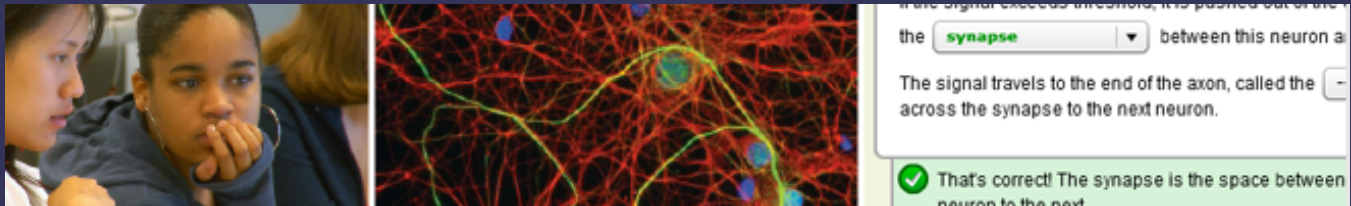


Open Learning Initiative Courses in Community Colleges: Evidence on Use and Effectiveness



Julia Kaufman, Sarah Ryan, Candace Thille & Norman Bier¹
Open Learning Initiative
Carnegie Mellon University

October 2013

¹ Julia Kaufman and Sarah Ryan are research scientists at the Open Learning Initiative (OLI). Candace Thille is the founding director of OLI, and Norman Bier is the current director of OLI at Carnegie Mellon. John Rinderle from OLI also provided much helpful assistance and feedback for this report. Additionally, Brian Junker and Howard Seltman provided invaluable statistical advice. The authors would also like to thank the many faculty and students who participated in this evaluation, as well as the funders - Bill & Melinda Gates Foundation, the Lumina Foundation, the William and Flora Hewlett Foundation and the Walter S. Johnson Foundation - who made this work possible.

Introduction

Despite high costs for a college education and increasing debts that students face after they graduate, the demand for effective post-secondary higher education continues to grow. In 2013 alone, college enrollment grew by 15.4% across the U.S.² The increase in the number of students seeking post-secondary education has brought with it a corresponding increase in learner variability: a larger population of students with broader experiences, differing perceptions and diverse achievement levels.³

In the face of dramatic challenges in access and attainment, policy makers, foundations and administrators have looked to technology as a mechanism for controlling costs and increasing student success. There has been a dramatic increase in the use of online courses over the past decade,⁴ and an unprecedented interest in Massive Open Online Courses (MOOCs) in the past year. In the majority of cases, online courses have been executed in a manner that attempts to replicate the experience of a traditional classroom, in effect leveraging the technology as a mechanism for scaling lectures to a broader population of students. Although the MOOC approach has generated great excitement, outcomes have thus far not lived up to expectations.⁵

A different approach to leveraging technology in support of learning can be found in the work of the Open Learning Initiative (OLI). OLI is an open education research project that began in 2002 at Carnegie Mellon University with a grant from The William and Flora Hewlett Foundation. Unlike many open education projects, OLI is not a collection of materials created by individual faculty to support traditional instruction. The original goal of the project was to develop web-based learning environments that support individual learners, who do not have the benefit of an instructor, to achieve the same learning outcomes as students who complete traditional courses.

To meet the goal of supporting independent learners, OLI built on Carnegie Mellon's institutional strengths in cognitive science, software engineering, and human-computer interaction. OLI courses are developed by teams composed of disciplinary faculty, learning scientists, human-computer interaction experts, and software engineers in order to make the best use of multidisciplinary knowledge for

² State Higher Education Executive Officers, State Higher Education Finance, FY 2012.

³ The Chronicle of Higher Education. "Who Are the Undergraduates?", December 12, 2010

⁴ November 2011 report by the Sloan Consortium and the Babson Survey Research Group.

⁵ Ry Rivard. "Udacity Project on 'Pause'", Inside Higher Education, July 18, 2013.

designing effective learning environments. The OLI design team articulates an initial set of student-centered, observable learning objectives and designs the instructional environment to support students in achieving those outcomes.

Although originally designed to support individual learners, OLI courses are increasingly used by instructors inside and outside of Carnegie Mellon as a complement to their instructor-led courses.

OLI learning environments have been used in numerous studies to better understand effective approaches to enhancing learning with technology and to gain an understanding of how the courses support learners and educators. Studies at Carnegie Mellon have shown that students using the OLI statistics course were able to achieve the same or better learning outcomes in half the time of peers in the traditional course.⁶ A subsequent independent trial of the Statistics course was conducted by ITHAKA S+R across six public institutions. Using a randomized approach with statistically reliable control and treatment groups, ITHAKA S+R reported that students using OLI had equivalent learning outcomes and completed the course 25% more quickly.⁷

The successes that prior studies have demonstrated in accelerating learning in four-year colleges raises the question of how the OLI approach could be successfully used to support community colleges in addressing the challenges currently facing higher education. In the face of shrinking state budgets, community colleges are being asked to increase attainment rates and reduce the cost of instruction. They are pinched for money, squeezed for space and find themselves under unprecedented pressure to see to it that the growing number of students who matriculate on their campuses each year complete their degrees. Yet, over half (54%) of all community college credits are non-productive (i.e., courses in which students fail, withdraw or that do not contribute to program completion).⁸

To explore how the OLI approach might be used and adapted to address some of the challenges confronting community colleges, OLI launched the Community College Open Learning Initiative (CC-OLI) in 2009. Jointly funded by the Bill & Melinda

⁶ Lovett, M., Meyer, O., & Thille, C. (2008). *The Open Learning Initiative: Measuring the effectiveness of the OLI statistics course in accelerating student learning*. Journal of Interactive Media in Education.

⁷ Bowen, W.G., Chingos, M.M., Lack, K.L., & Nygren, T.I. (2012). *Interactive Learning Online at Public Universities: Evidence from Randomized Trials*. ITHAKA.

⁸ “*The Institutional Costs of Student Attrition*” by Delta Cost Project, 2012, and “*Winning by Degrees*” by McKinsey & Company, 2010.

Gates Foundation, the William and Flora Hewlett Foundation, the Lumina Foundation and the Walter S. Johnson Foundation, CC-OLI is an attempt to apply the lessons learned and success from OLI's earlier work to the unique challenges faced by 2-year institutions. CC-OLI focused on adapting or creating learning environments to support gateway courses (i.e., high-enrollment courses with low student success rates that are critical to many degree paths and that are often an obstacle to degree completion). CC-OLI's goals were to include community college faculty in the development, use and evaluation of these courses and to scale to at least 40 community colleges and 100 classrooms, all with an overarching goal of demonstrating a measurable increase in student success.

As of spring of 2013, the CC-OLI project has created community college-targeted courses in Statistics, Introduction to Biology, Anatomy & Physiology and Introduction to Psychology, along with pilot modules in English Composition. Participation in these development efforts has been widespread, with 51 faculties from 48 institutions participating in development. Thus far, 39,000 students in classes offered at 142 academic institutions, have used these courses.

While both the accelerated learning studies and the ITHAKA S+R studies reported similar or improved learning outcomes associated with the introduction of OLI materials into the classroom, neither study investigated differences in outcomes between groups of students who used the OLI materials minimally and those who made more extensive use of the learning resources. Similarly, the earlier studies did not explore faculty use of the courses. We know that the way faculty use the learning assets and the instructor tools available within OLI varies greatly. Gaining an understanding of how use of the OLI resources is associated with students' learning outcomes would have real value, both for refining the learning environment design and supporting faculty in incorporating OLI resources into their practice.

The current report summarizes the findings from a study of the use of the CC-OLI courses by faculty teaching at 24 community colleges across the United States. The report particularly delves into the relationship between faculty/student use of OLI resources and students' learning outcomes. Beyond the specific findings with regard to CC-OLI, the study also reports on some of the challenges in executing large-scale evaluation studies in higher education.

Goals of the Study

We had three main goals for the current study that had implications for how we designed our study and analyzed our data. We outline those goals alongside our rationale for each goal and the research methods we used to gather and analyze data addressing those goals.

Goals	Rationales	Methods
To understand the differences among faculty and students who use OLI materials versus those who choose not to use OLI materials	To better understand the population served by OLI so that they can be served better by online resources	Descriptive analyses of: <ul style="list-style-type: none"> • Faculty and student surveys • Baseline data from pre-tests
To compare achievement gains for students who use OLI versus those who do not	To investigate whether students of faculty who agree to use OLI demonstrate achievement gains over and above those who do not use OLI	Quasi-experimental propensity score matching to compare OLI and non-OLI students: <ul style="list-style-type: none"> • Use of student survey and achievement data to estimate propensity scores • Use of propensity scores to match similar OLI and non-OLI students • Comparisons of achievement gains among matched students
To understand implementation differences among faculty and students who opt to use OLI and the associations between implementation and achievement gains	<p>To learn about the full variation of faculty implementation choices and those choices most associated with achievement gain in order to support faculty and to make recommendations regarding best OLI practices</p> <p>To consider how large differences in faculty and student use of OLI matter for students' achievement gains</p>	Hierarchical linear models and qualitative analyses to understand the relationship between OLI use and achievement: <ul style="list-style-type: none"> • Investigation of relationships between faculty and student use of OLI, as gauged by surveys and OLI system data • Investigation of main survey and OLI system variables associated with students' OLI achievement gains • Qualitative comparisons of OLI use among faculty with highest and lowest student achievement gains

We first describe our study methods in more detail below, including a description of the Open Learning Initiative courses and the components of those courses we studied. We then provide main findings in relation to the goals listed above. Those main findings include:

- Key differences in demographics, academic performance, and preferences defining faculty and students who use OLI compared to those who do not
- Positive but non-significant effect of OLI courses on achievement gains for a subset of students of faculty who agree to use OLI with one of their class sections but not the other
- Huge variations in use of OLI courses among faculty and students
- Large achievement gains among faculty and students who engage in “optimal use” of OLI resources as they have been measured for this study

Data and Methods

About Open Learning Initiative Online Courses

The Open Learning Initiative (OLI) draws on the expertise of college faculty from particular content areas alongside software developers to create high-quality online resources. The courses on which we focused for this evaluation were introductory “gateway” courses critical for graduation and student success: Anatomy & Physiology, Biology, Psychology and Statistics.

All OLI courses are designed to engage students to put new knowledge into practice and support students to assess their own progress. Activities and assessments embedded within the courses help to ensure that students understand concepts rather than simply memorize facts. Embedded learning activities include “Learn by Doing” (LBD) and “Did I Get This” (DIGT) activities. LBD activities are meant to help students practice new skills or develop understanding of a new concept. While engaged in a LBD activity, students receive automated support and feedback. DIGT activities are meant to help students assess whether they have mastered a skill or understood a specific concept. Upon completing a DIGT self-assessment, a student receives context specific and immediate feedback. In addition to the low-stakes LBD and DIGT activities, the OLI courses include regular high stakes short quizzes and longer exams.

OLI courses provide tools for educators to easily assess what students are learning and what concepts and skills students have not yet mastered. Faculty have access to the Learning Dashboard, which includes information about student participation and performance both at the class and individual student level. Instructors can use the Dashboard to view the predicted knowledge state of the class or an individual student on each learning objective. Faculty can also take advantage of the OLI course Grade book to view student performance on graded assessments.

Instructors can use the OLI course builder to customize a course by selecting and sequencing the OLI materials and controlling access and due dates. In addition to customizing their OLI course, faculty in the study were also free to decide whether or not to require students to use any or all of the OLI material and whether or not use of OLI would count toward the students' grades.

Study Recruitment and Participation

With the assistance of Lumen Learning, the evaluation team sent email to over 1800 faculty and administrators in almost 1000 community colleges inviting them to participate in the Fall 2012 pilot study and/or Spring 2013 community college evaluation study. Faculty who teach introductory courses in Anatomy & Physiology (A&P), Biology, Psychology and Statistics were invited to participate either by 1) using the CC-OLI resources to teach their courses or 2) teaching their courses as they normally would (i.e. without using OLI). Both OLI and non-OLI participants were asked to administer pre- and post-assessments to their students, complete a faculty survey, and ask their students to complete a student survey. All faculty who agreed to participate were offered a monetary incentive. About 20% of contacted faculty responded to the initial email and indicated interest in learning more about the study and/or committed to participate. Faculty who expressed an interest were contacted by the evaluation team with a request for some additional information about their courses in order to enroll them in the study. Faculty who had originally indicated interest in the study but who did not respond to follow-up emails from the evaluation team, were contacted one to two additional times to solicit their participation.

Twenty faculty from eleven community colleges participated in the Fall pilot evaluation, which focused on testing the assessment and survey instruments and conducting initial analyses using a sample of faculty and students from the Anatomy & Physiology and Psychology courses. Thirty-seven faculty and 1603 students from 24 community colleges across the United States participated in the Spring study. Table 1 provides additional details about participation. As indicated in Table 1, fewer faculty were interested in participating as a non-OLI instructor. For this reason, faculty who were interested in participating as an OLI instructor were asked if they would be interested in teaching one section using OLI resources and one section of the same

course without using OLI resources. Three faculty agreed to teach an OLI and a non-OLI section for an additional monetary incentive. Another faculty member agreed to use OLI to teach his Psychology course, while not using OLI to teach his Statistics course. Students were officially part of the study if they completed a consent form and a pre-test.

Table 1. Faculty and Student Participation in the CC-OLI Evaluation

	OLI faculty	Non-OLI faculty	Non-OLI /OLI faculty	OLI students	Non-OLI students
Anatomy & Physiology	7	3	1	286	134
Biology	6	1	1	243	89
Psychology	7	1	0	467	17
Statistics	7	2	1	286	81
TOTAL	27	7	3	1282	321

Data Collection

Pre- and post-assessments. For each course, we worked with an external assessment group to obtain approval for using assessment items from recognized, reliable and valid assessments of content knowledge in each specific subject area. We used the same pre- and post-assessment for each course:

- The Anatomy & Physiology (A&P) exam consisted of 42 assessment items drawn from a 100-item examination developed by (and used with permission from) the Human Anatomy & Physiology Society (HAPS). The HAPS examination is intended to test students on content covered over two terms of A&P (often labeled A&P I and A&P II in the community colleges where we conducted this evaluation). We thus relied on several content experts to choose 42 questions from the examination that would typically be covered in the first term of a college-level A&P course. The selected items were not matched to the topic areas in the OLI course.
- The Biology exam consisted of 40 items randomly drawn from a version of the Advanced Placement Biology examination and used with permission from the College Board. The selected items were not matched to the topic areas in the OLI course.
- The Psychology exam consisted of 40 items randomly drawn from a version of the Advanced Placement Psychology examination and used with permission

from the College Board. The selected items were not matched to the topic areas in the OLI course.

- The Statistics exam consisted of the 40 items on the Comprehensive Assessment of Outcomes in Statistics (CAOS) test, used with permission of the assessment authors.⁹ The items were not matched to the topic areas in the OLI course.

Student surveys. Approximately 75% of participating students (n=1199 out of 1603 students) completed a student survey at the beginning of the spring term. The survey was administered at the beginning of the term in order to ensure the highest rates of participation and to ensure the inclusion of any students who might later drop out of the study or the course. As an incentive to complete the survey, students were promised entry into a drawing to receive one of three iPad Minis. The survey included demographic questions about students' background and college workload; students were also asked about their college major, college grades, their use of technology and their study skills.

Faculty surveys and syllabi. All but one faculty member (n=36) completed a faculty survey mid-way through the spring term. The survey asked faculty to provide information about their background, their instructional practices and about their course section(s) participating in the evaluation. Additionally, OLI faculty were asked to respond to a set of questions related to their use of OLI materials as well as how they asked their students to use the OLI materials. Thirty-two faculty also provided copies of their syllabi.

OLI system data. For 93% of students enrolled in the course sections taught by faculty using OLI (n=1188 out of 1282 students), the OLI system collected the following data: system logins, pages accessed, average page views, completion of quizzes and activities, performance on quizzes and activities, errors on quizzes and use of hints. The OLI system collected data from all 30 OLI faculty about their logins, pages accessed and viewed, completion of quizzes and activities, and use of the OLI Grade book and Dashboard tools.

Data Analyses

We analyzed survey data to provide descriptive information about the differences between OLI versus non-OLI faculty and students. We used "quasi-experimental" propensity score methods to compare achievement gains among OLI students similar

⁹ delMas, R., Garfield, J. Ooms, A. & B. Chance. (2007). Assessing students' conceptual understanding after a first course in Statistics. *Statistics Education Research Journal*, v6, n2, pp. 28-58.

to students who did not use OLI. Finally, we relied on hierarchical linear models and further qualitative analyses to understand which elements of OLI use and implementation had the greatest associations with student learning gains.

To understand the factors associated with pre-to-post gains across courses, we used standardized percent correct on the post-test as our outcome variable and included standardized percent correct on the pre-test as an independent variable. We also calculated IRT-adjusted pre-test and post-test scores using a 2-parameter IRT model for each course's test. However, because IRT-adjusted scores are meant to reflect item difficulty within particular exams, we only used the IRT-adjusted scores for within-course analyses.

Findings

PART I. How OLI Students and Faculty Differ from Non-OLI Students and Faculty in the Study

We begin this section by describing the differences between students who used OLI course materials and those who did not. Students of faculty who chose to use OLI were considered OLI students. We also note some measured differences among faculty who did and did not use OLI, although the small sample size kept us from gaining a full understanding of differences between OLI versus non-OLI faculty.

We should also note that our results may not be representative of what might be found across all U.S. community colleges. Faculty participants in our study came from community colleges across twelve U.S. states, including states in the Eastern, Western, Southern and Midwestern areas of the country. However, our combined (OLI and non-OLI) student sample includes fewer white students (45%) and more Hispanic students (31%) relative to the broader population of community college students in the U.S., which is 54% white and 16% Hispanic.¹⁰ The percentage of African American students in our sample is similar to that of the larger U.S. community college population (14%).

Compared to their non-OLI counterparts, OLI students had higher baseline pre-test scores, and they reported higher comfort using online resources and more time spent on academic work.

¹⁰ Demographic data regarding U.S. community college enrollment taken from American Association of Community Colleges (2012), 2012 Community College Fast Facts, Washington D.C., <http://www.aacc.nche.edu/AboutCC/Documents/FactSheet2012.pdf>

Using survey and assessment data, we uncovered several differences between OLI students and their non-OLI counterparts. Specifically OLI students had higher pre-test achievement scores, and they reported more of a preference for using online course materials and higher number of hours spent doing academic work in prior classes compared to their non-OLI peers.

According to faculty who responded to an informal poll, students sometimes knew before enrolling whether or not their course would have an online component. Thus, some of these difference between OLI versus non-OLI students such as the reported comfort level with online material, might be explained because certain students self-selected into courses with an online component.

Table 2 reports on all major variables for which we uncovered a statistically significant difference between OLI and non-OLI students. In addition to the higher comfort with online resources and more hours reported spent on prior academic work, OLI students were somewhat older than their non-OLI counterparts and more likely to be Hispanic and not white. OLI and non-OLI students did not differ significantly with respect to some critical characteristics, including gender, household income and the level of parental education.

We also used independent t-tests to compare faculty who opted to use OLI with those faculty who did not use OLI. In these comparisons, we labeled any faculty who taught both an OLI and non-OLI section of their course as “OLI faculty,” given our interest in comparing the characteristics of faculty who were not willing to use OLI at all to those who were. However, because we could only compare the small number of non-OLI-only faculty who completed particular items on the faculty survey (n=5) to OLI faculty who completed those same items (n=33), these comparisons almost certainly lack the statistical power required to understand true differences between OLI and non-OLI faculty. Nonetheless, we did find that OLI faculty reported spending less time lecturing compared to non-OLI faculty (p<.05 in independent t-tests, where equal variances between OLI and non-OLI groups cannot be assumed).

Table 2. Variables with Significant Differences for OLI vs. Non-OLI Students

Variable	Non-OLI students		OLI students		P-value in independent t-test
	N	Mean (SD)	N	Mean (SD)	
Hispanic (1=Yes)	196	.22 (.41)	988	.32 (.47)	.002*
White (1=Yes)	196	.58 (.49)	988	.43 (.50)	<.001
Years of age	195	23 (6)	994	24 (8)	.009

Pretest score (standardized percent correct)	320	-.13 (.92)	1278	.03 (1.02)	.011
Time spent on academic work (1=no hours; 6=>30)	197	3.1 (1.0)	996	3.5 (1.2)	<.001*
Comfort working in an online environment (1=Strongly disagree; 4=Strongly agree)	194	2.9 (0.9)	993	3.1 (0.7)	.002*
Preference for face-to-face classroom meetings vs. online-only courses (1=Strongly disagree; 4=Strongly agree)	195	3.5 (0.7)	992	3.1 (0.8)	<.001
Preference for using online materials in courses (1=Strongly disagree; 4=Strongly agree)	194	2.7 (0.9)	986	2.9 (0.7)	.024*
Preference for using a traditional text vs. online materials (1=Strongly disagree; 4=Strongly agree)	195	3.1 (0.9)	990	2.8 (0.8)	<.001*
Anticipated grade in [A&P/Biology/Psychology/Statistics] course (1=C- or lower; 6=A)	192	3.8 (1.1)	986	4.0 (1.0)	.01*

* p-value for independent t-test where equal variances between OLI and non-OLI groups cannot be assumed.

PART II. Comparing Achievement Gains of OLI vs. non-OLI Students: Quasi-Experimental and Regression Methods

We used quasi-experimental and regression methods to estimate the OLI “effect” on students’ achievement gains for the sub-sample of faculty who taught one section using OLI course materials and one section without using OLI course materials. The advantage of focusing on this sub-sample is the ability to “control” for differences in faculty and institutional characteristics that would be nearly impossible to fully account for if we attempted to compare OLI and non-OLI participants across the whole study sample. The sub-sample was composed of four instructors who agreed to teach both a "treatment" and "control" section of the same course. These four instructors taught one section of their course using OLI materials and they taught another section not using OLI. The drawback of focusing on the sub-sample of participants is that we have little power to detect significant differences between the OLI and non-OLI students’ achievement gains. Another drawback is that we do not know to what extent the information that the faculty member received from the OLI

grade book and dashboard in his or her OLI course influenced his or her teaching in the non-OLI course.

We used *propensity score matching* (PSM) as one tool to compare achievement gains of OLI to non-OLI students. Propensity score matching uses information from the group of individuals who did not participate in the intervention to evaluate what would have happened to those individuals who did in the absence of the intervention. PSM works to account for selection biases that are present because we did not randomize students to our “treatment” (OLI). As we discussed in the prior section, potential selection biases include that students using OLI were more comfortable using online resources, were somewhat older than their non-OLI counterparts and were more likely to be Hispanic and not white. To account for these biases, propensity score matching methods aim to find a non-OLI student that is similar to a student who used OLI resources. Then, an estimate of OLI’s impact is generated by calculating the difference in achievement for OLI versus non-OLI students. These differences are averaged across all participants, providing an estimate of the average treatment impact for those who participated in OLI.

The total sample size for the propensity score analysis was 186 students. Table 3 provides a breakdown of this sample by instructor and treatment status.

Table 3. Students Included in PSM Analysis by Instructor and Treatment Status

	Instructor 1	Instructor 2	Instructor 3	Instructor 4
Treatment (OLI)	8	32	34	10
Control (non-OLI)	21	32	29	20

A key advantage of matching students who used OLI with those who did not *within* instructor was that the treatment and control students were already matched on all instructor and institutional characteristics, both observed and unobserved. Therefore, it was possible to use only student-level covariates to estimate the propensity scores (or the probability of selection into OLI). We were able to include a rich set of covariates that we expected might predict both selection into OLI and/or the outcome given that all student who participated in the study, whether in an OLI or non-OLI course, were asked to complete a student survey. The covariates that were used to predict propensity scores are included in Appendix A, and a description of the steps we took to do our propensity score matching analysis is included in Appendix B.

In Table 4, the estimated “OLI effect” is presented for each of three models. The estimate of the OLI effect based on the results of Model 1 provides the most robust estimate we are able to obtain with the sample of students and faculty included in the

CC-OLI evaluation study. In addition to the fact that the estimated treatment effect is based on a matched sample of similar OLI and non-OLI students, this is the only model that accounts for all observed and unobserved instructor and institutional characteristics. In Model 1 we estimated the average effect of treatment on the treated (ATT) based on the propensity score matched sample. The ATT is the average effect of treatment on those subjects who ultimately received the treatment, or in this study, whose instructors agreed to use OLI.

In Model 1, the expected effect of OLI among students whose instructors agreed to participate in OLI is positive though not statistically significant ($ATT=.07, p=.24$). This result implies that OLI participants may experience a 7 percentage point improvement on the posttest relative to what their score would have been had they been in a traditional course setting. However, because this estimated effect for OLI does not reach statistical significance, we do not have a strong basis for concluding that the effect is different from zero. It is possible that the failure to reach significance is an artifact of the very small size of this sample.

For the purposes of comparison, we also provide the results from hierarchical linear models (HLMs) that could have been used to estimate the impact of OLI on student achievement gains among this subsample of 186 students. In Model 2 we use the (unmatched) sample of 186 students to estimate a hierarchical linear regression model in which the outcome (percent correct on the posttest) is regressed on an indicator of whether a student participated in OLI. This model also accounts for random variation among the four instructors of the students in this sample. In this model, the estimated OLI effect ($b=.05, p=.03$) is positive and significant; however, this model does not control for any pre-treatment student characteristics. In Model 3, after we account for all of the student-level covariates that were used in the matching procedure and random variation among the four instructors, the estimated OLI effect is essentially zero and non-significant ($b=.001, p=.95$).

Table 4. Estimated OLI effect for our Faculty Sub-Sample

	Estimated OLI effect ^a	<i>p</i> -value	Original <i>N</i>	Treated <i>N</i>	Matched <i>N</i>
Model 1 – ATT (matched)	.07(.06)	.24	186	84	84
Model 2 – HLM model (non-matched)	.05(.02)	.03	186	186	–
Model 3 – HLM model (non-matched)	.001(.03)	.95	186	186	–

^a Standard errors in parentheses; In Model 1, the estimate is the estimated difference in mean posttest percent correct between the matched treated and control groups, while in Models 2 and 3, the estimate is the estimated regression coefficient from the OLS or HLM regression model for the treatment status indicator.

PART III. Differences in OLI Implementation and Use among Faculty and Students

The prior sections focused on differences between OLI and non-OLI students based on whether a faculty member merely agreed to use OLI materials in his or her class section. In this section, we describe large implementation differences among faculty who agreed to use OLI. The large variation in faculty and student use and implementation of OLI provide a potential explanation for why we did not uncover significant differences between OLI and non-OLI students' achievement gains in our quasi-experimental comparisons: faculty used OLI in some less and more optimal ways, which – in turn – led to varying effects on student achievement.

We observed large variation in the extent to which faculty using OLI resources required students to complete OLI activities and assessments, graded students based on their OLI work, and prepared for class using OLI materials.

As illustrated in Figure 1, faculty reported wide variation in the percentage of the total required course tasks and grading that would be based on OLI materials. As can be seen in Figure 1, the highest number of faculty fell into the 0-25% range in terms of percentage of course tasks and required grade based on OLI. If OLI did influence gains in achievement, it would be much less likely to have an impact on achievement gains for the faculty that used OLI the least as part of their course tasks and assessment.

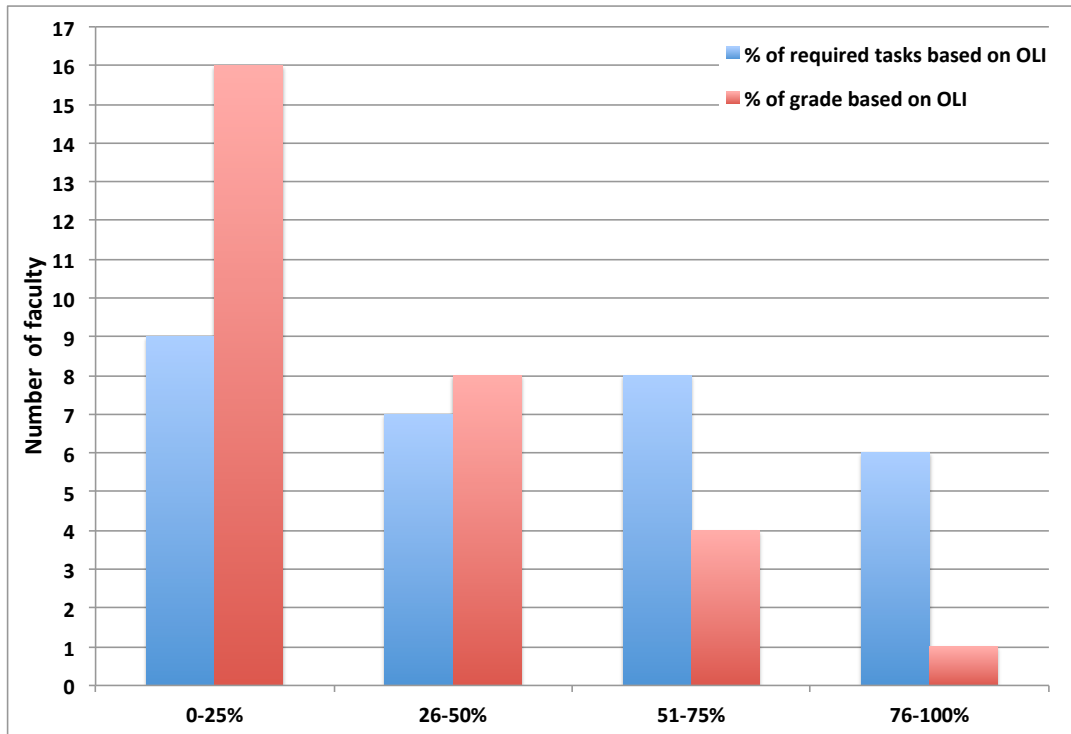


Figure 1. OLI Faculty Reports Regarding Use and Assessment Based on OLI

In preparing for their classes, faculty used OLI materials to varying degrees, both in terms of getting up to speed on the content within the modules and taking advantage of OLI tools, like the Grade book and Dashboard.

The majority of faculty reported reading through OLI explanatory content and completing quizzes and activities about once a month during the course of their instruction. Faculty were somewhat more likely to complete a quiz compared to completing a Did I Get This? or other kind of OLI activity in preparation for their lessons. They reported somewhat more use of the Grade book and Dashboard as compared to use of activities and quizzes. Figure 2 provides an overview of faculty reports about their use of OLI materials.

We have also captured information from the OLI system regarding instructor logins, page views and use of Grade book and Dashboard. In Table 5, we summarize some of the data regarding faculty use of the OLI system. When faculty taught two separate OLI courses with the same content, we calculated the average across their courses for Learning Dashboard and Grade book views. We took into account the total number of pages in each faculty member’s course when calculating percent of pages accessed and viewed. As can be seen in Table 5, the standard deviations for each indicator are almost as large or larger than the means themselves, suggesting a huge amount of variation with regard to faculty use of the OLI system.

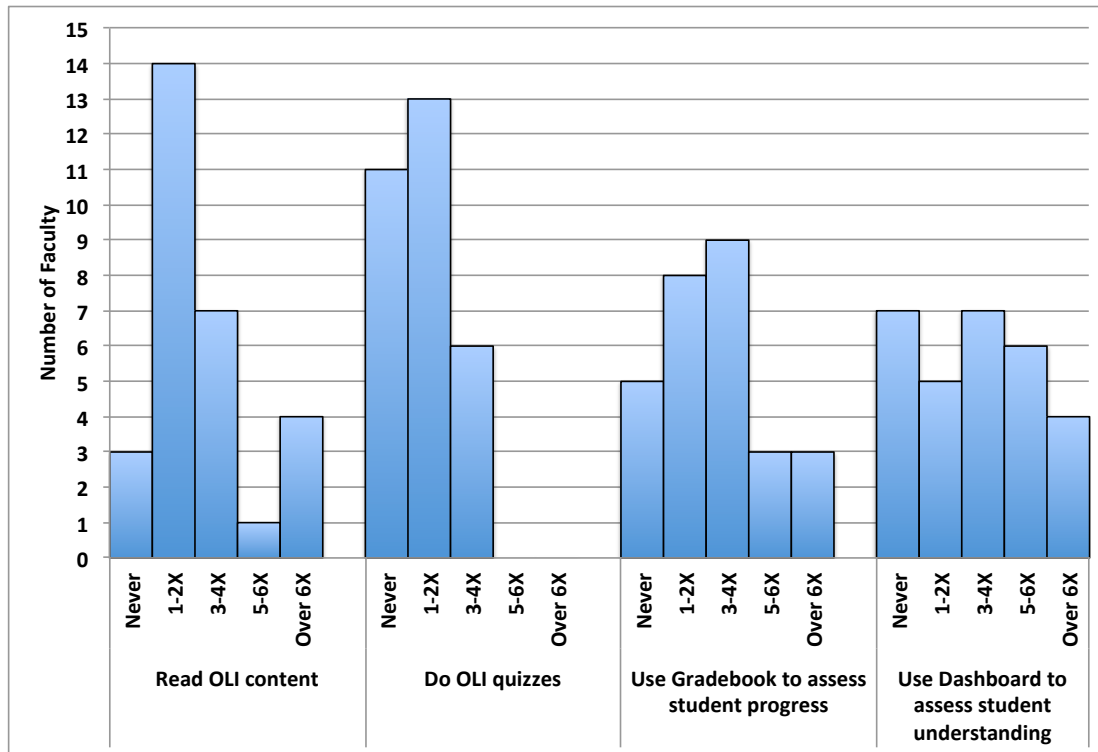


Figure 2. Faculty Times Per Month Using OLI Materials to Prepare for Class

Table 5. Means and Standard Deviations (SD) for Faculty Use of OLI System (n=30)

	Mean (SD)
Logins	75 (64)
Percent of pages of the OLI course instructor accessed	26% (25%)
Average number of views of a single page	.50 (.54)
Learning Dashboard views	39 (61)
Grade book views	26 (25)

Students used OLI resources to varying degrees; students in Psychology and Statistics used OLI somewhat more than students in Biology and Anatomy & Physiology.

Just as faculty reported variable use of OLI, students also used OLI resources to varying degrees. Our OLI system collects data about student performance, including the number of times students logged in, the percent of activities and quizzes students completed, the number of hints and incorrect feedback the students receive while engaged in learning activities, the extent to which students investigated their progress by looking at the “My Scores” reports and student Learning Dashboard.

As can be seen in Figure 3 below, the largest proportion of students logged into the OLI system 20 or fewer times, although there were considerable differences in the number of student logins by course. For example, in A&P, 73% of participating students logged into their OLI course 20 or fewer times. In Statistics, on the other hand, only 10% of participating students logged in 20 or fewer times.

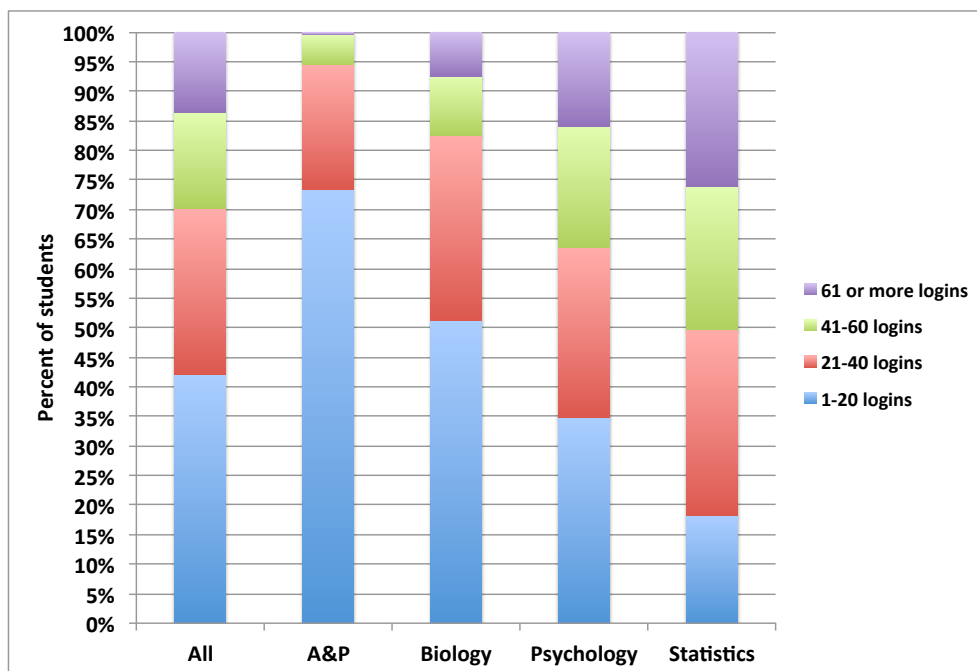


Figure 3. Distribution of Students Total Logins Overall and By Course

Table 6 reports on the mean percent of pages accessed, Learn By Doing and Did I Get This activities started, and quizzes completed. To calculate these percentages, we took into account total pages, activities and quizzes that instructors made available, which varied by course and instructor. As with logins, we see the trend of students using Statistics and Psychology course materials more so than students in A&P and Biology.

Table 6. Means (Standard Deviations) for Student Use of OLI Materials

	All	A&P	Biology	Psychology	Statistics
% Pages Accessed	49% (34%)	42% (33%)	38% (34%)	53% (34%)	58% (30%)
% Learn By Doing activities started	36% (33%)	33% (32%)	27% (33%)	37% (32%)	44% (33%)
% Did I Get This activities started	39% (34%)	33% (32%)	28% (33%)	37% (34%)	54% (34%)

% quizzes completed	54% (37%)	52% (39%)	41% (39%)	61% (35%)	55% (32%)
---------------------	--------------	--------------	--------------	--------------	--------------

PART IV. How Faculty and Student Use of OLI are Associated with Student Outcomes

In this section, we focus on how faculty and student use of the OLI system – in concert with other factors – is related to students’ achievement gains from the pre- and post-tests we administered. As illustrated in the figure below, our findings indicate that faculty use of OLI is indirectly associated with students’ achievement gains through its strong and significant positive relationship with students’ use and performance within the OLI system. Faculty use may also have a direct relationship with students’ achievement gains that we did not uncover due to the small number of faculty participating in the study. We used hierarchical linear models to understand relationships among faculty use, student use and student achievement gains.

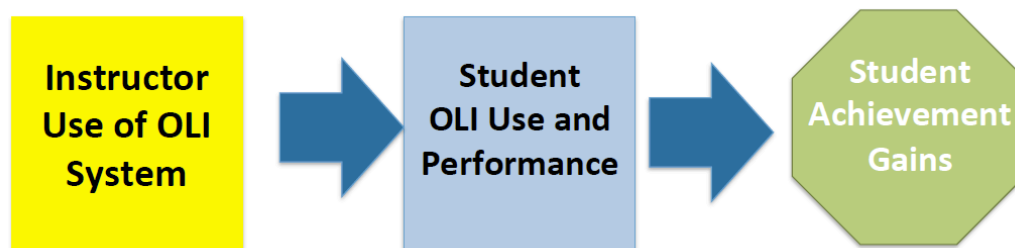


Figure 4. Relationships among Instructor OLI Use, Student OLI Use, and Student Achievement Gains

Instructor use of the Grade book and course pages accessed were significant predictors of multiple variables measuring OLI student use and performance. In turn, student use of OLI and performance on OLI activities and quizzes were some of the largest significant predictors of student learning gains from pre-test to post-test.

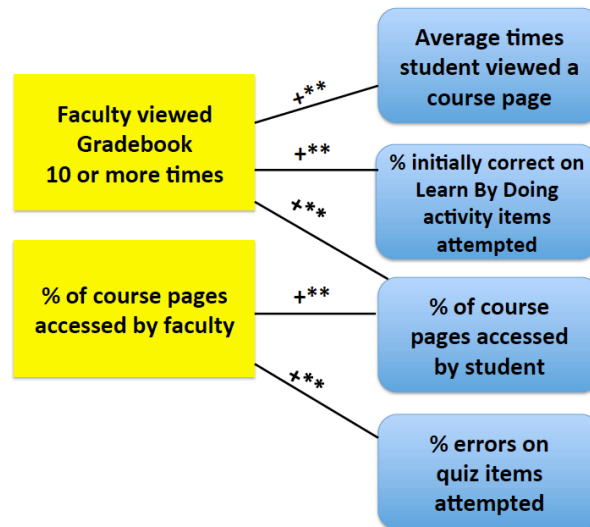
We used multiple hierarchical linear models to assess whether any faculty use variables were associated with student use and performance in the OLI system. As indicated in Figure 5, we found that two variables measuring instructor use of OLI

were significantly associated with multiple measures of students' OLI use and performance: faculty viewing their OLI Grade book more than 10 times and faculty accessing a higher percentage of pages in their course. As shown in Figure 5, those two faculty use variables were significant predictors of the following dependent variables in separate HLMs: (1) students accessing a higher percentage of online pages within OLI; (2) higher average student views per OLI page; and (3) higher correct response on initial try for Learn By Doing activity items. Additionally, we uncovered the somewhat counterintuitive finding that an instructor who accessed a higher percentage of course pages was significantly more likely to have students who made a higher percentage of errors on quiz items. HLMs controlled for additional independent fixed effects alongside faculty use, including students' pre-test score, whether a student was black or Hispanic and course (Biology, Psychology or Statistics; A&P) as comparison. Random effects in HLM models took into account both instructor-level and college-level variation.

Our HLMs suggest, for example, that a Hispanic student with a pre-test score one standard deviation above the average score, who has an instructor who looks at the Grade book at least 10 times and views 70% of the pages in a course, is predicted to access about 55% of the pages in the course; whereas another Hispanic student with a similar pre-test score, who has an instructor who does not use the Grade book and only views 5% of the course pages, is predicted to access only 22% of the pages in their course. Similarly, a Black student with a higher than average pre-test score, who has an instructor who looked at the Grade book at least 10 times, is predicted to get about 44% of their initial responses to Learn By Doing items correct versus 37% initial responses correct for similar students who do not have instructors who look at the Grade book or look at it fewer than 10 times.

Figure 6 reports on the extent to which the student variables that have a significant association with faculty use *also* predict student learning gains (students' standardized percent correct on post-test when controlling for standardized percent correct on pre-test). As can be noted in the figure, the percent of course pages that students access, percent of Learn by Doing activity items to which students gave initial correct responses, and percent of errors on quizzes are all significant predictors of learning gains, whereas students' average per page views was not significantly associated with learning gains when we accounted for other independent variables in our regression analyses.

Figure 5. Faculty OLI Use Variables Associated with Student OLI Use and Performance in HLMs*



+ = Significant positive effect; ** $p < .01$

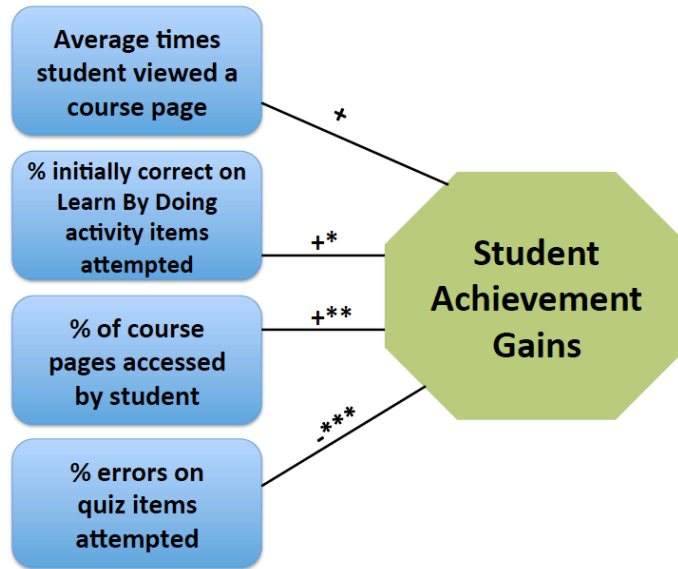
* HLMs for faculty use on student use/performance account for random instructor-level and institution-level variation and fixed effects for pre-test percent correct, as well as dummy variables for black, Hispanic, and course (e.g. Biology, Psychology, Statistics), in addition to the faculty use variables noted in the table.

We are unsure why students' initial percent correct on Learn By Doing activities is significantly associated with instructor use of the Grade book whereas students' initial percent correct or eventual percent correct on other kinds of activities or quizzes is not. We surmise that instructor use of the Grade book may be a proxy for their recommendation for students to engage in and do well in low-stakes activities like Learn By Doing activities, which – in turn - encourages students to attend to those activities.

In addition to the student use and performance variables included in Figure 6, significant predictors of standardized percent correct on post-test – over and above standardized percent correct on pre-test, which was always significant – included: less time students reported spending on non-academic tasks ($p < .05$) and instructor years of teaching experience ($p < .01$). Being black or Hispanic, as well as teaching experience, were also significant negative predictors of post-test scores in initial models, although these significant effects disappeared when we added all the OLI student use and performance variables in our final HLM.

Associations for all fixed effects in our HLMs are included in Appendix B.

Figure 6. Student OLI Use and Performance Associated with Student Achievement Gains in HLMs*



+ = Positive effect; - = Negative effect; * $p < .05$; ** $p < .01$; *** $p < .001$

*HLMs for student use/performance are on standardized percent correct on post-test, controlling for standardized percent correct on pre-test, as well as random effects for instructor/college and other fixed effects that led to best model fit, including course, black, Hispanic, students' reports of time spent on non-academic activities and instructor teaching experience

Thus, as an extension of our previous example, our analyses suggest that a Hispanic student with a higher than average pre-test score who has instructor who accessed their Grade book more than 10 times and views 70% of the course pages is predicted to access about 55% of the pages in the course and get about 56% of the Learn By Doing items correct on their first try. A student with that predicted percent pages accessed and LBD items correct is, in turn, predicted to have a standardized percent correct on their post-test of about .90. On the other hand, a similar student with an instructor who does not access their Grade book and only views 5% of the course pages is predicted to access only 24% of the course pages and get 50% of the LBD items correct on the first try. That student, in turn, is predicted to have a standardized percent correct on their post-test of about .64, a score that is about .26 standard deviations lower than .90.

Faculty’s experience with OLI and their consistent, less variable, use of OLI may have further encouraged students’ higher OLI use and, as a consequence, their learning gains.

We conducted further descriptive analyses by comparing the highest and lowest-performing faculty in our sample. Specifically, we focused on nine faculty with average student gains in standardized percent correct and/or IRT-adjusted scores above .50 (who we call high-performing faculty) and eight faculty with gains below .10 in both IRT-adjusted and standardized percent correct scores (who we call low-performing faculty). Readers should keep in mind that comparisons between these faculty do not include the fourteen “middle-performing” faculty whose students were neither high- nor low-performing.

In Table 7, we summarize how these high-performing faculty compared to low-performing faculty on a number of factors, including those that we explored in earlier regressions. These data, taken together, indicate that the high-performing faculty are those with more experience using OLI and more teaching experience in general. Although high-performing OLI faculty had a lower mean number of logins to the OLI system compared to low-performing OLI faculty, as well as slightly lower Grade book views, high-performing faculty were much more consistent in their logins and Grade book views. Specifically, high-performing faculty tended to log into the OLI system regularly from week to week, whereas low-performing faculty often logged in many times in a single week and rarely in the next week. Furthermore, standard deviation – a measure of variation – in regard to logins and Grade book views was twice as large for low-performing faculty compared to high-performing faculty, meaning that many more low-performing faculty either logged into OLI and viewed their Grade book a great deal or almost not at all whereas OLI logins and Grade book views were more consistent among high-performing faculty.

The high variation in use of OLI among low-performing faculty implies two types of low-performing faculty: (1) OLI users who spent a lot of time trying to familiarize themselves with OLI, which is likely, considering that low-performing faculty often had not used OLI previously; and (2) OLI users who were less motivated – for whatever reason – to explore OLI and instructor tools like the Grade book. On the other hand, students of high-performing faculty may have benefited from the greater relative experience of the faculty and their regular use of the OLI system.

Table 7. Differences between Highest and Lowest-Performing Faculty

	High-performing faculty (Faculty with average student standardized percent correct and/or IRT-adjusted achievement gains above .50, n=9)	Low-performing faculty (Faculty with average student standardized percent correct and IRT-adjusted achievement gains below .10, n=8)
Faculty used OLI resources before Spring term	78% (n=7)	25% (n=2)
Faculty years teaching experience	100% with 10 or more years experience 56% (n=5) with 20 or more years experience	25% (n=2) with 10 or more years experience 0% with 20 or more years experience
	Mean (Standard deviation)	Mean (Standard Deviation)
Faculty logins to OLI	56 (40)	81 (73)
Faculty OLI Grade book views	21 (10)	23 (20)
% OLI course pages student accessed	52% (15%)	40% (26%)
Average student views per page	1.2 (.44)	.93 (.64)
% initial response correct on LBD items attempted	58% (14%)	52% (19%)
% errors on quiz items	37% (12%)	43% (14%)

Lastly, as indicated in our earlier regressions, high-performing faculty had students who – on average – accessed more OLI pages, had higher average views per page, higher percent of initial responses correct on Learn By Doing items and lower percent of errors on quiz items compared to low-performing faculty. As with faculty use, we see higher variation in student use among students of low-performing faculty compared to high-performing faculty. These descriptive findings reinforce the notion that higher faculty use of OLI encouraged higher student use and more of students’ attention and care on activities and quizzes.

We did not find that faculty who assigned a higher percent of OLI materials as required reading or graded a higher percent of OLI activities had higher average student use and performance. In fact, many low-performing faculty reported assigning OLI as their main text for their course and reported basing a higher percentage of required readings and grade on students’ use and performance in OLI compared to high-performing faculty. However, as we noted above, the lower use of OLI among students of low-performing faculty compared to high-performing faculty did not reflect those higher requirements. Thus, simply “requiring” OLI may not be enough motivation for community college students to use it. Instead, students who use OLI more and perform better may be taking their cues from faculty who appear to them to be more familiar with the OLI system and who are using it to assess student learning.

Challenges for Research on Instruction in Higher Education Settings

We encountered a number of challenges over the course of the year we spent attempting to better understand how community college instructors and their students use the Open Learning Initiative (OLI) materials in hybrid mode and the impact of their use on student learning. We grappled with dilemmas that arose with respect to recruiting faculty and students, defining faculty implementation of OLI, and drawing appropriate comparisons between similar faculty and students in OLI and non-OLI conditions. We believe that the challenges we encountered are not unique and can serve to highlight issues that will be important for others to consider as they approach the study of online modes of instruction in higher education settings.

We decided from the outset that we would conduct a quasi-experimental analysis rather than run a randomized controlled trial of OLI. We made this decision for multiple reasons, but it was mainly driven by our desire to understand faculty implementation and students' learning gains in naturally-occurring environments where faculty made their own decisions about the extent to which they would use OLI.

We were only able to enroll a small fraction of the large number of faculty initially contacted by Lumen Learning to participate in the study. We especially did not receive enough interest from faculty who agreed to serve as "control" faculty, those who would not use OLI materials with their students. Low faculty participation meant that we lacked power to detect differences in learning gains between OLI and non-OLI groups through our quasi-experimental analysis.

Earlier work that OLI researchers had conducted with community college faculty revealed that in an average class, the instructors spend only about 15-20 minutes preparing for each class; all instructors reported prep times in the 0-30 minute range. The short prep time is likely due to the fact that many faculty in community colleges teach 3-5 classes each term. The faculty members interviewed in the earlier study were part of a project unrelated to OLI and had to apply to be part of that project, so the interview sample was likely to be biased toward faculty who were highly motivated and innovative. Less motivated, innovative faculty likely might report even short prep time.

The limits on the degree to which faculty could prepare to use the OLI materials for the current study was likely compounded by the short time, overall, that faculty had access to the OLI course material and instructor tools. Three of the four OLI courses in the study were courses that had only been completed and ready for use in the Fall of 2012. The Instructor Learning Dashboards associated with Biology and A&P

courses were not ready until the Spring 2013, when the evaluation was taking place. The effective use of the OLI instructor tools, the Grade book and Learning Dashboard, require some knowledge and skill on the part of faculty. As is the case with development of any new knowledge or skill, effective practice with opportunities for feedback and reflection followed by additional practice is critical. The current study only looked at student learning gains in a single term with students taught by faculty who were new or nearly new to OLI and had little time to prepare for shifting their teaching practice.

Our choice to run a quasi-experimental design also meant that faculty selected whether to use OLI or not on their own. While allowing this choice was important for measuring naturally-occurring differences in implementation, it meant that OLI faculty – and the institutions in which they taught – could be different from non-OLI faculty in multiple ways that could not necessarily be “controlled for” in a propensity score matching analysis. Perhaps more importantly, OLI faculty later reported that their students often knew that they would be using a course with an online component, which likely led to some considerable measured differences between OLI and non-OLI students, especially on the measure of comfort in using online courses.

In addition to the large differences in the way faculty in the OLI condition chose to use the OLI materials, there were large differences in the way faculty in both conditions chose to implement the pre and post-test, which made it difficult to make judgments about the efficacy of the OLI materials. For example, some faculty delivered the post-test as part of the final exam in the course while others delivered the post-test as a low stakes unrelated activity, sending very different signals to the students about how much effort to expend on the post-test.

All these large differences between OLI and non-OLI faculty and students, as well as large differences within each condition imply that differences in learning gains between the two groups could be attributed to any number of self-selection factors beyond their agreement to use OLI or not. In our propensity score matching analysis, we attempted to minimize those differences by only comparing students of faculty who agreed to teach both an OLI and non-OLI section. But, even in those circumstances, students sometimes knew that they were opting into or opting out of an online section, which likely led to self-selection biases on the part of the students, if not the faculty.

Summary and Implications

One of our goals for this evaluation was to understand differences between OLI versus non-OLI faculty and students and explore achievement gains between those two groups. We also set out to understand how faculty and students used OLI and how

that use was related to student achievement.

Our quasi-experimental comparisons suggested a positive but non-significant effect of Open Learning Initiative courses on achievement. The non-significance of the findings can be explained, in part, by the huge variation in faculty use of the OLI system. OLI faculty required OLI readings and graded them to greatly varying extents. Faculty also used the OLI system to differing extents, and that use was significantly related to students' use of OLI. Specifically, in hierarchical linear models accounting for variation within instructor and community college institution, the percent of all course pages that faculty accessed and faculty use of the OLI Grade book was significantly and positively associated with students' use and performance of OLI activities and quizzes. Students' OLI use and performance, in turn, was a large positive predictor of student achievement gains from pre- to post-test.

Descriptive analyses comparing a smaller group of faculty with highest versus lowest student achievement gains suggest that highest-performing faculty had more teaching experience and experience using OLI compared to lowest-performing faculty. Additionally, highest-performing faculty used OLI more strategically in their course while at the same time utilizing OLI tools like the Grade book regularly to assess student learning.

These findings have several implications for faculty utilization of online resources and further research on the use of online resources in college settings. Our analysis on implementation suggests that faculty and students can use OLI in ways that are more and less productive for student learning and achievement. Specifically, using OLI as a main text in a course or requiring more or less of student work in OLI does not appear important for encouraging good student use of OLI and related student achievement. Instead, faculty who familiarized themselves with OLI resources by looking at OLI course pages themselves and using the OLI Grade book to assess student learning were those faculty whose students also used OLI in ways that were positively associated with their achievement gains.

Our findings also suggest that researchers proposing to understand the efficacy of instructional interventions in higher education settings via experimental and quasi-experimental studies should first investigate differences in "treatment" implementation. Large differences in implementation like the ones we observed for faculty using OLI suggest that some "treatment" conditions in experimental treatment-control comparisons could be more similar to some control conditions than to other treatment conditions. When such diversity in implementation exists, experimental research is unlikely to uncover significant effects.

Study of implementation, rather than efficacy, is thus likely an important precursor to

considering any experimental study designs. Beyond providing evidence of the multiple forms a “treatment” intervention could take, implementation studies can point us to the key aspects of large interventions that most likely matter for what students learn. For example, in our study, faculty use of the OLI Grade book rather than the explicit assignment of OLI as a text was significantly associated with student use of OLI and – as a consequence – student achievement gains. If implementation studies found further evidence of a strong relationship between faculty use of the Grade book and student use and performance, OLI faculty trainings and workshops could emphasize and support Grade book use as an effective piece of the OLI system rather than the simpler interventions of simply assigning more work in OLI. Such trainings and workshops could support more consistent implementation across faculty and potentially set the stage for better experimental work.

Future Directions for Research

Our findings suggest that much research lies ahead in terms of understanding the relationship between faculty use of open online resources and student learning gains in higher education settings. We cannot expect faculty use of online course materials to lead to any improvements in teaching and learning without further research to understand faculty implementation of online course materials and – particularly – how developers of those materials can encourage optimal faculty implementation.

By studying how instructors take up and use OLI course materials – and how that use is related to student learning – our evaluation work has moved the field closer to better understanding optimal implementation of online resources. While our study points to important aspects of faculty and student use of OLI that lead student achievement gains, we have not yet arrived at a full understanding of what “optimal use” of OLI materials entails, which suggests several pathways for further research both using our data and engaging in future data collection efforts.

First, we conducted our analysis across multiple OLI courses that are quite different in terms of content, stage of course development, and particular design choices made by course development teams. Further research could explore the usefulness of particular kinds of activity and quiz items used within particular courses in association with a range of outcomes, including both use and performance outcomes within those courses and student performance on the pre- and post-tests used to assess learning gains for this study. In addition, researchers could focus in on the relationship between the students’ use of activities designed to meet specific learning objectives within OLI and their performance on subsets of pre- and post-test items that assess student knowledge related to those learning objectives.

Second, our findings on the indirect relationship between faculty use and student performance implies the possibility of further research using path analysis, structural equation modeling and/or other statistical tools to map out the connections between particular elements of faculty use, their associations with student use and – as a result – student performance gains. We are beginning to embark on some of that work now.

Finally, our work points to the need for additional research to investigate faculty's productive use of online resources and supports for that use, especially in higher education settings. In such settings, faculty are often given free reign to choose course materials that will best help students learn and may have little time or incentive to engage in additional training and professional learning activities to support the best use of course materials. If researchers can hone in on the elements of faculty use that matter most and the minimal support necessary to support that use, they can provide that information to leaders and faculty to help them make decisions about the most efficient and useful steps that faculty should take to use online resources effectively.

Our work indicates that tools like the OLI Grade book are helpful resources for faculty to assess student progress and knowledge and that those tools encourage higher student use of OLI. Producers of course materials should consider how such tools for instructors to track and assess student learning can be built into all online courses, not only to help faculty to plan, grade, and provide instruction that addresses gaps in students' understanding but also to motivate students to use the online course materials. If thoughtfully designed, those tools could also provide important data to all stakeholders, including designers and evaluators who are looking for clear evidence about what aspects of the course materials support student learning.

Appendix A

Table A-1. Descriptive Data for Propensity-Score Matched Sub-Sample Variables

	OLI Students	Non-OLI Students
	Mean/Mode ^a	Mean/Mode
Pretest Percent Correct	37 %	35 %
Gender	Female	Female
Race	Non-White	White
Native language	English	English
Year of birth	1988	1989
Mother's education level	Some college	Some college
Father's education level	Bachelor's degree	High school diploma/GED
Combined family income	\$20,000-\$39,999	\$80,000 or more
Already have a degree	No	No
Full-time student	Yes	Yes
Number of credits this term	11	11
Cumulative GPA	3.50	3.5
Majoring in a STEM field	No	No
This course is required for my major	No	No
Time spent each week preparing for classes	6-10 hours	6-10 hours
Time spent each week working for pay	0 hours	21-30 hours
Time spent each week caring for children or other family	0 hours	0 hours
Time spent each week on activities on campus	0 hours	0 hours
Time spent each week commuting	1-5 hours	1-5 hours
Number of science and math courses taken at this institution	0	1
Cumulative GPA in science and math courses	3.5	3.0
Whether repeating this course	No	No
Grade expected in this course	A	A
Does not give up easily when faced with technology-related obstacles	Agree	Agree
Is comfortable working in an online environment	Agree	Agree
Prefers face-to-face classroom	Agree	Agree Strongly

meetings compared to online-only		
Prefers using physical textbook instead of online text	Agree	Agree Strongly
Believes is responsible for own education	Agree Strongly	Agree Strongly
Adjusts own behavior and time management in order to complete course requirements	Agree	Agree
Is good at completing tasks independently	Agree Strongly	Agree

N=186 (OLI N=84; Non-OLI N=102)

^a Most covariates used in the estimation of propensity scores are categorical in nature and the modal response is provided for both groups on these covariates.

^b This is the outcome measure and is provided for descriptive purposes only; this variable was not included as a covariate in the estimation of propensity scores.

Table A-2. Descriptive Data for Variables used in final hierarchical linear models (OLI faculty and students only)

Variables in HLM on Student Use and Performance in OLI	N	Mean (Standard Deviation)	
(1) Standardized Pretest Percent Correct	1278	.03	(1.02)
(2) Black student	988	14%	(35%)
(3) Hispanic student	988	32%	(47%)
(4) Biology (compared to A&P)	1281	22%	(42%)
(5) Psychology (compared to A&P)	1281	36%	(48%)
(6) Statistics (compared to A&P)	1281	22%	(42%)
(7) Grade book viewed 10+ times by instructor	30	26	(25)
(8) % OLI course pages instructor accessed (out of all pages instructor made available to students)	30	26%	(25%)
Variables in HLMs on Student Standardized Post-test Percent Correct			
Variables 1-6 above			
(7) Time student reported spending on non-academic work per week (e.g. childcare, outside activities, etc.) [1=None; 6 = More than 30 hours]	1195	2.5	(.8)
(8) Average times student viewed an OLI course page	1186	1.2	(1.1)
(9) % OLI course pages student accessed (out	1186	49%	(34%)

of all pages instructor made available to
students)

(10) % correct initial responses on Learn By Doing activity items attempted	1177	59%	(24%)
(11) % errors on quiz items attempted	1080	37%	(21%)
(12) Faculty years teaching experience [1=1-3 years; 5=More than 12 years)	30	4.1	(1.3)

Appendix B.

Analyses Steps for Propensity Score Matching

All propensity score analyses were conducted using R version 3.0.1 (<http://www.r-project.org/index.html>). Our analytical process involved a series of steps. The first step involved the selection of covariates to be used in the estimation of propensity scores, followed by the actual estimation of propensity scores. Next, each individual in the treated (OLI) group was matched to an individual in the non-treated group based on the propensity scores. The quality of the matching was assessed using a number of balance checks and, finally, the effect of OLI use was estimated for the students enrolled in an OLI course. Each of these steps is discussed in greater detail below.

The overall response rate on the student survey among all students who participated in the study was 80%. With almost all methods used to estimate propensity scores, however, it is necessary to have full coverage on all included covariates; otherwise, any cases with missing information on one or more covariates are dropped from the analysis. Therefore, prior to the estimation of propensity scores, we used multiple imputation to fill in missing values on these student characteristics. Five datasets were multiply imputed using the "mice" package in R. Propensity score matching and estimation of the treatment effect were conducted using each of the five imputed datasets and then the results were pooled (averaged).

Matching was conducted using the "Matchby" function in R, which is part of the "Matching" package. Matches between control and treatment students based on their estimated propensity scores were restricted to occur only among students who had the same instructor, which ensured perfect balance on faculty and institutional characteristics. We used one-to-one (one treated matched to one control) matching in which each treatment student was matched to the "nearest neighbor" control student, or the control student with the most similar estimated propensity score. Matching with replacement was used. Matching with replacement often results in better matches because controls that look similar to many treated units can be used more than once. Following the matching process, the function "MatchBalance" was used to assess the extent to which the matching process was successful in achieving balance between the treated (OLI) and non-treated students on all of the included covariates.

Analyses Steps for Hierarchical Linear Models

We used hierarchical linear models both to investigate (1) the relationships between faculty use and student use/performance in the OLI system and (2) variables related to students' achievement gains. All hierarchical linear models were conducted using R version 3.0.1 (<http://www.r-project.org/index.html>). To arrive at our final models, we

tested many hierarchical linear models (HLMs) using the lme4¹¹ package in R. Given the nested nature of the data (i.e. students nested in teachers and in institutions), we reasoned that HLM was the best choice for our analysis. In our final HLMs, we included the mixed effects that provided us with the best model fit, based on deviance statistics, and we omitted variables that did not have significant effects or make a difference for the fit of the models.

Table 1-B includes information on all fixed effects in separate HLMs for four main student-level dependent variables measuring student OLI use and performance. Random effects for teacher and institution were also included in all models. As can be seen, the last two fixed effects for instructor use of OLI are those included in Figure 5 of the report.

Table 1-B. Significant fixed effects in hierarchical linear models predicting student use and performance

	Dependent student-level variable			
	% course pages accessed	Average views per page	% of correct initial response on Learn by Doing items attempted	% errors on quiz items attempted
Standardized Pretest Score	+**	+*		_-***
Black student	_-***		_-***	_-***
Hispanic student	_-**	_-**		_-***
Biology (compared to A&P)				
Psychology (compared to A&P)		+*	+**	_-*
Statistics (compared to A&P)		+*	+***	
Instructor views Grade book more than 10 times	+**	+**	+*	
Percent of course pages instructor accesses	+**			+**
Student N	939	939	930	851
Instructor N	30	30	30	30
College N	22	22	22	22

+ = Significant positive effect; - = significant negative effect; * p<.05; ** p<.01; *** p<.001
 * Hierarchical linear models were used in the above regressions to account for random teacher-level and

¹¹ Bates, D., Maechler, M. & Bolker, B. (2011). *Lme4: Linear mixed-effects models using s4 classes*: Available at <http://cran.r-project.org/web/packages/lme4/index.html>

institutional variation, in addition to the fixed effects listed for each model.

Table 2-B includes information on all fixed effects in our final HLM on student standardized percent correct on the post-test (controlling for standardized percent correct on the pretest and thus, in effect, examining factors associated with pre-to-post gains). Random effects for teacher and institution were also included in the model. In the final model below, we accounted for 30% of the variation in post-test score.

The effects for student use of OLI system are those reported in Figure 6 of our report.

Table 2-B. Significant fixed effects predicting OLI student standardized post-test score in hierarchical linear model

	Independent variables	Effect
	Standardized pre-test score	+***
	Biology (compared to A&P)	
	Psychology (compared to A&P)	
	Statistics (compared to A&P)	
Student survey variables	Black	
	Hispanic	
	Time spent on non-academic work (e.g. childcare, outside activities, etc.)	-*
Student use of OLI (from system data)	Average views per page	
	% OLI course pages accessed	+**
	% items initially correct on Learn by Doing assessments	+*
	% errors on quizzes	-***
Teacher survey variables	Teacher years of experience teaching course	+**
Student N		575
Instructor N		27
College N		20

+ = Significant positive effect; - = significant negative effect; * p<.05; ** p<.01; *** p<.001

