due solely to an advantage for blended classes, in which online and in-person instruction was combined. The authors point out that the blended classes often had additional learning elements and increased instructional time compared to the control classrooms, making it difficult to conclude that it was the online element per se that was having the beneficial effects. This parallels arguments that it is instructional time and techniques that matter, not the medium of instruction itself (Clark, 1994). Nonetheless, different media have different affordances and challenges – and effective education requires knowing how best to take advantage of the strengths of the instructional medium, whether online or face-to-face. As educators increasingly rely on online learning, there is a critical need for research that provides insight into the factors that support effective online education.

One of the ways in which online learning can potentially enhance education is through personalized or individualized instruction. The idea of using technology to adapt learning to

a student's needs is not new. For example, Skinner (1958) suggested that "teaching machines" could be used to enhance a student's learning, in part by being adaptive, presenting "just that material for which the student is ready" (p. 971). With Skinner's teaching machines, it is content and pace that are adapted to an individual student's learning. With modern technology, there is a much greater option for what can be modified, and therefore, considerable variability in what is actually meant by "personalized" instruction. B. Means et al. (2014) point out that it has been used to mean everything from adapting content to match students' interest, to allowing students to choose the method of learning to match their preferences, to adapting the pace and content of what is taught to match students' knowledge. In the National Educational Technology Plan, the US Department of Education (2016) identifies personalized learning as taking place when the pace of learning, the instructional approach, learning objectives, and instructional content are optimized for each learner, matching their needs and interests.

Given the variation in what is considered to be personalized education, it is not surprising that the evidence in support of personalized learning has been mixed: The effectiveness of adapting instruction depends on which learner features are being adapted for and how the personalization is being implemented. For example, in their review of the literature, Pashler, McDaniel, Rohrer, and Bjork (2008) found very little evidence that adapting instruction to match a student's learning style had any effect on learning. On the other hand, Homer and Plass (2010) found evidence that matching level of interactivity in online simulations to learners' executive functions can positively affect learning. The authors found that, controlling for prior knowledge, students with higher levels of executive functions had better learning outcomes with more interactive simulations, while students with lower levels of executive functions had better learning outcomes with less interactive simulations.

Another emerging area of work on adaptive learning systems attempt to take into account the affective state of the learner. For example, Affective AutoTutor (D'Mello & Graesser, 2010) is a modified version of the intelligent tutoring system AutoTutor that

attempts to automatically detect and adapt to the emotional state of the learner. AutoTuto uses contextual cues from the ongoing dialogue taking place in the tutoring system, as well as gross body cues and facial expression (detected via cameras) to infer learner affect. The intelligent tutoring system will then respond, for example, with an empathic and supportive comment – with compatible expressions being made by the on-screen AutoTutor agent (e.g., the agent may say, “I know this is difficult material, but I think you’re on the right track!”, while displaying an empathetic look). In their review, D’Mello and Graesser (2012) report that adapting for learners’ emotions in Affective Autotutor results in significantly greater learning outcomes compared to the standard version of AutoTutor.

The number of studies exploring adaptive learning technology has been growing over the last decade Muñoz et al. (2022). The reviews by Marten et al and Munoz et al revealed the variety and types of adaptive learning technology as well as the varying emphasis ranging from student self-regulation to learning efficiency, the effectiveness of adaptive learning technology on student academic performance is not clear. An older meta-analysis conducted Verdu and colleagues (2008) identified 15 empirical studies with effect sizes ranging from 0.1 to 1.0. In the nine empirical studies examining academic outcomes identified by Lim, Lim, and Lim (2023) conducted since 2016, seven had statistically significant results, one had a null finding, and one had mixed results (positive for chemistry, null for other subjects).

This study was part of the Bill and Melinda Gates Foundation Adaptive Learning Market Acceleration Program (ALMAP) which funded 16 projects to examine 15 adaptive learning technologies. Results from Education (2016) indicated mixed results. Four of the 15 studies showed higher course grades with the remaining had no statistically significant impact on grades. The context of using adaptive technology often had an impact on results. For example, studies where adaptive learning replaced lectures had a positive impact on posttests whereas already fully online courses saw a smaller, but still positive, impact.

Although there is a growing body of research supporting the personalized learning for

different learner characteristics, the most robust data come from research on adapting instruction to account for learners' prior knowledge. For example, Kalyuga and his colleagues have identified the expertise reversal effect (e.g., Kalyuga, Ayres, Chandler, & Sweller, 2003), which occurs when learning scaffolds that help less knowledgeable students actually hinder learning in more expert learners. For example, instruction with worked examples provides learners with a problem and the solution to the problem. This approach has been shown to be quite effective for novice learners, but can induce a high cognitive load and hinder learning in more experienced learners (Kalyuga, 2007; Kalyuga et al., 2003).

Adapting instruction to students' knowledge is very much a part of effective face-to-face instruction. Good teachers will ask students questions to find out what they know and don't know, and then adapt their instruction to correct students' misconceptions and to teach things that are just beyond their current knowledge, i.e., that are in the students' "zone of proximal development" (Vygotsky, 1978). In effect, good teachers have a model of their students' conceptual understanding and then use this model to guide instruction.

Similarly, effective online learning systems must also have a model of students' knowledge, and then adapt instruction to match this model. Ideally, this model should update as students learn so that it is an accurate representation of the students' knowledge (i.e., there should be ongoing assessment that should feed into the system). The goal of the current study is to examine the potential benefits of adding one such adaptive learning system, Cerego, to online college courses.

The Cerego Learning Engine

Cerego is a commercial learning engine (free to individual users) that allows users to study existing courses (sets) or design their own learning sets by defining specific items. The system models learners' memory of all items in a learning set at any given time, and, using a spaced repetition algorithm, schedules rehearsal of the items to reduce forgetting. Spaced

repetition algorithms present students with more difficult items more frequently and less difficult items less frequently. For Cerego, students are presented all items but frequency and timing for when students are represented an item is determined by their previous responses. Figure 1 shows a screen shot from the Cerego learning system for a set used in an introductory biology course. For this item, students are shown a diagram of a cellular process and asked whether or not they know it. If they say that they don't know it, they are shown the correct answer. If they say that they do know it, they are then asked to identify the cellular process, and are then given feedback as to whether or not they have provided the correct answer. The Cerego system keeps track of whether or not the student knew the correct answer and adjusts the learner's memory model accordingly.



Figure 1. Screen Shots from Cerego Adaptive Learning Technology

There is a broad range of topics available in Cerego sets, and users have the option of learning one of the existing sets or creating their own. For the present study, sets were created for three asynchronous online college courses: two biology courses and one math course. Each course is divided into 8 distinct modules, and learning sets were created that mapped directly onto the course content for each module (Table 1). Use of Cerego was optional for students in the courses.

These sets were designed by teachers from the schools in which the system was being used. During a usual study session, new items are introduced and later tested. In addition to the new items, previously learned items are rehearsed as well. Learners can specify how many items they want to learn at a time, and are given a system generated schedule for when they should return to continue studying. The type of adaptivity provided by the system can therefore be described as adaptivity in the core learning mechanic and learning progression based on current knowledge (Homer & Plass, 2010).

Table 1

Course Topics by Module.

Module	Biology A	Biology B	Mathematics
1.00	What is Biology? Introduction to Chemistry	Introduction to the Science of Life	Unit Conversions and Metric Units
2.00	Biomolecules: Carbohydrates, Lipids/Proteins, Nucleic Acids/Cells	Introduction to Chemistry	Percents and Earning Interest
3.00	Energy of Life/Photosynthesis	The Cell	Compound Interest in Savings Plans and Investments
4.00	Midterm Examinations	Energy	Compound Interest in the Other Direction: Loans, Credit Cards, and Mortgages
5.00	Mitosis/Meiosis	Inheritance, Chromosomes and the Idea of a Gene	Becoming a Critical Consumer of Statistical Information
6.00	Patterns of inheritance/Molecular Biology of the gene/Biotechnology	DNA	Mathematics in the Arts and Nature
7.00	Evolution of Biological Diversity/Ecosystems	Evolution	The Mathematics of Voting
8.00	Final Examinations	Bacteria, Viruses and Prions	Understanding Uncertainty in Diagnostic Testing

Method

Participants

Students from three existing asynchronous online undergraduate college courses served as participants for the study from 2015 to 2017. The courses included Biology A ($n = 193$), Biology B ($n = 220$) and Mathematics ($n = 1,347$). The students ranged in age from 18 to 65 (mean age = 36), with roughly equal numbers of male and female students in each course. The students came from a diverse ethnic background, with 57% identifying as White, 20% identifying as African American, 3% identifying as Asian American, 12% identifying as Hispanic, and 8% indicating “other” when asked about their ethnic identity. All students were fluent in English. See Appendix A for descriptive statistics.

Materials and Procedures. For this study, the use of Cerego was designed to supplement already existing courses which are prescriptive in nature. That is, the course content, readings, homework, exams, and discussion topics, are the same from section-to-section. Under this format, instructional faculty serve as facilitators to answer student questions, grade, and monitor discussions. These courses are entirely asynchronous and divided into eight modules. Table 1 outlines the topics for each module. The Biology course was offered only in an 8-week format but the Mathematics course was offered in both an 8-week and a 15-week format. Given that fewer than 9% of students opted for the 15-week format, and once we verified there were no statistically significant differences between students in these two formats, the analysis was conducted with a single mathematics cohort.

Data for this study was collected across six eight-week terms (starting January, March, May, July, September, and November) and three 15-week terms (starting January, May, and September). Biology A represents the original course prior to the start of the study and ran in January, March, and May. This course had two large assessments given in weeks four and eight. This course underwent a major revision, in part to better align the course content to the content developed in Cerego, as well as to change the assessment model to weekly

assessments. Biology B represents the revised course that ran in July, September, and November. As a result, Biology A has three outcome measures (midterm, final, and final grade) whereas for Biology B and Mathematics we have nine outcome measures (weekly quiz and final average).

At the beginning of each term, students were sent an introductory email and survey along with information on how to login to Cerego. This information was also included in the course as part of the introductory materials. A short video was also produced to give students an introduction to using the Cerego system. A total of 16 modules were created in Cerego that mapped directly to the course modules (Table 1). Students self-selected to utilize Cerego, however an incentive of 10 points for each module quiz were offered for students who used Cerego for at least one hour during that module. These 10 points were then subtracted for the analysis conducted here, therefore the total score corresponds to the students' final course average without the incentive points. Additionally, demographic data was collected from the institutions student information system, time spent using Cerego from the software vendor, and a post-survey to assess student satisfaction with Cerego were collected.

Analytic Strategy

Randomized control trials (RCT) are generally considered the gold standard for estimating causal effects. However, in this study, randomizing students to treatment and control groups was not feasible. To address the selection bias introduced by the lack of randomization, propensity score analysis (PSA) using matching and stratification (P. R. Rosenbaum & Rubin, 1983) was used. The propensity score is defined as the “conditional probability of assignment to a particular treatment given a vector of observed covariates.” Research comparing the use of propensity score methods with randomized experiments have shown that causal estimates from observational studies using propensity score methods are generally consistent with those from randomized experiments (Cook, Shadish, & Wong, 2008; Shadish, Clark, & Steiner, 2008).

In practice propensity score analysis is typically conducted in two phases. In Phase I, propensity scores are estimated using logistic regression (see Table 2 for summary of logistic regression results). The fitted values from the logistic regression are used as estimates of the propensity scores. Both matching (Sekhon, 2011; Stuart, 2010; Stuart & Rubin, 2008) and stratification (Raudenbush, Hong, & Rowan, 2003) was used to estimate effects (P. R. Rosenbaum, 2012) (see also Rosenbaum, 2012). Matches were formed such that the absolute distance between propensity scores between treated and control students were minimized after stratifying on gender and ethnicity. For stratification, quintiles were calculated on the propensity scores defining five strata. To check balance, dependent sample t-tests were performed for each covariate and standardized effect sizes estimated for each stratum for the stratification method. Statistical significance is not generally recommended as the sole approach for checking balance (Paul R. Rosenbaum, 2002) as statistically significant differences are often detected in large samples even if the standardized effect size is small. Appendix B provides balance plots showing the overall standardized effect size before (in red) and after (in blue) stratification. In virtually all cases, the standardized effect sizes were substantially reduced after stratification except in cases where balance was present prior to stratification. Moreover, the standardized effects for all covariates after stratification and matching are below 0.1, or accounting for less than 1% of variance.

In Phase II, average treatment effects are estimated using dependent sample t-tests for the matched method. For the stratification method, independent sample t-tests for each stratum which is pooled to provide an overall estimate of the treatment effect.

Results

Data was retrieved from the College's student information system, online learning management system, and the online survey software. Once data was merged, missing covariates were imputed using a multiple imputation procedure (S. van Buuren, 2012) using the mice (Stef van Buuren & Groothuis-Oudshoorn, 2011) package in R (R Core Team,

2023). To test whether data were missing at random, a binary shadow matrix was created whereby values missing prior to imputation were coded with 1s and non-missing values were coded with 0s. A separate logistic regression was performed combining the imputed matrix with the shadow matrix. Income for Biology B was the only missing covariate indicator that had a statistically significant regression coefficient. Since this had a negligible impact on the estimated propensity scores, it was not included in the final propensity score estimation.

With the imputed dataset, propensity scores were estimated using logistic regression (see Appendix C) where treatment was defined as a student using Cerego sometime during the course. Balance was checked using covariate balance assessment plots (see Appendix B) using the **PSAgraphics** package (Helmreich & Pruzek, 2009) in R. These plots show the standardized effect size before (in red) and after (in blue) propensity score adjustment. The standardized effect size was reduced for all covariates where the initial effect size was large and with one exception for Biology B, the adjusted effect sizes are below 0.1 indicating that each covariate accounts for less than 1% of the variance indicating sufficient balance was achieved.

Students who used Cerego at any point during the semester were classified as being in the treatment group for estimating the propensity scores. This is also the treatment indicator for estimating the effect on overall course grade. However, stratification and matching were performed separately for each module outcome where treatment is determined by whether the student used Cerego for that particular module. As a result, a student may be in the treatment for one outcome analysis but in the control for another.

Both stratification and matching were used to estimate effects. Matching was conducted using the **Matching** package (Sekhon, 2011). Specifically, each treatment student was matched to one control student who matched exactly on gender (also ethnicity for mathematics) then by propensity score within a specified caliper of 0.15 standard deviations. For stratification, independent t-tests were performed for each strata and aggregated to

provide an overall effect size estimate. For matching, dependent sample t-tests were performed. Table 3 provides numeric results for all statistical tests and Figures 2, 3, and 4 present the results for Biology A, Biology B, and Mathematics, respectively. The Figures also provide the number of students who used Cerego for that module along with the average number of minutes Cerego was used by students. In most cases, these results indicate there is little or no statistically significant result on students' use of Cerego on their overall course grade. However, for most of the quiz grades, there are statistically significant results indicating that students who used Cerego scored anywhere from 5 to 13 points higher on that module's quiz than their non-using counterparts.

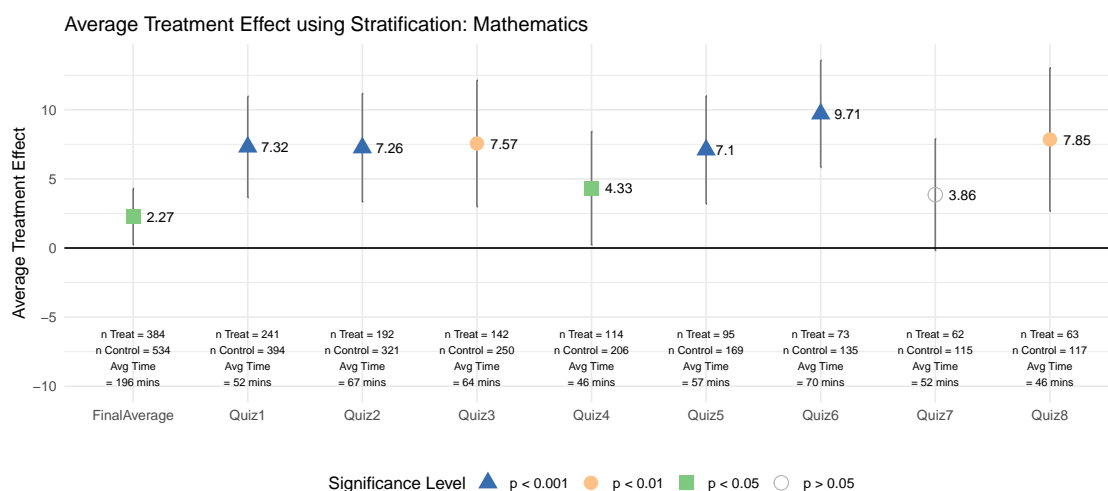


Figure 2. Mathematics Results

Student Satisfaction

At the completion of the courses students were surveyed about their satisfaction and use of Cerego within the course. The survey was divided into two parts: part one asked students about their overall satisfaction with Cerego using a five point Likert scale from strongly disagree to strongly agree; and part two asked students how Cerego compared to other learning modalities on a four point Likert scale ranging from not true to always true.

Figures 5, 6, and 7 provide the results for each question on the student satisfaction

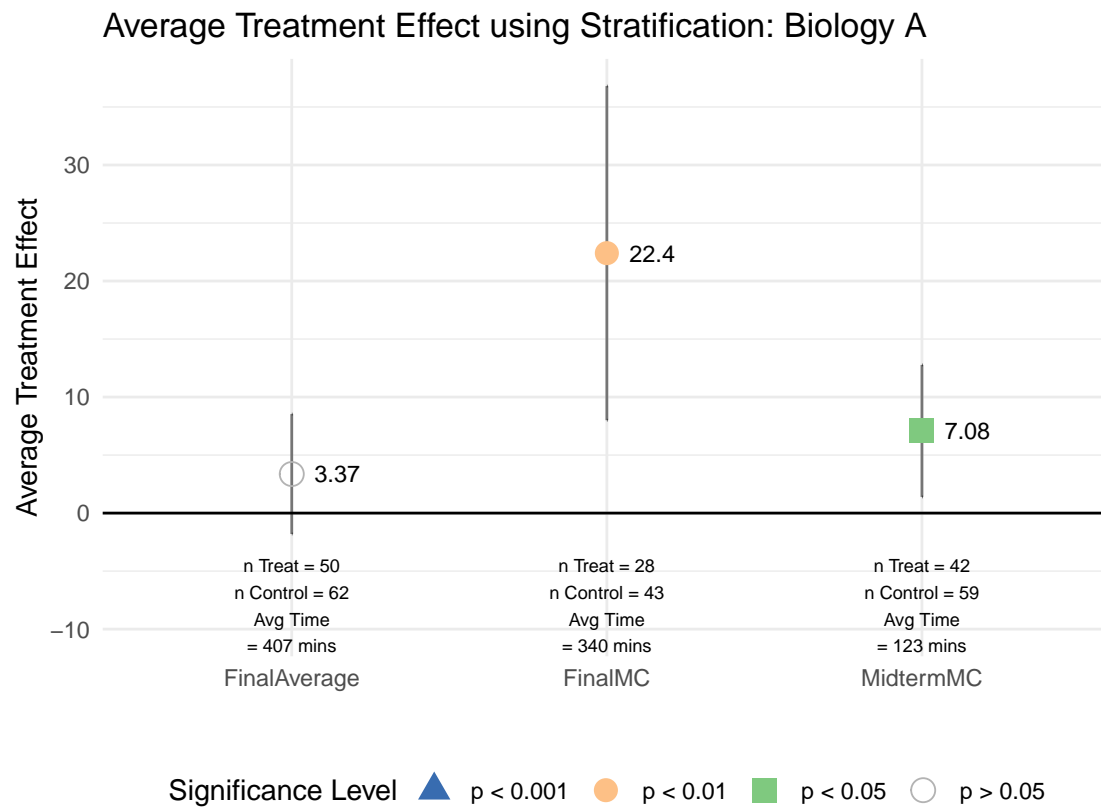


Figure 3. Biology A

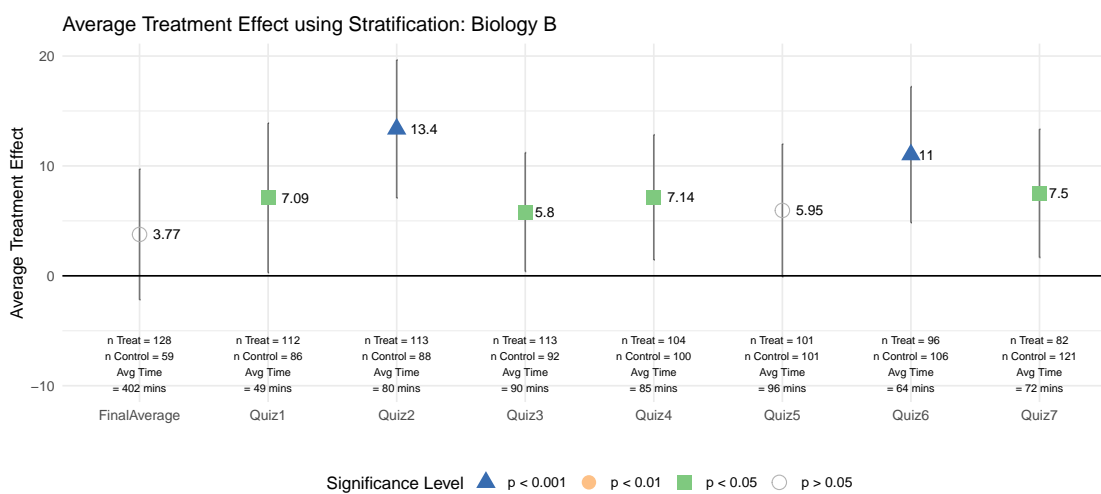


Figure 4. Biology B Results

Post Semester Student Survey Results

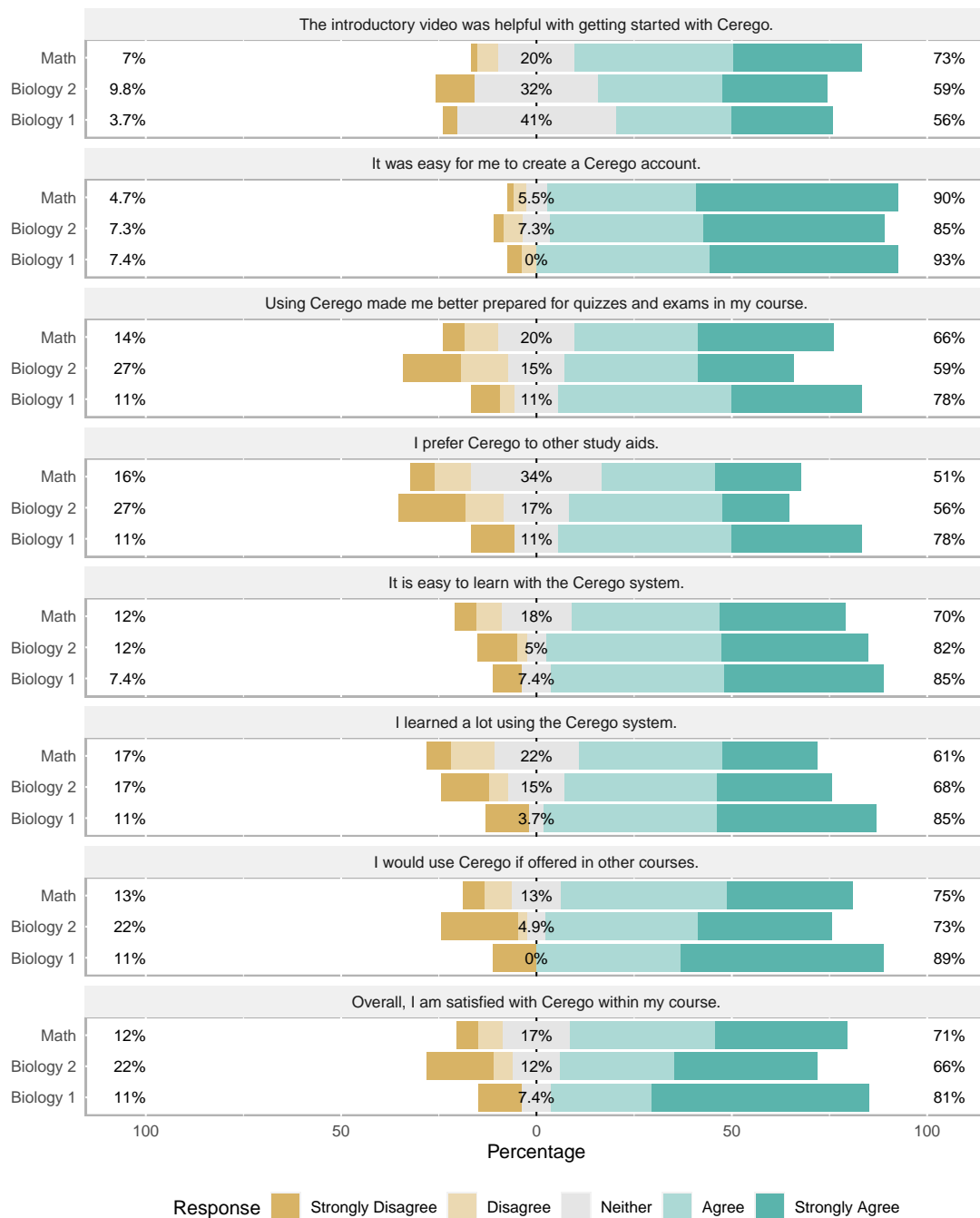


Figure 5. Student Satisfaction with Cerego

What do you think of Cerego?

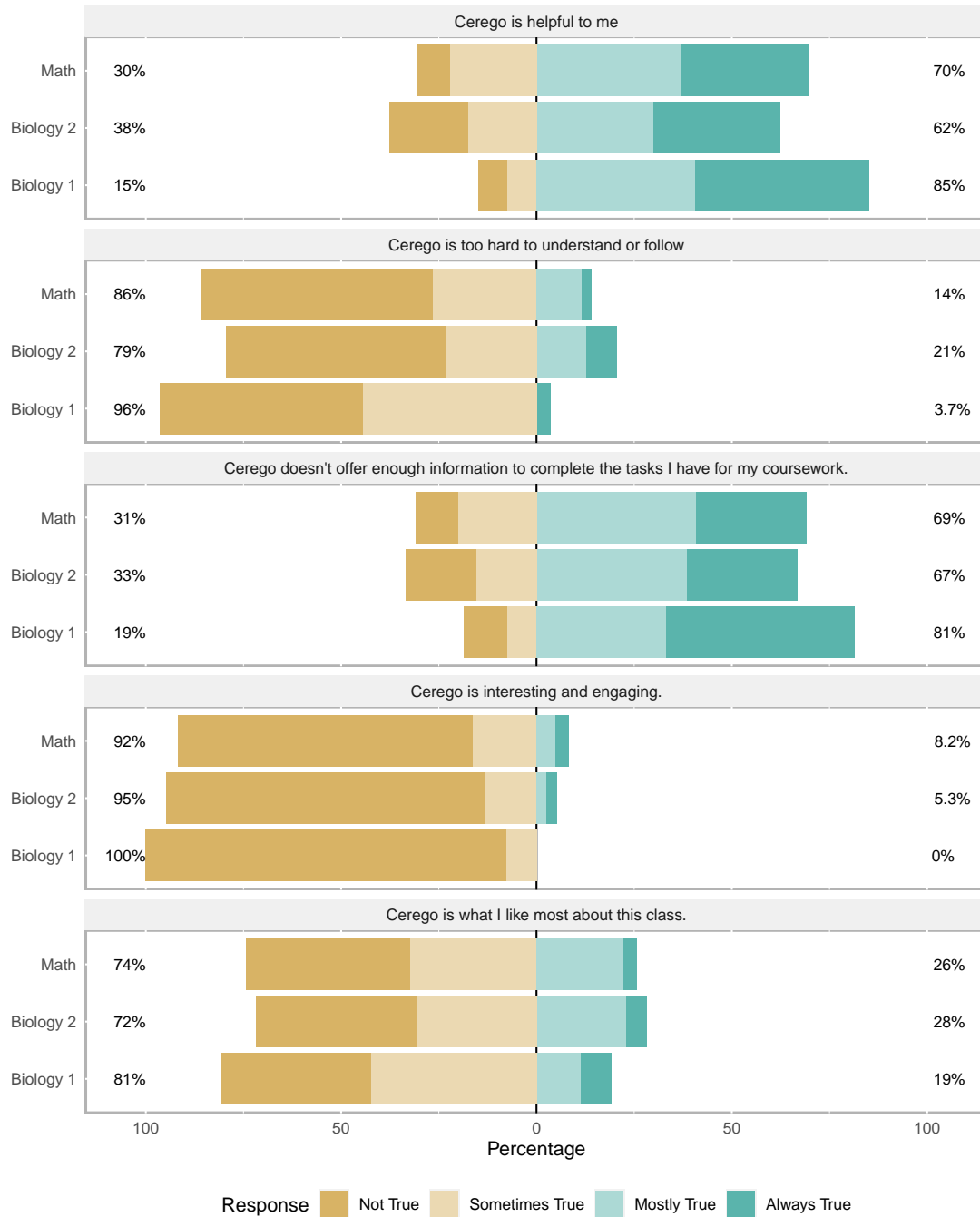


Figure 6. What do you think of Cerego?

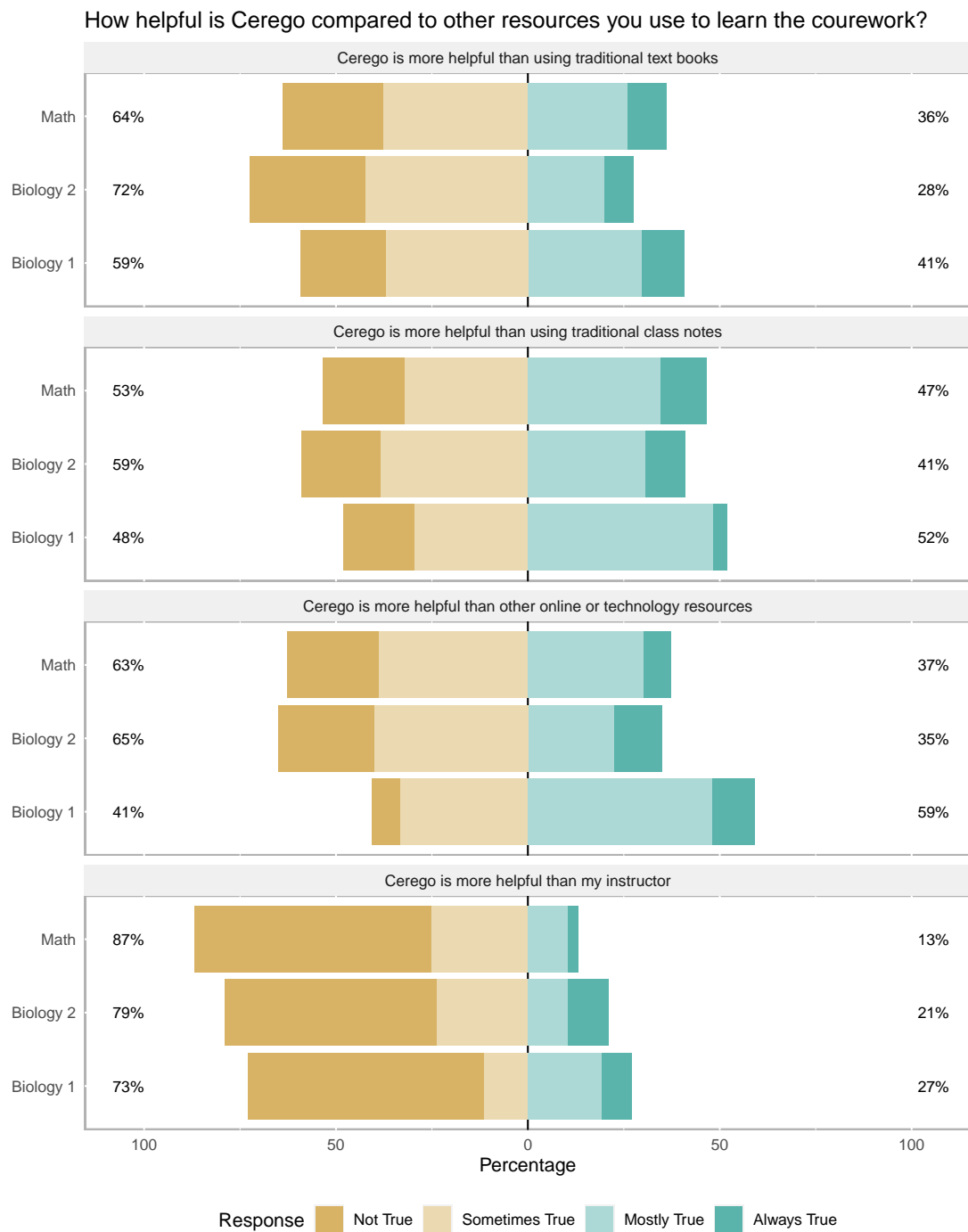


Figure 7. How helpful is Cerego compared to other resources you use to learn the coursework?

survey. Approximately 71% of students were overall satisfied with Cerego within the course and only 14% dissatisfied. In terms of how Cerego compares to other instructional features in the courses, results are mixed. Although they were satisfied the results suggest that Cerego is best situated as a supplement instead of a primary instructional strategy.

Discussion

With the proliferation of computers, tablets, and smartphones, utilizing them for education is an obvious endeavor. It is possible that technology will help educators better differentiate instruction in order to instruct students within their zone of proximal development (Vygotsky, 1978). The results of this study provide evidence that adaptive learning technology can have a positive impact on students' learning especially when implemented in asynchronous online courses where there are no other adaptations of the curriculum presentation.

The choice of two different subject areas also indicate that different content areas may benefit more from adaptive learning technology such as Cerego. When developing the content for Cerego, there were more issues with the Mathematics items than the Biology items in part due to the fact that the mathematics curriculum emphasized process whereas Biology had a larger portion of the curriculum focused on definitions and core concepts. Cerego utilizes a flashcard metaphor, which is a learning strategy often used when memorization is desired. Although there are positive results for Mathematics, future studies may wish to examine the method of adaptation in relation to the type of content.

Few studies of adaptive learning technologies explored both student academic performance and student satisfaction. Our data not only shows the positive impact of adaptive learning technology on student academic outcomes, but students are also generally satisfied with the technology. However, the survey results suggest that adaptive technology should not necessarily replace current instructional strategies but instead supplement current

pedagogy.

This study had a number of limitations, first is the issue of causality. As with many studies in education, randomization is often not possible or feasible. For this study we were required use observational data with propensity score methods, although it should be noted that propensity score analysis has been shown to provide consistent results with randomized designs (Cook et al., 2008; Shadish et al., 2008). Since this was a planned observational study we were able to ensure that important covariates were available for adjusting for selection bias. However, as with any quasi-experimental design, the results are potentially sensitive to any unmeasured confounders.

Additionally, this study did not take full advantage of all the features Cerego provides. Specifically, Cerego provides a rich instructor dashboard where instructors can review the performance of individual items, modules, and individual students. The purpose of these tools is to provide instructors with information they can use to adapt their teaching to students' performance. Given the positive trends for students who utilized Cerego, future studies may wish to integrate the tool more fully into the curriculum.

This study is an important contribution to the literature on adaptive learning technology. This study utilized a strong quasi-experimental design with a relatively large sample size that assesses both academic achievement and student satisfaction across two important subjects for STEM majors. Our results suggest that the use of adaptive technology is a promising for increasing student outcomes, especially on localized, module (or unit) assessments. Although additional research is necessary, these results support that adaptive technologies, and Cerego in particular, are useful tools for instructors to provide for students but should considering making them a more integral component of the curriculum.

References

- Allen, I. E., & Seaman, J. (2013). *Changing course: Ten years of tracking online education in the united states*.
<http://www.onlinelearningsurvey.com/reports/changingcourse.pdf>: Sloan Consortium.
- Bernard, R. M., Abrami, P. C., You, Y., Borokhovski, E., Wade, A., Wozney, L., & Huang, B. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research*, 74(3), 379–439.
- Clark, E. R. (1994). Media will never influence learning. *Educational Technology Research and Development*, 42(2), 21–29.
- Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, 27(4).
- D'Mello, S., & Graesser, A. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction*, 20(2), 147–187.
- D'Mello, S., & Graesser, A. (2012). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(4).
- Education, S. (2016). *Lessons learned from early implementations of adaptive courseware*.
- Helmreich, J. E., & Pruzek, R. M. (2009). PSAGraphics: An R package to support propensity score analysis. *Journal of Statistical Software*, 29(6), 1–23. Retrieved from <http://www.jstatsoft.org/v29/i06/>
- Homer, B. D., & Plass, J. L. (2010). Expertise reversal for iconic representations in

- science visualizations. *Instructional Science*, 38(3), 259–276.
- Johnson, S. D., Aragon, S. R., Shaik, N., & Palma-Rivas, N. (2000). Comparative analysis of learner satisfaction and learning outcomes in online and face-to-face learning environments. *Journal of Interactive Learning Research*, 11(1).
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, 19(4), 509–539.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31.
- Lim, L., Lim, S. H., & Lim, W. Y. R. (2023). Efficacy of an adaptive learning system on course scores. *Systems*, 11(31).
- Maki, R. H., Maki, W. S., Patterson, M., & Whittaker, P. D. (2000). Evaluation of a web-based introductory psychology course: I. Learning and satisfaction in on-line versus lecture courses. *Behavior Research Methods, Instruments, & Computers*, 32(2), 230–239.
- Martin, F., Chen, Y., Moore, R. L., & Westine, C. D. (2020). Systematic review of adaptive learning research designs, context, strategies, and technologies from 2009 to 2018. *Education Tech Research Dev*, 68.
- Means, B., Bakia, M., & Murphy, R. (2014). *Learning online: What research tells us about whether, when and how*. New York: Routledge.
- Means, Barbara, Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2009). Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies. *US Department of Education*.
- Muñoz, J. L. R., Ojeda, F. M., Jurado, D. L. A., Peña, P. F. P., Carranza, C. P. M., Berrios, H. Q., . . . Vasquez-Pauca, M. M. (2022). Systematic review of adaptive learning technology for learning in higher education. *Eurasian Journal of Educational Research*, 98.
- Neuhauser, C. (2002). Learning style and effectiveness of online and face-to-face

- instruction. *The American Journal of Distance Education*, 16(2), 99–113.
- Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles concepts and evidence. *Psychological Science in the Public Interest*, 9(3), 105–119.
- R Core Team. (2023). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Raudenbush, S. W., Hong, G., & Rowan, B. (2003). *Studying the causal effects of instruction with application to primary-school mathematics*. Research Seminar II: Instructional; performance consequences of high poverty schooling.
- Rosenbaum, Paul R. (2002). *Observational studies*. Spring.
- Rosenbaum, P. R. (2012). Testing one hypothesis twice in observational studies. *Biometrika*, 99(4), 762–774.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Sekhon, J. S. (2011). Multivariate and propensity score matching software with automated balance optimization: The Matching package for R. *Journal of Statistical Software*, 42(7), 1–52. <https://doi.org/10.18637/jss.v042.i07>
- Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association*, 103(484), 1334–1343.
- Skinner, B. F. (1958). Teaching machines. *Science*, 128(3330), 969–977.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25, 1–21.
- Stuart, E. A., & Rubin, D. B. (2008). *Best practices in quantitative methods* (J. Osborne, Ed.). Sage Publications.
- van Buuren, S. (2012). *Flexible imputation of missing data*. CRC Press.

- van Buuren, Stef, & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3), 1–67.
<https://doi.org/10.18637/jss.v045.i03>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wise, B. (2010). The online learning imperative: A solution to three looming crises in education. *Education Digest: Essential Readings Condensed for Quick Review*, 76(3), 52–58.

Appendix A

Appendix A: Descriptive Statistics

Descriptive Statistics for Mathematics

Table A1

Descriptive Statistics for Mathematics

Variable	**Control**	**Treatment**
__Gender__		
Male	535 (59%)	195 (48%)
Female	375 (41%)	213 (52%)
(Missing)	4	3
__Ethnicity__		
Black	180 (20%)	71 (17%)
Hispanic	117 (13%)	47 (11%)
Other	100 (11%)	29 (7.1%)
White	517 (57%)	264 (64%)
__Military__	450 (49%)	157 (38%)
__DegreeLevel__		
Associate	611 (67%)	287 (70%)
Bachelor	238 (26%)	104 (25%)
Master	3 (0.3%)	2 (0.5%)
None	62 (6.8%)	18 (4.4%)
__Income__		
01-LT25K	75 (8.8%)	33 (8.6%)
02-LT35K	125 (15%)	51 (13%)
03-LT45K	131 (15%)	53 (14%)
04-LT55K	121 (14%)	60 (16%)
05-LT70K	137 (16%)	57 (15%)
06-LT85K	132 (16%)	53 (14%)
07-LT100K	54 (6.4%)	37 (9.7%)
08-LT120K	46 (5.4%)	21 (5.5%)
09-GE120K	29 (3.4%)	17 (4.5%)
(Missing)	64	29
__Pell__	182 (20%)	84 (20%)
__Employment__		
Unemployed	108 (13%)	40 (11%)
Part-Time	63 (7.7%)	33 (8.8%)
Full-Time	646 (79%)	301 (80%)
(Missing)	97	37
__ESL__	833 (94%)	373 (94%)
(Missing)	28	14
__FirstGeneration__	314 (34%)	148 (36%)
__Age__	36 (29, 42)	37 (31, 44)
__GPA__	3.00 (2.50, 4.00)	3.50 (2.75, 4.00)
(Missing)	201	95
__TransferCredits__	27 (3, 48)	24 (2, 48)
__EarnedCredits__	6 (3, 12)	7 (3, 13)

Descriptive Statistics for Biology A

Table A2

Descriptive Statistics for Biology 1

Variable	**Control**	**Treatment**
__Gender__		
Male	91 (75%)	41 (66%)
Female	31 (25%)	21 (34%)
(Missing)	7	2
__Ethnicity__		
Black	28 (22%)	15 (23%)
Hispanic	12 (9.3%)	6 (9.4%)
Other	20 (16%)	4 (6.3%)
White	69 (53%)	39 (61%)
__Military__	82 (64%)	36 (56%)
(Missing)	1	0
__DegreeLevel__		
Associate	47 (36%)	18 (28%)
Bachelor	62 (48%)	37 (58%)
Master	4 (3.1%)	1 (1.6%)
None	16 (12%)	8 (13%)
__Income__		
01-LT25K	7 (6.3%)	3 (5.7%)
02-LT35K	17 (15%)	6 (11%)
03-LT45K	15 (13%)	12 (23%)
04-LT55K	17 (15%)	8 (15%)
05-LT70K	26 (23%)	7 (13%)
06-LT85K	14 (13%)	10 (19%)
07-LT100K	6 (5.4%)	4 (7.5%)
08-LT120K	6 (5.4%)	3 (5.7%)
09-GE120K	4 (3.6%)	0 (0%)
(Missing)	17	11
__Pell__	24 (19%)	13 (20%)
__Employment__		
Unemployed	13 (13%)	4 (7.4%)
Part-Time	4 (3.8%)	0 (0%)
Full-Time	87 (84%)	50 (93%)
(Missing)	25	10
__ESL__	114 (95%)	55 (95%)
(Missing)	9	6
__FirstGeneration__	40 (31%)	16 (25%)
__Age__	33 (28, 40)	33 (28, 41)
(Missing)	1	0
__GPA__	3.07 (2.40, 3.67)	3.33 (2.64, 3.90)
(Missing)	27	16
__TransferCredits__	47 (24, 73)	49 (11, 77)
__EarnedCredits__	7 (3, 14)	9 (3, 13)

Descriptive Statistics for Biology B

Table A3

Descriptive Statistics for Biology 2

Variable	**Control**	**Treatment**
__Gender__		
Male	48 (77%)	93 (62%)
Female	14 (23%)	58 (38%)
(Missing)	3	4
__Ethnicity__		
Black	21 (32%)	32 (21%)
Hispanic	7 (11%)	17 (11%)
Other	9 (14%)	15 (9.7%)
White	28 (43%)	91 (59%)
__Military__	44 (69%)	83 (54%)
(Missing)	1	0
__DegreeLevel__		
Associate	25 (38%)	55 (35%)
Bachelor	31 (48%)	82 (53%)
Master	1 (1.5%)	6 (3.9%)
None	8 (12%)	12 (7.7%)
__Income__		
01-LT25K	7 (12%)	15 (11%)
02-LT35K	7 (12%)	23 (17%)
03-LT45K	10 (17%)	16 (12%)
04-LT55K	9 (15%)	16 (12%)
05-LT70K	7 (12%)	27 (20%)
06-LT85K	8 (14%)	14 (11%)
07-LT100K	5 (8.5%)	9 (6.8%)
08-LT120K	3 (5.1%)	10 (7.6%)
09-GE120K	3 (5.1%)	2 (1.5%)
(Missing)	6	23
__Pell__	14 (22%)	39 (25%)
__Employment__		
Unemployed	7 (14%)	31 (24%)
Part-Time	0 (0%)	6 (4.7%)
Full-Time	43 (86%)	90 (71%)
(Missing)	15	28
__ESL__	55 (90%)	138 (95%)
(Missing)	4	10
__FirstGeneration__	26 (40%)	44 (28%)
__Age__	35 (29, 41)	34 (29, 41)
(Missing)	1	0
__GPA__	3.00 (2.62, 3.64)	3.25 (2.70, 3.75)
(Missing)	14	38
__TransferCredits__	47 (19, 77)	44 (13, 63)
__EarnedCredits__	9 (3, 16)	7 (0, 13)

Appendix B

Appendix B: Propensity Score Balance Statistics

The following figures summarize the balance achieved from propensity score analysis.

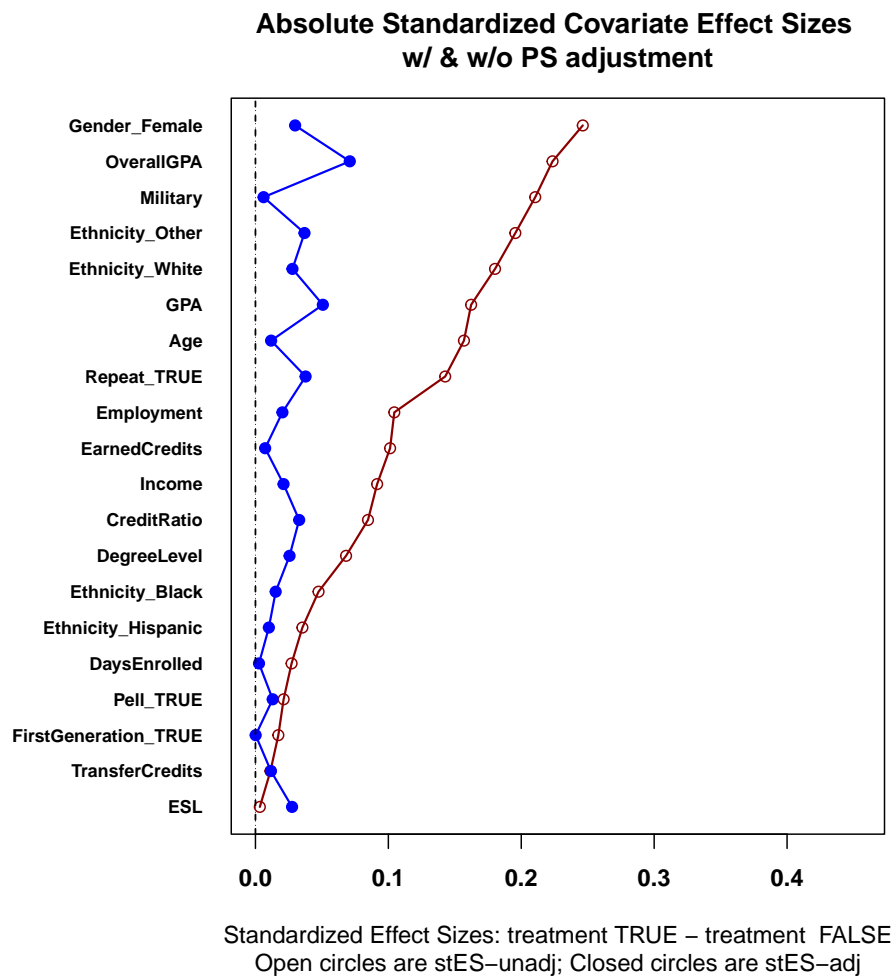


Figure B1. Absolute Standardized Covariate Effect Size With and Without Propensity Score Adjustment: Mathematics

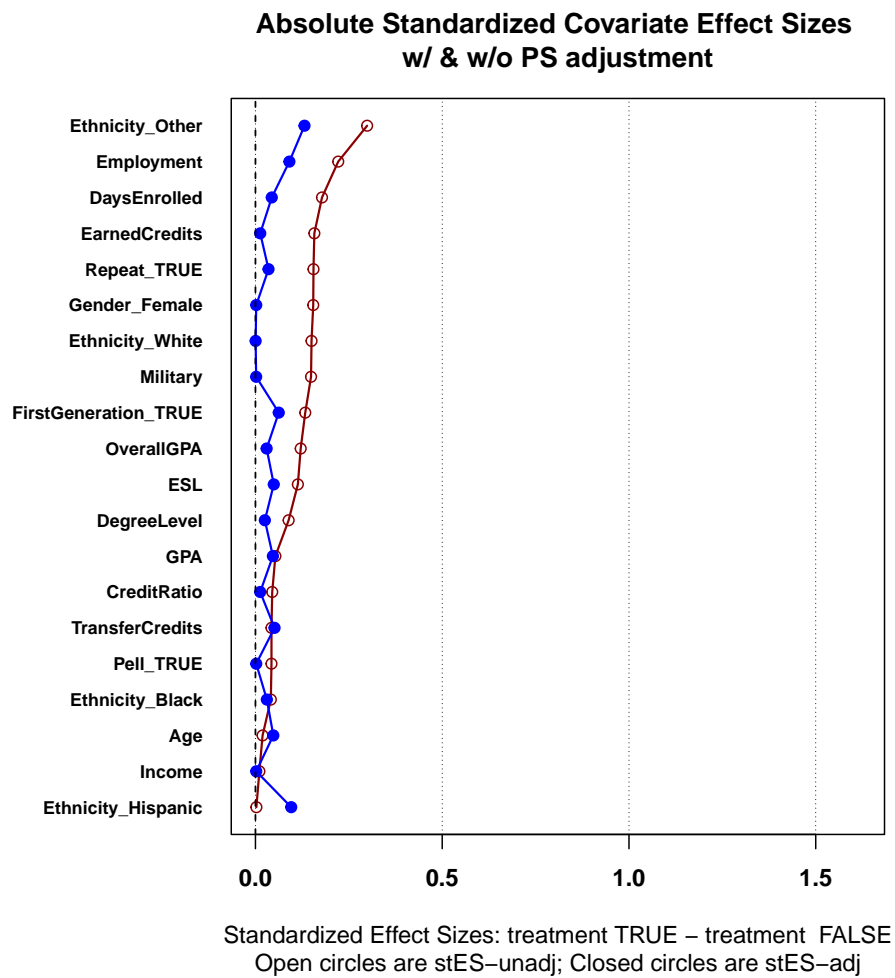


Figure B2. Absolute Standardized Covariate Effect Size With and Without Propensity Score Adjustment: Biology A

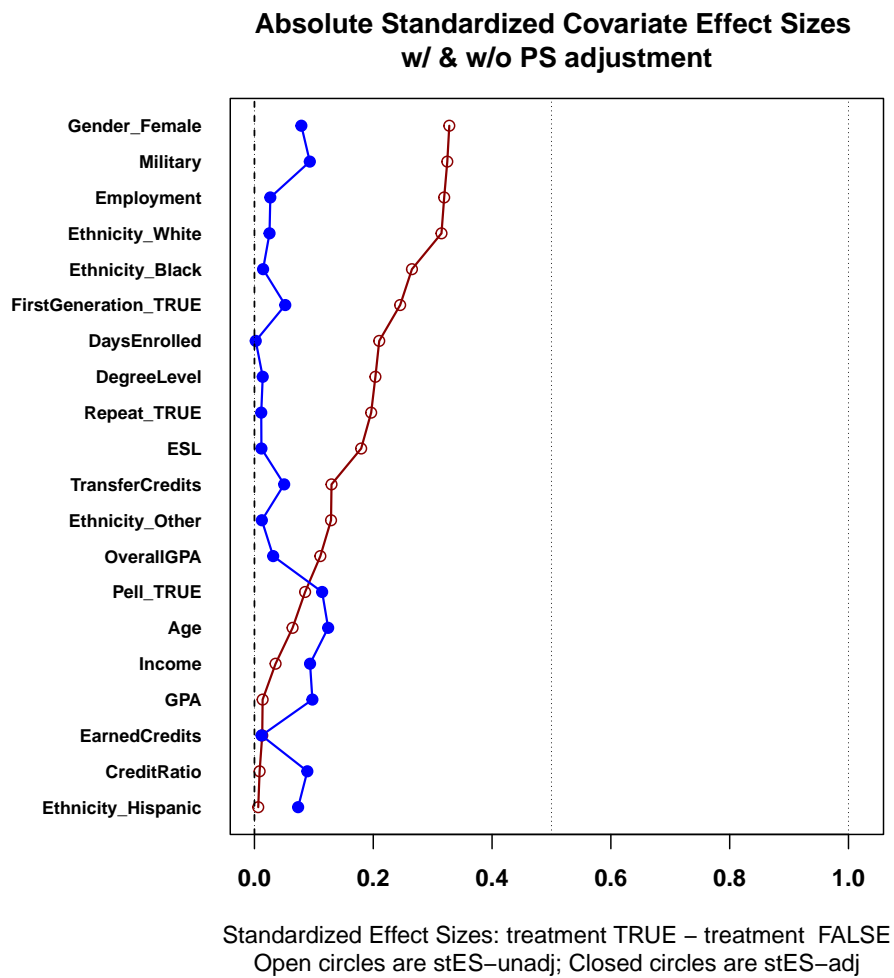


Figure B3. Absolute Standardized Covariate Effect Size With and Without Propensity Score Adjustment: Biology B

Appendix C

Appendix B: Logistic Regression Results

Characteristic	**log(OR)**	**p-value**	**log(OR)**	**p-value**	**log(OR)**	**p-value**
__Gender__	0.46	__0.003__	0.24	0.5	0.91	__0.000__
__Ethnicity__						
__White__	—		—		—	
__Black__	-0.12	0.5	0.18	0.7	-1.0	__0.000__
__Hispanic__	-0.09	0.7	0.05	>0.9	-0.32	0.000
__Other__	-0.55	__0.019__	-0.66	0.3	-1.1	0.000
__Military__	-0.26	0.11	-0.51	0.2	-0.19	0.000
__DegreeLevel__	0.12	0.5	0.09	0.8	0.42	0.000
__GPA__	0.22	0.072	0.30	0.3	-0.20	0.000
__OverallGPA__	0.20	0.073	0.10	0.8	0.26	0.000
__TransferCredits__	0.00	0.8	0.00	0.6	-0.01	0.000
__CreditRatio__	-0.78	__0.035__	-0.76	0.4	0.64	0.000
__EarnedCredits__	0.00	0.8	-0.02	0.3	-0.02	0.000
__Income__	0.00	0.9	-0.08	0.4	0.00	>0.9
__Age__	0.01	0.15	0.00	>0.9	0.03	0.000
__DaysEnrolled__	0.00	>0.9	0.00	0.4	0.00	0.000
__FirstGeneration__	-0.01	>0.9	-0.21	0.6	-0.46	0.000
__Employment__	0.16	0.085	0.41	0.2	-0.47	0.000
__ESL__	-0.04	0.9	0.55	0.5	0.55	0.000
__Pell__	-0.07	0.6	0.02	>0.9	0.08	0.000
__Repeat__	-0.58	0.11	-0.35	0.5	-0.68	0.000