(2016) In T. Fukawa-Connelly, N. Infante, K. Keene, and M. Zandieh (Eds.), Proceedings of the 19th Annual Conference on Research in Undergraduate Mathematics Education, Pittsburgh, Pennsylvania.

Exploration of the Factors that Support Learning: Web-based Activity and Testing Systems in Community College Algebra

Shandy Hauk
WestEd

Bryan Matlen
WestEd

A variety of computerized interactive learning platforms exist. Most include instructional supports in the form of problem sets. Feedback to users ranges from a single word like "Correct!" to offers of hints and partially to fully worked examples. Behind-the-scenes design of such systems varies as well – from static dictionaries of problems to "intelligent" and responsive programming that adapts assignments to users' demonstrated skills, timing, and an array of other learning theory-informed data collection within the computerized environment. This short paper presents background on digital learning contexts and describes the lively conversation with attendees at the conference session. The topics were the research design and early results of a cluster-randomized controlled trial study in community college elementary algebra classes where the intervention was a particular type of web-based activity and testing system.

Key words: Adaptive Tutoring System, College Algebra, Multi-site Cluster Randomized Controlled Trial

Research Questions

Funded by the U.S. Department of Education, we are conducting a large-scale mixed methods study in over 30 community colleges. The study is driven by two research questions:

Research Question 1: What student, instructor, or community college factors are associated with more effective learning from the implemented digital learning platform?

Research Question 2: What challenges to use-as-intended (by developers) are faculty encountering and how are they responding to the challenges as they implement the learning tool?

Background and Conceptual Framing

First, there are distinctions among cognitive, dynamic, and static learning environments (see Table 1). Learning environments can vary along at least two dimensions: (1) the extent to which they adaptively respond to student behavior and (2) the extent to which they are based on a careful cognitive model.

Table 1. Conceptual framework of the types of instruction based on adaptability and their basis in a theory of learning.

		Static	Dynamic
Is a particular model of learning explicit in design and implementation (structure and processes)?	No	Text and tasks with instructional adaptation external to the materials	Adaptive tutoring systems (Khan Academy, ALEKS, ActiveMath)
	Yes	Textbook design and use driven by fidelity to an explicit theory of learning	"Intelligent" tutoring systems (Cognitive Tutor)

Static learning environments are those that are non-adaptive without reliance on an underlying cognitive model – they deliver content in a fixed order and contain scaffolds or

feedback that are identical for all users. The design may be based on intuition, convenience, or aesthetic appeal. An example of this type of environment might be online problem sets from a textbook that give immediate feedback to students (e.g., "Correct" or "Incorrect").

Dynamic learning environments keep track of student behavior (e.g., errors, error rates, or time-on-problem) and use this information in a programmed decision tree that selects problem sets and/or feedback based on students' estimated mastery of specific skills. An example of a dynamic environment might be a system such as ALEKS or the "mastery challenge" approach now used at the online Khan Academy. For example, at khanacadmy.org a behind-the-scenes data analyzer captures student performance on a "mastery challenge" set of items. Once a student gets six items in a row correct, the next level set of items in a programmed target learning trajectory is offered. Depending on the number and type of items the particular user answers incorrectly (e.g., on the path to six items in a row done correctly), the analyzer program identifies target content and assembles the next "mastery challenge" set of items. Above and beyond such responsive assignment generation, programming in a "cognitively-based" dynamic environment is informed by a theoretical model that asserts the cognitive processing necessary for acquiring skills (Anderson et al. 1995; Koedinger & Corbett, 2006). For example, instead of specifying only that graphing is important and should be practiced, a cognitively-based environment also will specify the student thinking and skills needed to comprehend graphing (e.g., connecting spatial and verbal information), and provide feedback and scaffolds that support these cognitive processes (e.g., visuo-spatial feedback and graphics that are integrated with text). In cognitively-based environments, scaffolds themselves can also be adaptive (e.g., more scaffolding through examples can be provided early in learning and scaffolding can be faded as a student acquires expertise; Ritter et al., 2007). Like other dynamic systems, cognitively-based systems can also provide summaries of student progress, which better enable teachers to support struggling students. Some studies have shown the promise of cognitively-based dynamic environments in post-secondary mathematics (Koedinger & Sueker, 1996).

Method

The study we report here is a multi-site cluster randomized trial (note: because the study is currently underway, we purposefully under-report some details). Half of instructors at each community college site are assigned to use a particular adaptive web-based system in their instruction (Treatment condition), the other half teach as they usually would (Control condition). The primary outcome measure for students' performance is an assessment from the Mathematics Diagnostic Testing Program (MDTP), which is a valid and reliable assessment of students' algebraic knowledge (Gerachis & Manaster, 1995).

Using a stratified sampling approach to recruitment, we first conducted a cluster analysis on all 112 community college sites eligible to participate in the study (i.e., in a state that was a study partner and offering semester-long courses in elementary algebra that met at least some of the time in a physical classroom or learning/computer lab). The cluster analysis was based on college-level characteristics that may be related to student learning (e.g., average age of students at the college, the proportion of adjunct faculty, etc.). This analysis led to five clusters of colleges. Our recruitment efforts then aimed to include a proportionate number of colleges within each cluster. The primary value of this approach is that it allows more appropriate generalization of study findings to the target population (Tipton, 2014). Recruitment for our first cohort of participants yielded a study sample of 38 colleges similar to the overall distribution across clusters that was the target for the sample (see Figure 1).

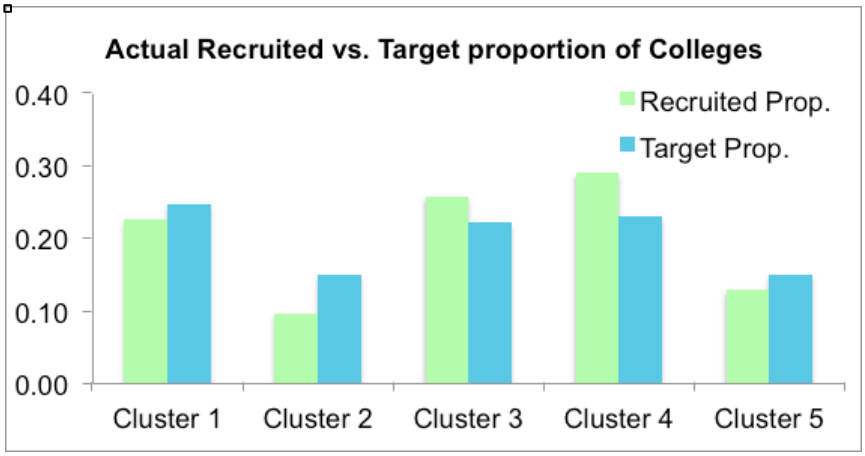

Figure 1. Recruited sample proportions and target sample proportions across clusters.

Sample for this Report

Initial enrollment in the study included 89 teachers across the 38 college sites. For this report on early results, we used the data from the participating students of 30 instructors across 19 colleges. Student and teacher numbers related to the data set reported on here are shown in Table 2.

Table 2. Counts of Teachers, Students, and Colleges in the Study.

Condition	*Teachers*	*Students*	*Colleges*
Control	19	147	15
Treatment	11	80	10
Total	*30*	*227*	*19*

Quantitative Analysis

The primary aim of the quantitative analysis was to address Research Question 1, how and for whom the particular adaptive computer environment might be effective. To this end, ultimately we will employ Hierarchical Linear Modeling (HLM) on the full data set. Models will include interaction terms between instructors' treatment assignment and covariates at different levels (e.g., students' history of course-taking, self-concept of ability), to explore the moderating impact of tool use on student learning. The primary post-test outcome measure is the MDTP elementary algebra assessment. A different but related MDTP pre-algebra diagnostic served as the measure of students' baseline knowledge. For this report, we have focused on the MDTP post-test as an indicator of algebraic knowledge.

Qualitative Analysis

To address Research Question 2, a great deal of textual, observational, and interview data are still being gathered. These data allow careful analysis of the intended and actual use of the learning environment and the classroom contexts in which it is enacted – an examination of implementation structures and processes. Indices of specific and generic fidelity derived from this work also will play a role in HLM generation and interpretation in the coming year.

Preliminary Results

Fall 2015 was the first full semester of data gathering for the project. It was our "practice" semester in that researchers were refining instruments and participant communication processes while treatment condition instructors were trying out the web-based learning tool with their classes for the first time. The "efficacy study" semester takes place in Spring 2016.

At the Conference: Poster Conversations

At the time of the conference, we had early results from the practice semester that suggested an aptitude by treatment interaction. Specifically, students in the Treatment group who started out with lower scores relative to the group mean on their algebra readiness pre-test, showed more benefit than Control group students (i.e., Treatment group students from the lower scores group had higher scores, relative to the group mean, on their post-test in elementary algebra). Some discussions in the poster session at the conference revolved around this interaction. For instance, one conference participant reported finding a similar result using a web-based technology: In his study, lower ability students exhibited higher grades when they were required to use the web-based tutor than when they were not. In another discussion, a conference participant hypothesized that instructors need to gain familiarity with technology before they can effectively use web-based learning tools for teaching. Indeed, after a semester of practice, Treatment but not Control instructors in our study reported an increase in their ability to use technology for teaching mathematics. Though not statistically significant ($p = .12$), the difference was consistent with the conference participants' hypothesis. Another key set of conversations at the poster were about the idea of an adaptive system that was based on a relatively stable "learning trajectory" or "genetic decomposition" as compared to a "cognitively-based" model approach that includes variability within a trajectory or decomposition, depending on the student, as the mechanism to guide selection algorithms when diagnosing and responding to student work in the computerized learning environment. We believe interactions such as those at the poster help to improve communication between the cognitive science research community and the RUME community.

Since the Conference: Updated Results

Since the conference, we have cleaned more data and have conducted analyses on this updated set. These analyses indicated that the aptitude by treatment interaction that was reported on the poster was no longer statistically significant: Estimate = -0.04, $p = 0.71$. Nevertheless, findings may continue to change as we continue to collect data in our efficacy semester.

Here we can add information about a new analysis of post-semester test scores that corrected for instructor clustering and students' scores on their algebra readiness pre-test. This analysis indicated that students in the treatment condition (adjusted $M = 23.80$, unadjusted $SD = 6.67$, $N = 80$) performed higher on their post-test than students in the control condition (adjusted $M = 22.45$, unadjusted $SD = 8.27$, $N = 147$), albeit these mean scores, at about 1 point difference, were not statistically different (Estimate = 0.93, $p = 0.62$). The effect size for this difference was Hedges' $g = .12$, which is considered small, but within expectation for efficacy trials of this type and is worth noting (Cheung & Slavin, 2015; Hill et al. 2008). As mentioned, this analysis included only a subset of students (data cleaning is ongoing) and results may continue to change as we collect, clean, and add more data to the analysis. Figure 2 shows box-plots of pre-test and adjusted post-test scores.

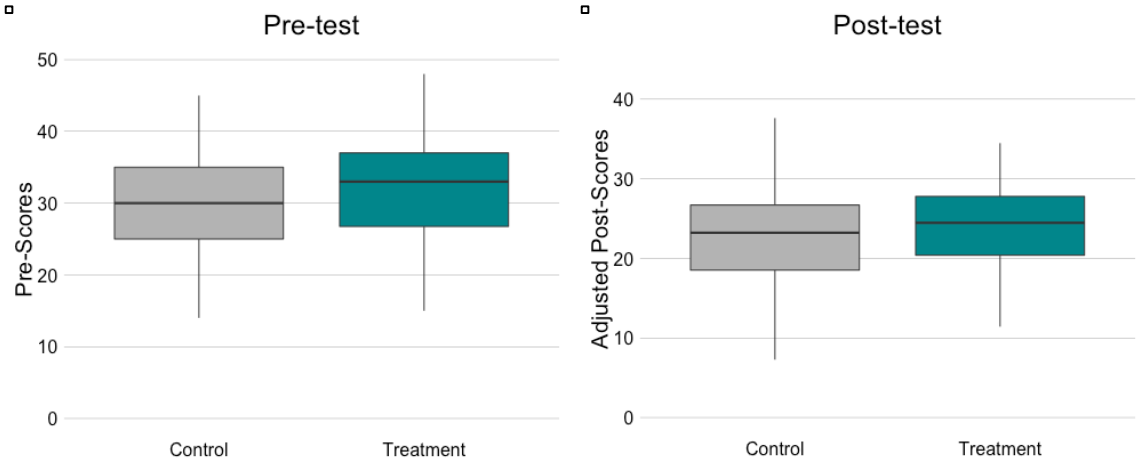

Figure 2. Left: Box-plot of pre-test scores. Right: Box-plot of adjusted post-test scores.

Next Steps

We will continue this study with a second cohort of new participants who will repeat the year-long study in the 2016-2017 academic year. Our specific objectives in the upcoming year are to (1) complete data collection from the first cohort for the primary efficacy study (i.e., data for hundreds of students for Spring 2016), (2) continue reporting findings from the Spring 2016 efficacy study of cohort 1, (3) recruit a second cohort of participants for another practice semester and efficacy study in 2016-17, and (3) begin the practice semester of the study with second cohort of participants.

Of particular interest is how the spread of information shown in Figure 3 might look for the efficacy (Spring 2016) data set. We look forward to having more to report and new questions to discuss at the 2017 conference.

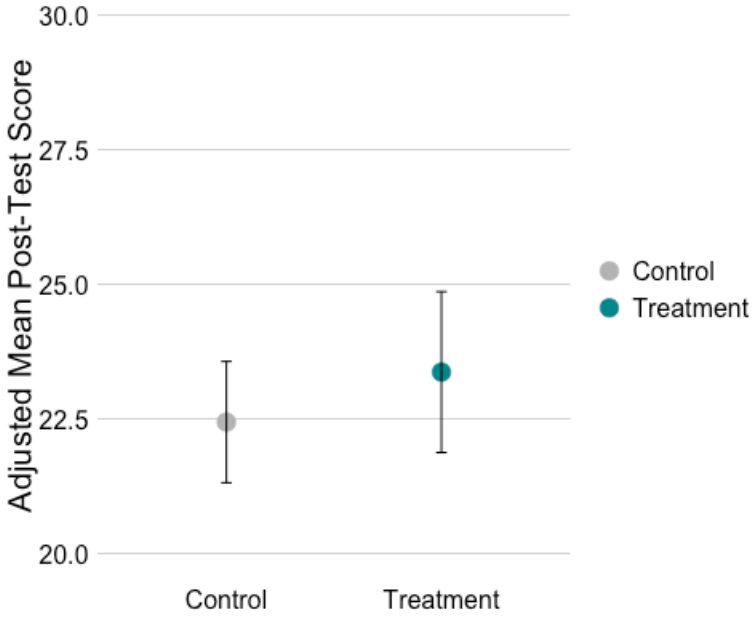

Figure 3. Adjusted mean post-test score by condition. Vertical bars represent standard errors of the means.

Acknowledgement

This project is supported by a grant from the U.S. Department of Education, Institute of Education Sciences (IESR305A140340). Any opinions, findings, conclusions or recommendations are those of the authors and do not necessarily reflect the views of the Federal Government.

References

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*(2), 167-207.

Cheung, A., & Slavin, R.E. (2015, September). How methodological features affect effect sizes in education. Baltimore, MD: Johns Hopkins University, Center for Research and Reform in Education. Source: http://www.bestevidence.org/methods/methods.html

Gerachis, C., & Manaster, A. (1995). *User manual*. Mathematics Diagnostic Testing Project, California State University/University of California. Retrieved from http://mdtp.ucsd.edu/approvalstatus.shtml

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2(3),* 172- 177.

Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences*. Cambridge, MA: Cambridge University Press.

Koedinger, K.R., & Sueker, E.L.F. (1996). PAT goes to college: Evaluating a cognitive tutor for developmental mathematics. In *Proceedings of the Second International Conference on the Learning Sciences* (pp. 180–87). Charlottesville, VA: Association for the Advancement of Computing in Education.

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, *14*(2), 249-255.

Tipton, E. (2014). Stratified sampling using cluster analysis: A sample selection strategy for improved generalization from experiments. *Evaluation Review*, *37*(2), 109-139.