SJSU Plus AUGMENTED ONLINE LEARNING ENVIRONMENT

PILOT PROJECT REPORT

September 2013

Prepared for Principal Investigator: Elaine D. Collins, Ph.D

***Research Team**: The research was conducted by the Research and Planning Group for California Community Colleges (RP Group) in collaboration with the Office of Institutional Effectiveness and Analytics at San José State University (SJSU).*

***RP Group Team:** Rob Firmin, Ph.D.; Eva Schiorring, MPP; John Whitmer, Ed.D.; Terrence Willett, M.S.*

***SJSU:** Sutee Sujitparapitaya, Ph.D*

Partial support for this work was provided by the National Science Foundation's Transforming Undergraduate Education in Science, Technology, Engineering and Mathematics (TUES) program under Award No. 1321336. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Executive Summary

Introduction: The report that follows presents highlights from a study of Augmented Online Learning Environments (AOLE) delivered by San José State University (SJSU) during the Spring 2013 semester through a collaboration with Udacity, a Silicon Valley-based provider of Massive Open Online Courses (MOOCs). The AOLEs included: a remedial-algebra survey course (Math 6L); introduction to college-level algebra (Math 8); and introduction to college-level statistics (Stat 95).

Research Design and Approach: The study was guided by three research questions:

1. Who engaged and who did not engage in a sustained way and who passed or failed in the remedial and introductory AOLE courses?
2. What student background and characteristics and use of online material and support services are associated with success and failure?
3. What do key stakeholders (students, faculty, online support services, coordinators, leaders) tell us they have learned?

The study was fueled by four different data sources: (i) SJSU's student data base; (ii) Udacity platform data on student engagement with course content and support services; (iii) responses from three student surveys; and (iv) student, faculty and other key stakeholder perspectives on AOLE obtained through interviews and focus groups.

The first three data sources were integrated into a flat file that the research team used to perform a two-tiered quantitative analysis which included: (1) 35 contingency tables on student characteristics, effort expended, support use, and pass/fail (See Section 5); (2) six logistic-regression models, defined by matriculation and course, on pass/fail on 18 independent variables (see Section 6). The quantitative methods employed are fully described in Appendix A. Key stakeholder perspectives, obtained through the qualitative research component, helped provide context and inform the direction of the quantitative analysis (see Section 7).

Findings: The research found that matriculated students performed better than non-matriculated students and that, in particular, students from the partner high school were less successful than the other AOLE students. Pass rates varied significantly with course taken and by persistence of student effort as seen in the following table and figure.

Course	**% Passed**
MATH 6L Matriculated	29.8%
MATH6L Non-matriculated	17.6%
MATH 8 Matriculated	50.0%
MATH 8 Non-matriculated	11.9%
STAT 95 Matriculated	54.3%
STAT 95 Non-matriculated	48.7%
Total	**33.3%**

Table S.1. Pass Rates by Course & Matriculation

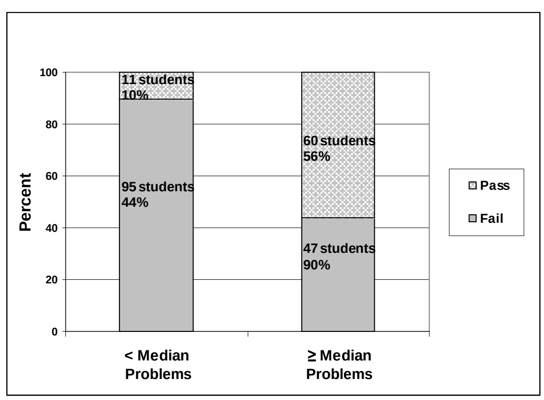


Figure S.1. Example of Student Effort Effect on Passing: Problem Sets Submitted

The statistical model found that measures of student effort trump all other variables tested for their relationships to student success, including demographic descriptions of the students, course subject matter and student use of support services. The clearest predictor of passing a course is the number of problem sets a student submitted. The relationship between completion of problem sets and success is not linear; rather the positive effect increases dramatically after a certain baseline of effort has been made. Video Time, another measure of effort, was also found to have a strong positive relationship with passing, particularly for Stat 95 students. The report graphs these and other relationships between variables examined by the logistic-regression models and pass/fail.

While the regression analysis did not find a positive relationship between use of online support and positive outcomes, this should not be interpreted to mean that online support cannot increase student engagement and success. As students, Udacity service providers and faculty members explained, several factors complicated students' ability to fully use the support services, including their limited online experience, their lack of awareness that these services were available and the difficulties they experienced interacting with some aspects of the online platform. It is thus the advice of the research team that additional investigations be conducted into the role that online and other support can play in the delivery of AOLE courses once the initial technical and other complications have been addressed.

Conclusion: The low pass rates in all courses should be considered in light of the fact that the project specifically targeted at-risk populations, including students who had failed Math 6L before Spring 2013 and groups demonstrated by other research to be less likely to succeed in an online environment. Previous studies (see Section 1) have found that these students do less well in online than in face-to-face courses. Further, student groups in at least one major study (Jaggars and Xu, 2013) who were found to experience the greatest negative effect from taking courses online share many of the characteristics found among the AOLE partner high school students in particular, a group with very low pass rates in Spring 2013.

Overall, much was learned during and from the first iteration of AOLE and improvements are already in progress in the second AOLE iteration. Perhaps most importantly, the faculty members who taught these courses, although they had to contend with major difficulties along the way, believe that the content that has been developed has tremendous potential to advance students' critical thinking and problem solving abilities. One faculty member summed it up this way: "Udacity has brought to the table ways to make the courses more inquiry-based and added real life context."

Table of Contents

Introduction

This report introduces findings generated by a study designed to inform the development and implementation of Augmented Online Learning Environments (AOLE) providing college preparation and introductory courses. The subject of the study is AOLE courses delivered by San José State University (SJSU) during the Spring 2013 semester through a collaboration with Udacity, a Silicon Valley-based provider of Massive Open Online Courses (MOOCs), including: a remedial/algebra survey course (Math 6L); introduction to college-level algebra (Math 8); and introduction to college-level statistics (Stat 95). These courses, referred to as SJSU Plus, were offered to matriculated SJSU students and non-matriculated students, including 60 students from a partner high school serving economically disadvantaged youth.

The report that follows is divided into seven sections. Section I places the research within the context of related studies on online learning. Section 2 discusses the research methodology and limitations. Section 3 describes the enrolled students, both within each course and within subgroups of interest. Section 4 presents pass rates overall and by course. Section 5 presents findings from contingency table analysis of associations between pass rates in the courses and student characteristics, student effort in the AOLE and use of online support. Section 6 introduces a statistical model of functional relationships that may help predict success and guide improvements in course design. Section 7 summarizes stakeholder perspectives obtained through surveys and interviews. Following the conclusion and recommendations, the report presents several attachments, including a detailed explanation of the methods used for the statistical analysis (Appendix A) and comprehensive tables of coefficients (Appendix B).

Section 1: Context

This study contributes to a growing body of research on academic achievement of students in online courses. In a recent meta-analysis of studies comparing online to in-person courses (2009), the U.S. Department of Education found that there was no significant difference between student achievement in online and traditional courses, although a small increase was noted for hybrid courses that combined the two methods. However, when these studies were further analyzed by at-risk student population demographics lower achievement was found in online courses (Jaggars & Bailey, 2010). This finding has been recently expanded by Xu and Jaggars (2013) in a study that investigated how different types of students—including groups that share characteristics found, particularly, among students from AOLE's partner high school—perform in online learning courses. Using course grade and course completion as dependent variables, Xu and Jaggars found that while all students did less well in online courses, some student groups were more negatively affected from taking courses in this mode. These students were males, younger students, students with lower levels academic skills, and African American students. The study also found that the negative impact of online learning was exacerbated when groups of students comprised of those who adapt least well to online learning study together. It should be noted that the study, based on research conducted across Washington State on 500,000 online and face-to-face course enrollments (and 41,000 students), did not distinguish between different types of online learning environments, faculty preparation, or support services available to students. The Xu and Jaggars study confirms prior findings in previous smaller studies, which also found that students from at-risk demographic groups and introductory courses had lower performance in online courses compared to other students, thereby exacerbating the well-documented achievement gap in higher education (Kaupp, 2012; Xiu & Smith, 2011, Terenzini & Pascarella, 1998).

An emerging discipline called "Learning Analytics" (Buckingham Shum, 2012; Ferguson, 2012; Siemens, 2011) has begun to investigate online course logfiles and other measures to better understand student learning in technology-mediated environments. One area of research has investigated how students engage with these environments and the relationship with their achievement and/or learning (Fritz, 2011; Macfadyen & Dawson, 2010; Ryabov, 2012; Whitmer, Fernandes, & Allen, 2012). Analyzing student use of these technologies allows us to peer inside the learning 'black box' and answer fundamental questions such as: Did students attempt to use the online course materials and still fail, or did they lack the study skills to consistently logon and use the materials? The present study draws on and contributes to this research by looking at how students use the online course materials and online support services and how that use is related to the likelihood that they passed the course.

Also relevant to AOLE is findings presented in “Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses” (Kizilcec, Piech & Schneider, 2013). Analyzing student engagement in ways that often parallel the research approach presented in this summary, the Stanford-based team of authors identifies different trajectories of student interaction with three computer science courses offered as Massive Open Online Courses (MOOCs). They conclude their study by proposing that course designers use information derived from learning analytics to modify the delivery and experiment with supplementary features such as “task lists and calendar features for staying organized" (p. 9). The research findings in this study help to identify the features and uses of them that have the strongest relationship with achievement and would therefore be good candidates for such an early warning system.

Section 2: Methods and Limitations

Research Purpose and Questions**:** The research project was designed to identify what worked and where improvements could be achieved and to generate as much information as possible that can inform future iterations of AOLE courses at SJSU Plus. The research therefore documented a wide range of perspectives on how the AOLE courses were working and grounded this information in a deep analysis of how students engaged with the MOOCs and the supports provided, their persistence and the outcomes they achieved.

The research questions that drove the analysis were:

4. Who engaged and who did not engage in a sustained way and who passed or failed in the remedial and introductory AOLE courses?
5. What student background and characteristics and use of online material and support services are associated with success and failure?
6. What do key stakeholders (students, faculty, online support services, coordinators, and leaders) tell us they have learned?

Data and Methodology: The research used the following four data sources:

1. Data on student background and characteristics from SJSU's student data base; this information was only available for matriculated students
2. Data on how students used the Udacity platform, including the course content and support services; this information was available for all students
3. Survey data revealing student perceptions about the AOLE experience and what they believed worked and could be improved; three surveys were conducted with response rates ranging from 33-39%.
4. Interviews and focus groups with key stakeholders, including a student focus group; entry and exit interviews with faculty and Udacity online support services providers; and interviews with the SJSU+ and Udacity coordinators and leaders.

The research team integrated data from the first three primary data sources (SJSU Admission and Enrollment, Udacity Instructional Resources and Support Services, and Survey Responses) into a single integrated file, using student ID and email address as a primary key. Inconsistencies and errors were eliminated. Only students completing the course were included in the analysis; students who withdrew or received an incomplete were not included as there is not a valid dependent variable (e.g. pass/fail) for these students. This limitation resulted in the elimination of 61 students from the analysis.

Appendix A provides a detailed description of the specific quantitative methods employed. In summary, the analysis was performed in two sections:

1. Contingency tables of student characteristics, effort expended, support use, and pass/fail to describe the sample and explore variables for possible significance.
2. A logistic regression model of pass/fail on student characteristics, use of Udacity course resources effort and use of support services. This method enables legitimate statistical examination of a binary outcome variable (pass/fail). The analysis includes additional variables created by transforming original data to enhance insights.

Findings generated from the investigation of different stakeholder perspectives (data source 4.) provided context and helped inform the direction of the quantitative analysis.

Limitations and Countermeasures: One limitation concerns the small number of students participating in the AOLE courses. Consequently the contingency tables are constrained to a small number of cells to protect against conclusions based on random phenomena and overgeneralization. A multiple-comparisons correction for determination of significance of association provides protection from false significant relationships (Type I errors) arising from creation of many contingency tables. Logistic regression models require relatively large sample sizes for validity, so they have been created only for the following populations: a) all students; b) matriculated and non-matriculated students in all courses; c) all students in each of the three course subjects.

The research was launched after the courses had begun. The SJSU Institutional Review Board (IRB) approvals were not obtained for several weeks, delaying implementation of the student entry survey until the 5th week of classes and complicating the task of obtaining permission to participate in the study for under-age students. While a major effort was made to increase participation in the survey research within this population, the result was disappointing (response rates of 32% for Survey 1; 34% for Survey 2, and 32% for Survey 3). The research team compared the survey participants to the entire student population and found significant differences. Most importantly, students who succeeded are over-represented among the survey respondents. However, while the survey responses are not representative of the overall population, useful information can be extracted from this data source as described in Section 7.

Another limitation resulted from the difficulty of obtaining Udacity instructional resources use and support service access data in a format that could be analyzed by statistical software. It took several months to receive this data in a flat-file format. While most of the data were provided by the end of the Spring 2013 semester, clarifications, corrections and data transformations had to be made for many weeks thereafter, including resolving accuracy questions that arose once the analysis of the Udacity platform data began. The challenges that the research team encountered in this project have been paralleled in other research projects involving open online course providers and educational researchers, including one in which members of the present research team are involved. The result for AOLE is that the research lagged behind the implementation whereas, ideally, it would be running alongside, providing just-in-time information about what works and where improvements can be made. However, these data have now been validated empirically and the research team is confident in the results provided in this report. The lessons learned and processes refined could also be applied easily to future courses with a much smaller level of effort and increased timeliness.

Section 3: Student Characteristics and Background

This section begins with a description of all students enrolled in the three courses, distinguishing between matriculated and non-matriculated participants for age and gender descriptors. Information about Pell-eligibility was only available for matriculated students. Information about URM status was available for matriculated students and for non-matriculated students who participated in the AOLE surveys and who responded to a question about race and ethnicity. The introduction of all students enrolled in the three courses is followed by a description of students in each of the three courses, again comparing the background and characteristics of matriculated and non-matriculated students where this data is available. The section concludes with a description of students from the partner high school. These students are presented as a group because they represent 20% of all participants and 45% of non-matriculated students, and because they have characteristics and pass rates that merit separate attention.

Students Across All Courses

Out of a total of 274 students enrolled in AOLE courses, 249 students remained in the sample after data cleaning, which included removal of students enrolled in multiple courses and of those with no course activity. In addition, 36 students were removed who withdrew from the course or received a final grade of Incomplete. This left 213 students for the deeper analysis. This data set will be referenced throughout this report as a “research data file”.

Table 1. Matriculated vs. Non-matriculated Students by Age, Gender, URM Status, and Pell Eligibility

	Count	Percent of Row			
	Matriculated	Non-matriculated	Total	Matriculated	Non-matriculated
< 18 years	0	39	39	0%	34%
18-24	94	36	130	96%	31%
25-30	3	12	15	3%	10%
31-45	1	22	23	1%	19%
46 +	0	6	6	0%	5%
Female	58	51	109	59%	44%
Male	40	64	104	41%	56%
URM	45	9	54	46%	24%
Non-URM	53	29	82	54%	76%
No Info Available		77	77		
Pell Eligible	38			39%	
Non-Pell Eligible	60			61%	
Total	98	115	213	100%	100%

Matriculated and non-matriculated students by key background and characteristics: Participants included 98 matriculated and 115 non-matriculated students for a total of 213 participants (46% matriculated vs. 54% non-matriculated). Table 1 compares the age distribution of non-matriculated and matriculated students. Matriculated students were predominantly between 18-24 years of age with only a few students older than this age category. Non-matriculated students, representing 54% of all student completers, were more diverse by age with roughly one-third of students less than 18, another one third between 18-24, and 34% were 25 or older. Among the non-matriculated students, participants from the partner high school contributed 33 of the 39 students under 18 years of age and 11 (31%) among those 18-24 years of age. In terms of **gender**, matriculated students comprised 59% female and 41% male students. Among non-matriculated students, more than half were male (56%) compared to 44% female. Among the matriculated students, 46% were **URM** while 54% were non-URM. As noted earlier URM status was not available for most of the non-matriculated students (77% of 115 non-matriculated students). A total of 39% of matriculated students were Pell eligible. Information about **Pell eligibility** was only available for matriculated students.

Math 6L Students

The course: Math 6L is a 5-unit course in remedial topics that prepares students for college-level instruction. It is described by one of the instructors as “fast-paced.” The course is “a review of topics from elementary and intermediate algebra. Completion of this course with a "credit" grade indicates satisfaction of the ELM exam.

Matriculated students who do not pass Math 6L during their first semester will be allowed to repeat it once. If they do not pass this course by the end of their first year, they will need to complete this course at a community college before they are eligible to enroll at SJSU (http://www.sjsu.edu/aars/advising/freshmen/courseprogress/). However, due to budget restrictions, MATH 6L has only been offered at SJSU in fall semesters since Fall 2009 so students that don’t pass Math 6L in the fall semester do not have the option of retaking it in the spring semester at SJSU. The Udacity-SJSU 6L course offered an alternative for students in this situation. All matriculated students in the course had failed Math 6L before.

Table 2: Math 6L Matriculated vs. Non-matriculated Students by Age, Gender, URM Status, and Pell Eligibility

	Count			Percent of Row	
	Matriculated	Non-matriculated	Total	Matriculated	Non-matriculated
< 18 years	0	10	10	0%	29%
18-24	47	11	58	100%	32%
25-30	0	2	2	0%	6%
31-45	0	8	8	0%	24%
46 +	0	3	3	0%	9%
Female	27	16	43	57%	47%
Male	20	18	38	43%	53%
URM	28	4	32	60%	44%
Non-URM	19	5	24	40%	56%
No Info Available	0	25	25		
Pell Eligible	18			38%	
Non-Pell Eligible	29			62%	
Total	47	34	81	100%	100%

Math 6L students by key background and characteristics: As described in Table 2, Math 6L included 47 (58% of all Math 6L students) matriculated and 34 non-matriculated students (42% of all Math 6L students) for a total of 81 students. All matriculated students in Math 6L were between 18 and 24 years of **age**. Among non-matriculated students, slightly less than one-third were 18 years or younger (29% of all non-matriculated students) while another one-third (32% of non-matriculated students) were 18-24 years of age. The remaining students included 11 participants (33%) who were 31 years or older, including 3 students over the age of 46. More than half of the matriculated 6L students were female (57%), while for non-matriculated students more than half (53%) were male. 60% of matriculated students were from URM groups. 38% of matriculated students were Pell eligible.

Math 8 Students

The course: Math 8, College Algebra, reviews basic algebra, including complex numbers, functions, graphs, polynomials, inverse functions, exponential and logarithmic functions. Course entry requirements are satisfaction of ELM requirement.

Table 3. Math 8 Matriculated vs. Non-matriculated Students by Age, Gender, URM Status, and Pell Eligibility

	Count			Percent of Row	
	Matriculated	Non-matriculated	Total	Matriculated	Non-matriculated
< 18 years	0	18	18	0	43%
18-24	15	13	28	94%	31%
25-30	1	2	3	6%	5%
31-45	0	8	8	0	19%
46 +	0	1	1	0	2%
Female	6	15	21	37.5%	36%
Male	10	27	37	62.5%	64%
URM	5	0	5	31%	0%
Non-URM	11	3	14	69%	100%
No Info Available	0	39	39		
Pell Eligible	9			56%	
Non-Pell Eligible	7			44%	
Total	16	42	58	100%	100%

Math 8 students by key background and characteristics: As described in Table 3**,** Math 8 included 16 matriculated and 42 non-matriculated students for a total of 58 participants. Among matriculated students, all but one (94%) were 18-24 years of age. The non-matriculated students included 18 students (43% of non-matriculated students)

18 years of age or younger and 13 students (31% of non-matriculated students) between the ages of 18 and 24. The non-matriculated students also included 8 between 31 and 45 years of age. The remaining 2 non-matriculated students were between 25 and 30, while 1 student was more than 46 years of age. The gender distribution was similar for matriculated and non-matriculated students. Among the former students, 37.5% were female; among the latter group 36% were female. Matriculated students included 5 URM (31%) and 14 non-URM (69%) students. More than half of the matriculated students were Pell Eligible (56%).

Stat 95 Students

The Course: Stat 95, Elementary Statistics is an introductory course in statistics intended for majors in education, nursing, personnel administration, psychology, social service and sociology, and psychology. Course entry requirements are satisfaction of ELM requirement and two years of high school algebra.

Table 4. Stat 95 Matriculated vs. Non-matriculated Students by Age, Gender, URM Status, and Pell Eligibility

	Count			Percent of Row	
	Matriculated	Non-matriculated	Total	Matriculated	Non-matriculated
< 18 years	0	11	11	0%	28%
18-24	32	12	44	91%	31%
25-30	2	8	10	6%	21%
31-45	0	6	6	0%	15%
46 +	1	2	3	3%	5%
Total	35	39	74	100%	100%
Female	25	20	45	71%	51%
Male	10	19	29	29%	49%
URM	12	5	17	34%	19%
Non-URM	23	21	44	66%	81%
No Info Available	0	13	13		
Pell Eligible	11			31%	
Non-Pell Eligible	24			69%	
Total	35	39	74	100%	100%

Stat 95 Students by key background and characteristics: As described in Table 4, Stat 95 included 35 matriculated and 39 non-matriculated students for a total of 74 participants. Among matriculated students 32 (91%) were between the **ages** of 18 and

24, while three students (9%) were older. Non-matriculated students included 11 participants under the age of 18 (28%), 12 students between the ages of 18 and 24 (31%) and 8 students between 25 and 30 years of age. A total of 8 (20%) non-matriculated students were over the age of 31, including 2 (5%) over 46 years of age. In terms of **gender**, 71% of matriculated students were female (25 students), while among non-matriculated students just over half were female (20 students). Among matriculated students, just over one-third (34%) were URM. While the URM-status is not known for one-third of the non-matriculated students, the relatively high survey response rate among non-matriculated Stat 95 students means that URM-status is known for two-thirds of these students who answered a survey question about race and ethnicity. Overall, among the non-matriculated students, 54% were non-URM and 13% URM. Slightly less than one-third of matriculated students (31%) were Pell Eligible.

High School Partner Students

Since high school participation is of special interest and since a relatively large percentage of participants were from the main partner high school, these participants are described as a group by their enrollment in the three courses.

The partner high school is located in the San Francisco Bay Area. The *US News* report on high schools in California states that approximately 97% of students at this high school are from underrepresented groups and that 79% are economically disadvantaged.

The distribution of students from the partner high school across the three courses by background and characteristics is described in Table 5 below.

Table 5: Students from Partner High School by Course Enrollment, Age and Gender

Total # of high school students from partner school	Math 6L	Math 8	Stat 95	Total
	8	24	12	44
Age	5 (62.5%) were 15 3 (37.5%) were 16	5 (21%) were 16 10 (42%) were 17 8 (33%) were 18 1 (4%) was 19	1 was 16 (8%) 9 were 17 (75%) 2 (17%) were 18	
Male	5 (62%)	11 (46%)	7 (58%)	23
Female	3 (38%)	13 (54%)	5 (42%)	21

Students Who Withdrew or Received an Incomplete

The 36 students who withdrew or received an incomplete were primarily non-matriculated students (32, 89%), divided among the courses (Math 6: 12 students; Math

8: 19 students; Stat 95: 5 students). 16 of the students (44%) were from the partner high school.

By demographic criteria, 44% of these students were female (16 students) and 56% were male (20 students). Half of the students (50%) were 18-24 years of age. The second largest group by age was over 46 years of age (7 students or 19% of those who withdrew or did not complete). Four students were under 18 years of age and 4 were between 25 and 30 years old. The remaining 3 students were 31-45 years of age. All 4 matriculated students who did not complete were Pell-eligible.

Section 4: General Education Pass Rates (C or Better)

The General Education (GE) Pass Rates Across Courses for Matriculated and Non-Matriculated Students: Pass rates used in this report refer to the GE pass rate of C or better required for the four GE "basic skills" courses: written communication, mathematical concepts, critical thinking and oral communication. As indicated in Table 6, the overall pass rate across the three courses was 33%, but the pass rate varied widely within sub-groups and across courses. Overall, matriculated students enjoyed higher pass rates than non-matriculated students (42% vs. 26%). The difference between the two groups was particularly pronounced in Math 8 where 50% of matriculated and only 12% of non-matriculated students passed. In Math 6L the difference between the pass rates for matriculated and non-matriculated students was 12%. In Stat 95, the difference was 5.6%.

Table 6. Student Pass Rates by Course & Matriculation

Course	Count			Percent of Row	
	Pass/credit	**Fail**	**Total**	**Pass/credit**	**Fail**
MATH 6L Matriculated	14	33	47	29.8%	70.2%
MATH6L Non-matriculated	6	28	34	17.6%	82.4%
MATH 8 Matriculated	8	8	16	50.0%	50.0%
MATH 8 Non-matriculated	5	37	42	11.9%	88.1%
STAT 95 Matriculated	19	16	35	54.3%	45.7%
STAT 95 Non-matriculated	19	20	39	48.7%	51.3%
Total	**71**	**142**	**213**	**33.3%**	**66.7%**

Table 7 below was prepared by SJSU Institutional Effectiveness and Analytics and compares grade distribution in face-to-face courses offered during previous semesters and during Spring 2013 for Math 8 and Stat 95 to grade distribution in the AOLE courses based on the actual course enrollment. Please note that the discrepancy between the 2013 pass and credit rates presented in the actual course enrollment and in Table 6 above (from the research data file) results from the removal of 14 students from the group of participants included in the AOLE research due to the fact that 6 had enrolled in more than one AOLE course and 8 had no Udacity ID.

For Math 6L, no face-to-face course has been offered since Spring 2009 due to budget cuts. As a result, the grade distribution in Math 6L for Spring 2004-2009 is presented for comparison purposes with grades reported only for students who took the course for the second time. The reason for this selection is that all Matriculated 6L students had failed the course before and were taking it in the AOLE format for their second time.

As the research references presented earlier in this report suggest, students tend to perform better in face-to-face than in on-line courses. This pattern can be seen in Table 7 which juxtaposes student pass (credit) and no-pass (no credit) rates in the face-to-face against first AOLE courses.

In Math 6L, in Spring semesters between 2004 and 2009, the percent of students who took the course for the second time and received credit ranged from a high of 50% in Spring 2006 to 34% in Spring 2008. In the Spring 2013 AOLE Math 6L, all matriculated students were taking the course for the second time with 30% receiving credit for the course.

In Math 8, the percent of students in face-to-face courses who completed the course with a C or better between Spring 2010 and Spring 2013 ranged from a low of 52% in Fall 2010 to a high of 76% in Spring 2013. In the Spring 2013 AOLE version of Math 8, 50% of matriculated students received a C or better.

In Stat 95, the percent of students in face-to-face courses who completed the course with a C or better between Spring 2010 and Spring 2013 ranged a low of 71% in Fall 2010 to a high of 80% in Fall 2012. In the Spring 2013 AOLE version of the course, 54% of matriculated students received a grade of C or better.

Table 7. Grade Distribution and Historical Comparison for Math 6L, Math 8, and Stat 95: Face-to-Face vs. AOLE

Math 6L: Entry Level Math	Face-to-Face (F2F) for 2nd Time Takers Only							AOLE Spring 13	
	Spring 04	Spring 05	Spring 06	Spring 07	Spring 08	Spring 09		Matric	Non-matric
% Retention	**97.4%**	**100.0%**	**98.5%**	**100.0%**	**100.0%**	**100.0%**		**95.9%**	**74.0%**
% Credit	40.5%	54.3%	50.0%	44.8%	34.4%	48.9%		29.8%	16.2%
% No Credit	59.5%	45.7%	50.0%	55.2%	65.6%	51.1%		70.2%	83.8%

Math 8: College Algebra	Face to Face (F2F)						Spring 13 F2F	AOLE Spring 13	
	Spring 10	Fall 10	Spring 11	Fall 11	Spring 12	Fall 12		Matric	Non-matric
% Retention	**97%**	**100%**	**98%**	**99%**	**99%**	**99%**	**100%**	**89%**	**70%**
% C or Better	**58%**	**52%**	**64%**	**74%**	**68%**	**73%**	**76%**	**50%**	**17%**
% C- or Lower	42%	48%	36%	26%	32%	27%	24%	50%	83%

Note: To meet the SJSU GE requirements, C or better is required.

Stat 95: Elementary Statistics	Face to Face (F2F)						Spring 13 F2F	AOLE Spring 13	
	Spring 10	Fall 10	Spring 11	Fall 11	Spring 12	Fall 12		Matric	Non-matric
% Retention	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**100.0%**	**99.3%**	**97.3%**	**84.9%**
% C or Better	**72.8%**	**70.8%**	**78.3%**	**77.6%**	**78.0%**	**80.4%**	**74.6%**	**54.3%**	**46.7%**
% C- or Lower	27.2%	29.2%	21.7%	22.4%	22.0%	19.6%	25.4%	45.7%	53.3%

Note: To meet the SJSU GE requirements, C or better is required.

Section 5: Contingency Table Analysis

This section presents summary findings from a contingency table analysis of relationships between the dependent variable (course pass/fail) and three types of independent variables:

A. Course subjects and matriculation status
B. Student engagement with online content
C. Student use of support services

This analysis demonstrates the relationship between one variable, such as enrollment in one of the courses, and the likelihood that the student passed or failed. In order to create these tables, variables were transformed into binary categories. In the case of student use of online content and support services, these were usually based on a dichotomous variable derived from the average value (e.g., whether a student was above or below the average for that variable). These tables create results that are easy to interpret for a general audience. The Chi^2 measure of association between the variables is also provided for more statistically-oriented readers to enable comparison of the relationship strengths between tables.

A total of 35 relationships between the independent and dependent variables were tested in the analysis. The following tables contain some of the more important results, either through a high relationship or an unexpectedly low relationship. The results of these tables were also used to inform the regression analysis and modeling phase. Significance has been adjusted properly for multiple comparisons (see Appendix A).

This section begins with a summary table of all tabular results and their individual-table significance, along with the adjustment in significance for the fact that so many tables will produce an expected number of spurious associations purely by chance.

Pass rates are in the Row % rows of each table. Individual-table significance uses Pearson chi-square with degrees of freedom in ().

A. Course Subjects and Matriculation Status

This first table demonstrates the pass rates between the different courses in AOLE. There are strikingly different rates in the courses, from a low in Math 8 (only 22% passing), to a high in the Statistics course (with 51% passing). This may be due to different student demographics enrolling in these courses, as indicated in the previous section, or that students with more higher education experience are more likely to be successful in online courses.

Table 8. Course vs. Pass Rate (All Students)

Course	No Pass	Passed	Total
MATH 6	61	20	81
Row %	75.31	24.69	100
MATH 8	45	13	58
Row %	77.59	22.41	100
STAT 95	36	38	74
Row %	48.65	51.35	100
Total	142	71	213
Row %	66.67	33.33	100

Pearson chi^2 (2) = 16.6451 p = 0.000

Where Table 8 demonstrates the differences between pass rate by course for all students, Tables 9 and 10 show how pass rate by course varied by whether the student was matriculated at SJSU or not; it is clear from these findings that SJSU students had much higher pass rates (42%) than non-matriculated students (26%). This finding was particularly true in Math 8, which was the course with the highest enrollment of students from the partner high school.

Table 9. Course vs. Pass Rate (Matriculated Students)

Course	No Pass	Passed	Total
MATH 6L01	33	14	47
Row %	70.21	29.79	100
MATH 801	8	8	16
Row %	50	50	100
STAT 9501	16	19	35
Row %	45.71	54.29	100
Total	57	41	98
Row %	58.16	41.84	100
Pearson chi^2 (2) = 5.4716 p = 0.065			

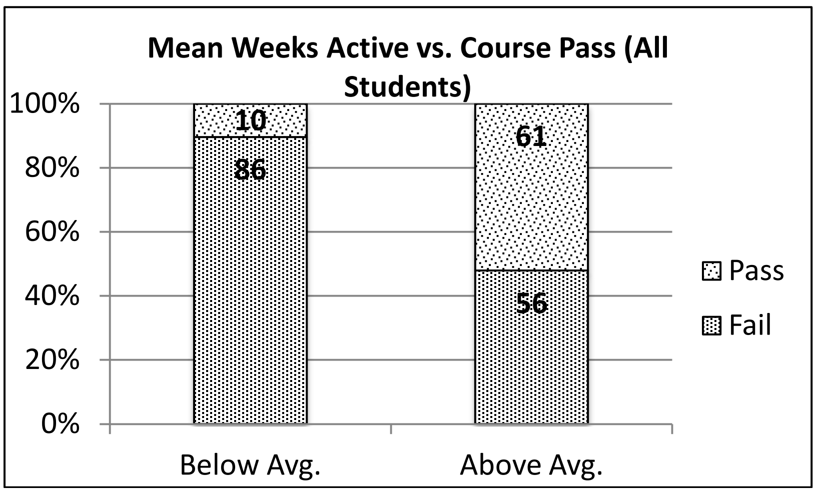
Table 10. Course vs. Pass Rate (Non-Matriculated Students)

Course	No Pass	Passed	Total
MATH 6L03	28	6	34
Row %	82.35	17.65	100
MATH 802	37	5	42
Row %	88.1	11.9	100
STAT 9502	20	19	39
Row %	51.28	48.72	100
Total	85	30	115
Row %	73.91	26.09	100
Col. %	100	100	100
Pearson chi^2 (2) = 15.9965 p = 0.000			

B. Student engagement with online content

The AOLE courses offered all lectures, problem sets, and interactive activities through the Udacity platform. A composite measure of time spent on these materials throughout the course was developed by creating a count of the number of weeks during which each student logged in for more than 30 minutes during one or more sessions. The mean value for all students was calculated for this variable, and each student's record was compared to determine if they were above or at/below this mean value. In Figure 1, it is clear that students who had less than the mean number of weeks active were very unlikely to pass the course. Students who were at or above the mean value were slightly more likely to pass the course.

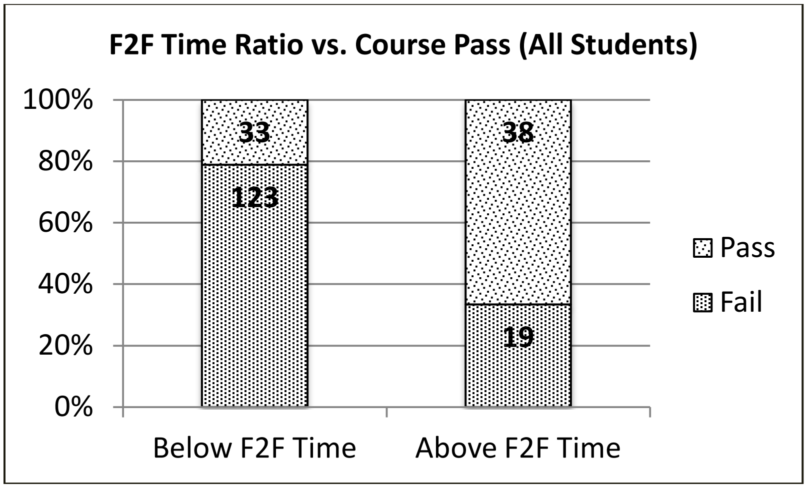
Figure 1. Mean Weeks Active vs. Course Pass (All Students)



As an alternative measure of time spent on course materials, a ratio was created that compared the time spent in the online course to the in-class time required for the equivalent course taught face-to-face (F2F). It standardizes effort for the time-intensity of each online course, and is defined as student time logged in/(course units x class-time weeks of the corresponding F2F course), where class-time weeks = SJSU Plus Weeks. This ignores F2F course homework time, a fact that boosts the ratio, but harmlessly across courses as it is proportional. When it equals 1.0 or greater the student has spent at least as much time logged in as the corresponding F2F course's classroom time.

In the ratio in Figure 2, it can be seen that students who spent less time in the online course compared to F2F were much less likely to pass the course, and the converse is true.

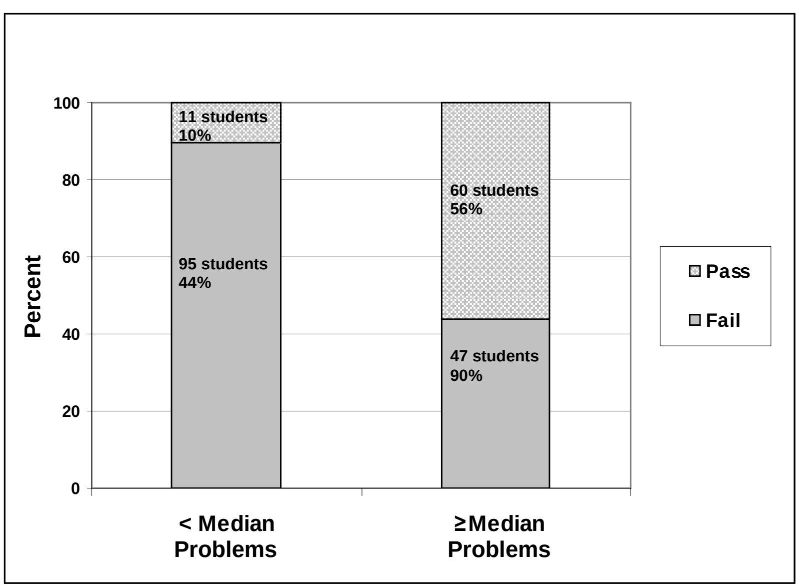
Figure 2. Face-to-Face Time Ratio vs. Course Pass (All Students)
= Total Student time logged in/Total class time of corresponding F2F course



To examine the value of specific activities within the course, the number of online problems attempted was calculated for each student. Figure 3 demonstrates the relationship between the number of online problems attempted by students and the likelihood that they would pass a course. Students were divided into whether they completed more than the average number of problems or less than the average number of problems, and that variable is plotted below by whether that student passed the course or not.

If a student completed less than the median number of problems, they were very likely to fail the course. By contrast, if a student was above the median value they were much more likely to pass the course, indicating that this was a high yield activity for both sets of students.

Figure 3: Problems Done vs. Course Pass (All Students)



C. Student Use of Support Services

In Table 11. significant positive interaction was found between early engagement with support services and pass rates. Students who had not engaged with online support services by week 5 were less likely to pass.

Table 11. All Student Contact with Online Support ≤ Week 5

Contact w. Online Support <=Week 5	No Pass	Passed	Total
No contact <=Week 5	95	32	127
Row %	74.8	25.2	100
Contact <=Week 5	47	39	86
Row %	54.65	45.35	100
Total	142	71	213
Row %	66.67	33.33	100
Col. %	100	100	100
Pearson chi^2 (1) = 9.3707 p = 0.002			

Additional tables are available upon request.

Section 6: Statistical Model

The contingency tables provide direct examination of student counts and the significance of associations between individual potential predictor variables and pass/fail. Many of these variables are significant when standing alone, but critical insight has been gained by analyzing them together to form contextual models of their relationships to student outcomes. Analyzing variables together in a model usually makes a number of stand-alone associations disappear when their effects are shown to be reflections of other variables or variable combinations. The goal for the modeling approach is to test pass/fail associations together and to reveal the strengths, directions and forms of significant relationships, with emphasis on model validity.

<u>Significant Predictor Variables</u>

The primary conclusion from the model, in terms of importance to passing the course, is that measures of student effort eclipse all other variables examined in the study, including demographic descriptions of the students, course subject matter and student use of support services. Although support services may be important, they are

overshadowed in the current models by students' degree of effort devoted to their courses. This overall finding may indicate that accountable activity by students—problem sets for example—may be a key ingredient of student success in this environment.

Note that some alternative measures of student effort may have achieved strong coefficients in less optimal models, but they could not be included in the examined-variables mix because they were redundant statistically. Statistically redundant variables play havoc with a model's ability to arrive at an optimal solution. Also, selected variables were practically unrelated to measures not defined by effort, allowing greater flexibility to achieve meaningful results than possible with excluded variables that demonstrate stronger relationships with measures other than effort.

Table 12 shows the 5 variables that are highly significantly related to pass/fail for 6 different groupings of the students. Four variables are measures of effort and the other variable measures use of support services. Their functional behaviors relative to pass/fail follows the table.

Table 12. Variables with Significant Contextual Associations with Pass/Fail

Variable	Variable Classification	Student Group(s) for which Variable is Significant	Expected Improvement in Odds of Pass over Fail per Unit of Variable Added	Strength of Net Effect*	Confidence Effect is not Random**
Problems Done	Degree of effort. When their levels are high they are indicators of early effort & persistence, which may proxy for required assignments.	All, Matriculated, Math 6L, Math 8	30.5% –36.7% per problem done	Strong	97.1% – 99.9%
Video Time		All, Matriculated, Non-matriculated, Stat 95	0.01% – 0.08% per video minute	Extremely strong for Stat 95, strong otherwise	99.1% - 99.9%
Weeks Active for at Least ½ Hour		Non-matriculated	34.2% per week	Strong	98.6%
Number of Sessions Logged In		Math 8	3.3% per session	Strong	98.1 %
Support Staff Characters Typed	Use of support services	Non-matriculated	-0.02% per character	Negative	98.0%

* Net of other variables' effects

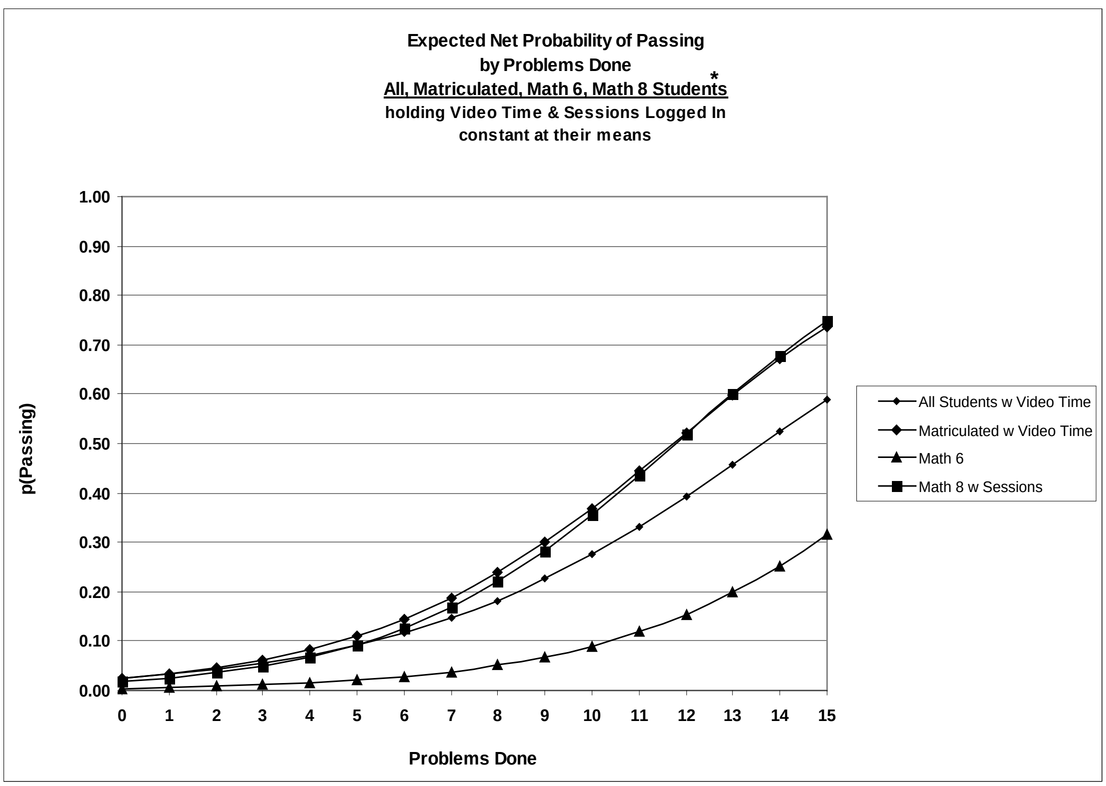
** Not the same as proof the hypothesis is true—that the effect exists unequivocally— (the error of confirming the consequent). But these are very high degrees of confidence that these models are significant and that the variables have an effect. See Blalock (1972) pp. 112-114

How the Predictors Behave by Student Grouping

All of the groupings exhibit significant relationships with one or more effort variables. The two significant predictors for the most groupings are Problems Solved and Video Time. Both variables exhibit strong positive relationships with passing; Video Time is especially strong for Stat 95 students. The functional relationships are best described with the aid of graphs.

Note on the terms “Expected” and “Net:” “Expected” refers to the fact that the modeled effects are best estimates given the data, representing what would be expected if the circumstances were repeated many times. “Net” applies to relationships that work with other independent variables in the same model. The graph shows the effect of the independent variable, e.g., Problems Done, on pass/fail as Problems Done increases, “net” of the effects of the other independent variables (if there are any), which are held fixed. In this report’s graphs the other variables are explicitly held constant at each of their means for the specific student group.

Figure 4. Expected Net Increase in Probability of Passing with Increase in Problems Done



* All variables are listed by model in the appendix. The graphed variables show effect net of the other variables’ effects. A model’s significant variables not included in a graph are being held constant at their mean values for each student group.

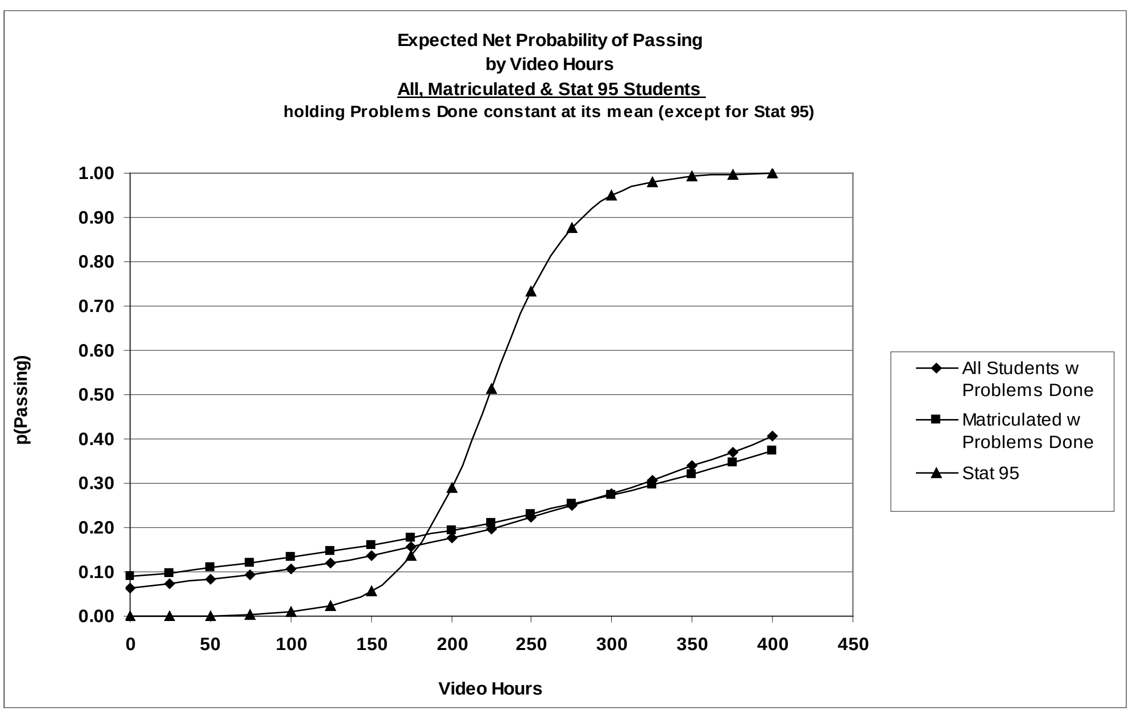
Problems Done is the total of problem sets completed by a student. It has a strong net positive relationship with the probability of passing for all students and matriculated students in conjunction with Video Time; Math 6L with Problems Done only; and Math 8 with Number of Sessions; with the effect of their companion independent variables held constant at their means. Its nonlinearity means that completing the first few problems in a course has little positive effect on passing, but each successive problem completed improves a student's chances to a greater extent than the previous problem completed. For example, when looking at all students together, completing problem 4 improves the chance of passing by 1.6 percent over completing 3 problems, while completing problem 15 improves it by 6.6 percent over completing only 14.

As can be seen in the graph, the net effect of Problems Done on the chance of passing is greatest as an overall relationship with students, and weakest for Math 6, but still important for those students.

Video Time

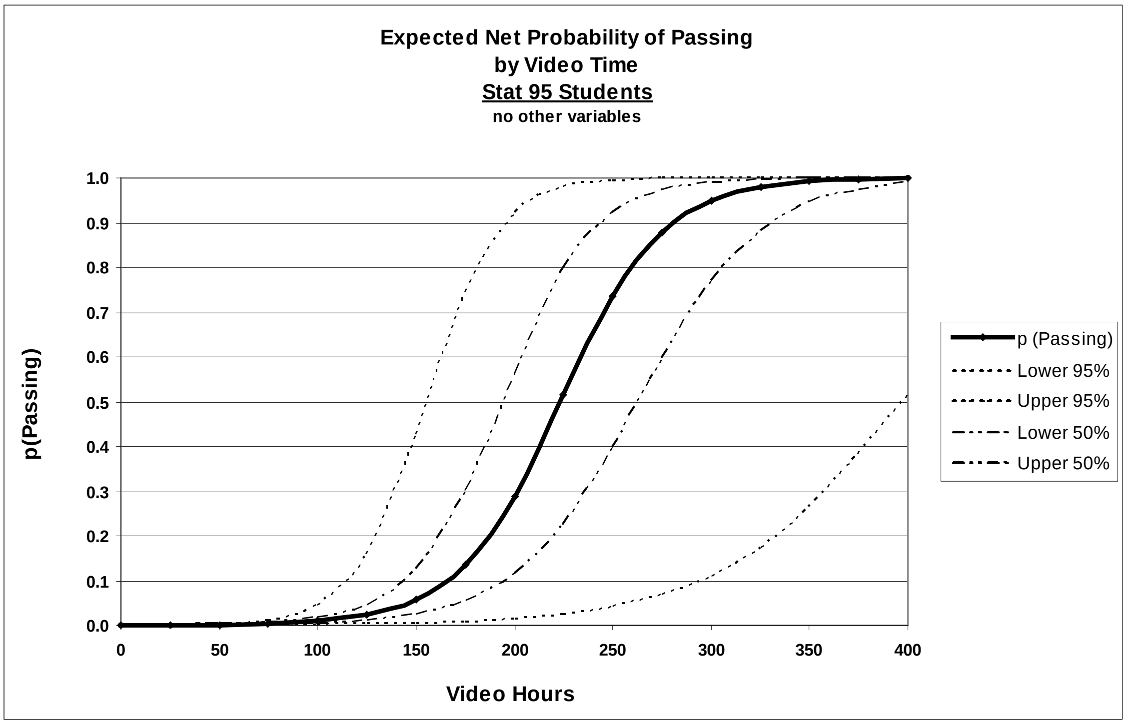
Video Time is how much time a student spent watching course lectures. Video Time was collected in minutes; its results are expressed here in hours for convenience of interpretation.

Figure 5. Expected Net Increase in Probability of Passing with Increase in Video Time



Expected Video Time net effect on pass rates is strong for all students together and for matriculated students, but less strong than Problems Done for these groups. The relationship with Stat 95 stands alone—no other variables were needed to create a valid model for that group. It is the strongest in the set of selected independent relationships. Also note that the Stat 95 students completed over 50 percent of their problems on average, while the Math 6L and Math 8 students completed less than 25 percent and less than 23 percent respectively.

Figure 6. Stat 95 Students: Expected Net Increase in Probability of Passing with Increase in Video Time



Video Time is the only significant predictor for the Stat 95 student group when the potential variable set was considered in the model context. The graph shows the expected relationship along with its 50% and 95% confidence intervals, meaning that the chance of passing is expected to lie between the two inner dashed curves one-half of the time, and within the two outer curves 95% of the time for repeated samples from a population with similar characteristics. The variable is highly significant (99.999 confidence) and predicts 89.5% of passes in the sample correctly (compared to 48.6% if one guessed based on the known margins).

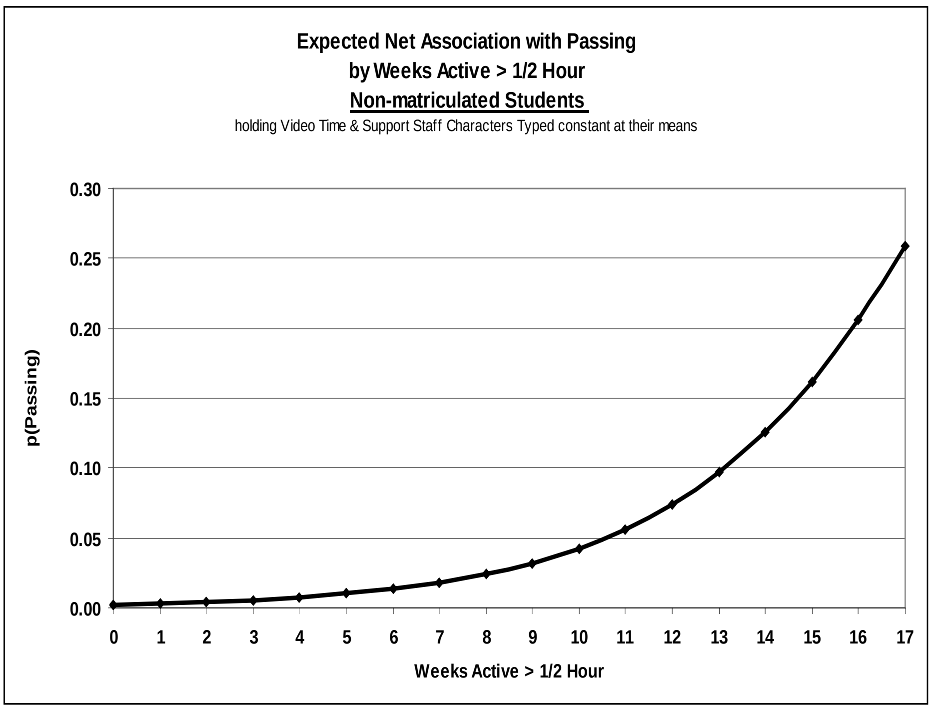
An example of the effect strength in Stat 95 is illustrated in Table 13. These numbers cannot be taken literally, because Video Time, as the only significant variable for this group, is not net of any other independents, and because the Stat 95 sample is fairly small for a logistic regression; but the general indication is clear.

Table 13. Example of Estimated Effect of Video Time Effort on Passing Stat 95

Increase in Video Time Hours	Expected Change in Chance of Passing Stat 95
100 – 200	28%
200 - 300	66%

Weeks Active for at Least ½ Hour and Number of Sessions Logged In

Figure 7. Non-matriculated Students: Expected Net Increase in Probability of Passing with Increase in Weeks Active ≥ ½ Hour per Week

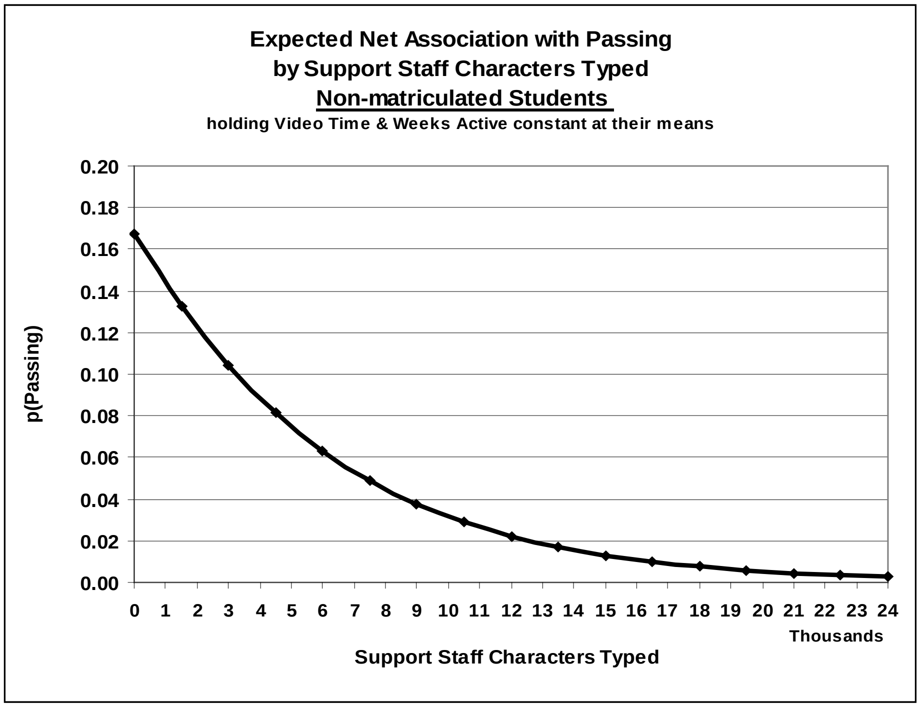


Weeks Active for at Least ½ Hour is another measure of effort and early and persistent effort as number of weeks increases. Its expected net effect is strong, but the mildest of the effort variables.

Figures 4 through 7 point to the same conclusion: student effort is the predictor of success, in magnitude, persistence and earliness of onset (by definition when students have higher numbers of problems done), as measured by: problems done, video time, weeks of greater than minimum time of effort, and number of sessions logged in.

Support: Support Staff Characters Typed

Figure 8. Observed Net Association between Support Characters Typed and Passing



Only one measure of support effect on student performance emerged as significant within the context of other variables, and only for non-matriculated students as a group: Support Staff Characters Typed. The other support measure included after multicollinearity testing is Meaningful Phone Calls, which is phone outreach by Udacity mentors to students falling behind and at risk of failing ("meaningful" refers to the fact that contact was made). Its best significance for all six student groups is 0.213, and in all procedures its odds ratios straddled 1.0 (positive and negative) by a wide margin. The Support Staff Characters Typed odds ratios also straddled 1.0 in the five models in which it was not significant (see Appendix B, Table B.3.)

Support Staff Characters Typed to non-matriculated students is consistently negative throughout its 95% confidence-limit range for non-matriculated students. The association exists empirically, but is relatively weak compared to the other significant independent variables. It is negative possibly because the most support was requested by those falling most behind, or because many students were unaware of support availability. It seems prudent to assume that the figure shows an association rather than a functional relationship—that specific circumstances in the Spring courses caused it to be negative, rather than to conclude that adding support causes poor performance. This is corroborated by consistent responses to the three surveys.

Non-significant Associations

Non-significant results are also important to perspective. Student characteristics including gender, being from the partner high school, URM status and Pell eligibility are not sufficiently significant within the context of other variables to be included in a parsimonious model that can explain most of the pass vs. fail outcomes. Small effects within context support the observation that student effort has the dominant effect on outcomes.

Model Diagnostics Summary

We examined 27 variables as potential predictors for pass/fail, including several created from the original set, (for example, weeks active more than a minimum time and hours logged-in relative to expected hours in class for the corresponding in-person course). Nine variables were eliminated from consideration as they are redundant statistically (see methodology appendix for these tests). Six logistic-regression models examined the remaining 18 potential predictors, finding 5 variables to be both significant statistically and meaningful in substance. 213 students are modeled (students who withdrew or received an incomplete are not in this part of the modeling analysis, as they have no pass/fail performance). Relationships are summarized as expected odds ratios of passing to failing, and of expected probabilities of passing. A complete table of model coefficients including notes on their meaning is in Appendix B.

The six models performed well and are parsimonious (not over-specified). There is confidence in results as all models' ability to reject the null hypothesis is greater than 99.9%. Ordinary least squares (OLS) R^2 is not valid to assess model strength in logistic regression. "Pseudo-R^2" measures designed for logistic regression are reported in Appendix B. Comparisons with OLS R^2 are difficult. But more intuitive metrics are available. For example, combined pass and fail correct-prediction rates range from 75% to 90% (all of which are better than the corresponding null models that predict the maximum category).

Full detail on contingency table and model methods and for all model coefficients may be found in Appendices A and B, including: approach, assumptions, multicollinearity tests, graphing equations, significant and insignificant coefficients, and multiple comparisons test.

Section 7: Stakeholder Input

Stakeholder input was collected through the following activities

- three student surveys conducted at the beginning, middle and end of the semester

- a focus group with students from the partner high school conducted at the end of the semester
- two focus groups with Udacity staff providing technical online support conducted at the beginning and end of the semester
- interviews with faculty conducted during the early part of the semester and at the conclusion of the semester
- interviews with coordinators and members of the leadership team from Udacity and SJSU conducted at the end of the semester

Additional interviews will be conducted during Fall 2013 and Spring 2014 with faculty and students to assess what they believe was learned when looking back at their Spring 2013 experience. These interviews will also enable faculty to describe how the lessons from the first iteration of AOLE courses were used to guide the development and implementation of the Summer 2013 and Spring 2014 AOLE courses.

This section presents highlights of information gathered from each different stakeholder group, beginning with the students.

Student Perspectives

Student perspectives were collected through three surveys and one focus group with students at the partner high school. While more than half of the students (110) included in the research group participated in at least one survey, the survey population was not representative of the entire student population studied by the research team. Most importantly, successful students were overrepresented among the survey population and almost no students from the partner high school completed the surveys (see Method Section). Nevertheless important insights can be gained from studying the survey responses.

The focus group, conducted at the end of the semester, included nine students. Two students were enrolled in Math 8 while the others were enrolled in Stat 95. While this group of students represents less than one-fourth of all students from the partner high schools, their voices provide a perspective not available elsewhere for the purpose of this research.

The surveys and the focus group protocols placed considerable emphasis on questions related to students' awareness of and engagement and satisfaction with the support services. As a result, the survey and focus group findings may help shed light on why use of support services did not rise to the top among the variables examined to explain student outcomes (see Section 6).

Most importantly, many students lacked experience with online education environments and did not fully understand what support services were available to them. In looking back at their experience, many students indicated that if given another chance to take the courses they would use the support services more.

As an illustration, in Survey 1—conducted during the fifth week of instruction—students were asked how well they understood the range of support services available to them. Slightly less than half of the students indicated that they "partially understood" the support services.

When asked in Survey 3 what they would do differently if they could start over, one of the top choices identified by students in Survey 3 was to use online support services more (51% across the board). Among all the groups of matriculated and non-matriculated students in the three courses, matriculated students in Math 6L (64%) and in Stat 95 (60%) were most likely to point to more online support as a change they would make if given a second chance.

Students' lack of online experience was revealed in Survey 1 where 39% of survey respondents reported not having taken any online course before, including 57% of matriculated students. This information should be considered in light of the fact that successful students were disproportionately represented among survey respondents. Particularly among students at the partner high school—a group that represented 20% of the students included in the research—previous experience with online instruction may well have been less than what was the case among the survey population. In addition to the fact that students from the partner high school were younger than their peers, the Udacity online support providers reported that many of the students did not have computers at home.

In the focus group, students described some of the challenges they had encountered navigating the online learning environment. These challenges included engaging with both the SJSU and the Udacity websites, a complication that was mentioned as well by many survey participants. Students in the focus group and in the surveys were unclear about what information to access and provide to each of the two websites, for example where to find information about grades. Additional complications arose from the fact that some e-mail communication about course requirements did not reach some of the students. "Sometimes I don't receive e-mails that other students do," one student commented in the focus group. She added: "Before an exam [the email] would go to my spam and then the day before the exam they told me I never registered because I did not know about it. I could not take the exam." Another student commented: "I got e-mails saying log into your SJSU ID and I didn't have one. I tried to register for something but was not able to do it. The e-mail said this is only for Udacity students."

Many students in the focus group and in the surveys also explained how difficult it had been for them to get used to taking exams online both in terms of making the necessary arrangements to register and in terms of feeling strange and in some cases uncomfortable about the online proctoring. One student in the focus group noted:

“They told us we have to register 72 hours in advance and we got the e-mail [with the announcement about the 72 hour requirement] 9 hours before the exam."

Most of the focus group students enrolled in Stat 95 also explained that they were taken by surprise by "Stat Crunch" assignments which several thought had been added "on top of lessons." Again the element of surprise may have resulted from the students’ difficulty in navigating the two different websites. "We did not know that we had the stat crunch. We didn't know until they e-mailed us the grades and they were really low."

In response to a question about what suggestions they had for improvements students in the focus group said that they would like to have "all the work in one website so we don't have to go to multiple links." Students also expressed the desire for more time, particularly as the lessons got harder later in the semester. "Now [late in the semester] lessons are getting harder, but we still have the same amount of time [to complete the assignments." Another student commented: "[the longer lessons] make me feel I don't want to do it. The lessons are really long and I feel I won't have time to finish it all-- it seems like a lot."

In the surveys and in the focus group, many students expressed the desire for more face-to-face opportunities with faculty and other students although, as several instructors noted (see below), very few students took advantage of the opportunity to engage in video conferencing with their instructors. Extrapolating from the survey responses, the reason may well be that many students were not aware that this source of help was available to them. Another reason may have been that many students are reluctant to engage directly with faculty. As Barbara Cox found in her study of how faculty and students communicate (Cox, 2012), there is, among under-prepared students in particular, often a great fear of engaging with faculty. Hence, office hours tend to be an extremely underutilized resource especially among students who might be able to benefit the most from this kind of one-on-one engagement with their instructors.

Yet, one of the top-rated changes students identified in Survey 3 was "more help with course content". In this area, there was almost no difference between survey responses from matriculated and non-matriculated students with 80% of respondents from both groups rating "more help with content" as a "very important" or "important" change they would like to see Udacity and SJSU make.

This finding, when corroborated with input from faculty (see below) points to an area that may require additional attention. Because while students did not use the opportunity to video-conference with instructors, many students e-mailed their professor, but not with questions about content. Instead, student email communications focused, across the three courses, on questions related to course requirements, assignments and other technical or process-related issues.

The statistical model pointed to the critical importance of effort. In Survey 3 students indicated that they recognized the need for a sustained effort and the danger of falling behind. In fact, when asked what they would change if starting the semester over knowing what they know now, one of the top choices was to "make sure I don't fall behind." Almost two-thirds of survey respondents (65%) in Survey 3 pointed to this change, including 82% of matriculated students in Math 6L and 75% of matriculated students in Math 8. In Stat 95, where students were less likely to fall seriously behind because of stricter adherence to deadlines (see below), 60% of both matriculated and non-matriculated students identified "not falling behind" as a change they would make.

Students were also asked in an open-ended question about ideas for how both the overall course and the support services could be improved. The question about overall improvements was posed in the first and last surveys with the first survey yielding much more comprehensive narrative responses than students provided in the last survey. Overall 49 of 66 survey respondents submitted narrative responses with the most frequently made suggestion (mentioned by 8 students) concerning the introduction of the support services so that more students would know that they were available and how they could be accessed. One student suggested: "Let them be known and give an introduction to them as well as their purpose;" another student proposed "a quick 10 questions quiz to welcome students to using Udacity". The student, who was very thankful that the class was available because he had failed a class once already and could not afford to take it again, explained that he found the forum and chat supports "only after much frustration and me getting stuck in going to the web to find it." One student suggested that the course introduction include more information about the level of commitment required. And one student suggested that the courses include a requirement that students try to use one of the support services.

In response to the Survey 3 question about what they would like for Udacity and SJSU to do differently a large majority of students also pointed to "more information upfront about what kind of support services are available" (87% of matriculated students and 72% of non-matriculated students) and "more information about how to find and access support services" (79% of matriculated and 72% of non-matriculated students).

Six students responded by expressing their enthusiasm for the support services or for the course in general. A student in Stat 95 noted: "I find the existing support services prompt and excellent. So far the online chat support through Udacity has been fantastic. It is like being in a lab with support staff who walk around and help students." Another student commented "I think there is enough support and everyone who I have asked for help through it has been amazing."

Six other students had suggestions concerning the forum. One commented that it would be helpful if the forums targeted specific courses and were separated by function. Other students had similar suggestions for how to improve the organization of the forum to make it more user-friendly and organized. Three students expressed difficulty using the whiteboard.

In terms of overall improvements, no clear themes emerged among 42 responses received in Survey 1. Seven students said their class" is working well", or noted that they had only a few suggested improvements. “Everything in my opinion is well put together" one student noted. Four students make comments about the pace of the course. Two Stat 95 students said they would like to be able to download material—one wanted an official online book with content designed to fit the lesson plans. One said “I’m so used to traditional classes, it is uncomfortable not having any physical materials to review or follow along with." In Math 8, a couple of students said that for feedback they would like to know which answer they had gotten wrong, instead of just being informed that there was an error. This point was also brought up by a couple of students responding to the question of how to improve support services.

Perspective of Student Online Support Providers

In focus groups conducted both at the beginning and at the conclusion of the semester, Udacity Online Support Providers (OSP) explained that the intention of the online support is to help students become "unstuck." One OSP described the intended users as students who "have the right background or knowledge and are ready to succeed in class, but they are just confused about something or frustrated, and they get stuck on something and can't get past a certain point for whatever reason. Getting these students unstuck—that is where we are most effective." During the first focus group, the OSPs recounted that most of the support was provided to a relatively small number of students from the partner high school who were enrolled in the two math courses. Their impression was that many of the students lacked adequate preparation for the courses and were very unlikely to succeed. "If you look at the number of words that we use to type back and forth, out of all the students about half of those words went to five students. " The situation was exacerbated by the fact that the students in the partner high school were studying together without an actual instructor present, a situation that subsequently was resolved by the partner high school with the assignment of a staff person to the study sessions. Also contributing to complicate the situation was the fact that many students in the partner high school did not have computers at home, and therefore were only able to study when they had access to the computer lab at school.

Relating this information to the findings from Section 6, it is possible that the large amount of online support provided to students who were experiencing great difficulty during the first part of the semester and who were unlikely to succeed contributes to an explanation for the lack of a strong positive correlation between the use of online support and success.

In the second focus group conducted toward the end of the semester, the OSPs explained that they spent much less time with underprepared students than was the case earlier in the semester. They identified two other types of users that they were now able to spend more time serving. One was students who "will contact us quite a bit for a period of time and then forget about us, or focus on another class for a while and then come back and use us intensively before an exam or before problems that are due and then disappear again. Another group comprised "some students who ask one or two questions over the course of the semester."

In the focus groups, the OSPs also explained that they devoted time to developing supplemental materials and study guides to help students before exams. They noted that they were in an extremely good position to identify where students needed help or additional and alternative ways to approach the material because of their first-hand knowledge of where students got stuck. The OSPs observed that they "track everything" and that they were accumulating material and information throughout the first AOLE courses that would be used to strengthen subsequent iterations.

The OSPs also noted that they had a regular flow of course content traffic matching the weekly deadline in Stat 95 and that the enforcement of deadlines in this course was helping students stay on track. "If you don't give [the students] good enough deadlines they don't know how to function." One OSP added: "The other classes have been a little more flexible and I think that hurts more than it helps."

Faculty Perspectives:

As noted earlier faculty interviews are still in progress and a full account of the findings from this research will be provided in the final report. Accordingly just a few highlights from the initial faculty interviews will be provided here to provide additional insight into some of the points made by students and OSPs.

Faculty noted that students almost never asked them questions about content. One faculty member explained that she received and spent a lot of time responding to hundreds of e-mails from students asking questions about "things they should know already if they had read the green sheet (course syllabus)," a resource that included guidelines on course assignments and requirements, precisely the kind of information that students subsequently requested in the e-mail correspondence with faculty. Students' failure to read instructions also resulted in problems with the midterm and proctoring. One instructor noted that she had out of necessity learned to write colorful boldfaced e-mails to draw students' attention. The same instructor said that with the exception of one student who asked for help with a math problem all other inquiries concerned procedural matters. This instructor also noted that she was wondering whether students were aware of the opportunity to schedule office hours. During the entire course only one student scheduled such a session with her. Another instructor from a different course echoed this point. "I had one question about content all semester-it was all about technical issues."

Faculty also corroborated points made by the OSPs about the fact that some students, in particular students from the partner high schools, lacked adequate preparation for the AOLE courses. One of the Math 6L instructors explained that: "Math 6L is designed to be a review and a refresher of algebra skills." Some of the partner high school students were not prepared for this kind of assignment.

Faculty members also felt that students did not know about the course online support and other services available to them until well into the semester. One instructor noted that he also did not know for a long time that "there were a bunch of services available to students." He added: "I had a vague notion that there was outreach by Udacity to students," but it took a while before he realized that Udacity had hired online support providers.

Faculty also noted the difficulties they experienced in communicating information to students who "do not respond to e-mails or read entire e-mails." It is possible that the students' lack of experience with online education contributed to this added communication challenge. As noted earlier, some students appeared to be unable to retrieve e-mails, a situation that may have been further complicated by the need for students to interact with two websites.

Coordinator and Leader Perspectives

The leaders and coordinators underscored in interviews that their goal for AOLE is to help more students, particularly those who come from under-resourced backgrounds, gain access to affordable and high-quality education. All the interviewees expressed continued commitment to working together to achieve this goal. Each leader also expressed deep satisfaction with the commitment they have seen from their partner in the AOLE collaboration. In addition, the leaders commented on the leadership, initiative and very hard work the instructors have made to the AOLE courses.

The interviewees agreed that much has been learned from the first AOLE iteration and explained the many ways in which changes have been made to strengthen the summer AOLE courses.

One AOLE leader identified as one lesson learned that: "In the future participating high schools and other institutions need to be more involved, as do parents and others in a position to provide support to students." Overall, there was a desire to engage more internal and external stakeholders in each phase of the process of development and delivery.

Conclusion and Recommendations

In the conclusion we return to the three research questions that drove the analysis presented in this paper.

1. Who engaged and who did not engage in a sustained way and who passed or failed in the remedial and introductory AOLE courses?

The research found that matriculated students performed better than non-matriculated students and that, in particular, students from the partner high school were less successful than the other AOLE students. Further, students in Stat 95 were more successful overall than students in the other courses.

The disappointing low pass rates in all courses should be considered in light of the fact that the project specifically targeted at-risk populations, including students who had failed Math 6L before and groups demonstrated by other research to be less likely to succeed in an online environment. As noted in Section 1, research has found that these students do less well in online than in face-to-face courses. Further, student groups in at least one major study (Jaggars and Xu, 2013) who were found to experience the greatest negative effect from taking courses online share many of the characteristics found among the AOLE partner high school students in particular, a group with very low pass rates in Spring 2013.

2. What student background and characteristics and use of online material and support services are associated with success and failure?

The analysis presented in this paper found that effort, measured in a variety of ways, trumps all other variables tested for their relationships to student success. The clearest predictor of passing a course is the number of problem sets a student completed. The relationship between completion of problem sets and success is not linear; rather the positive effect increases dramatically after a certain baseline of effort has been made.

The most successful students across all the courses, those in Stat 95, had mandatory assignments contributing to the final grade that they had to submit every week. This tight structure and regular accountability may—as both the OSPs and faculty members pointed out—have helped this group of students overcome, to some degree, their lack of online preparation. This group also demonstrates the steepest relationship between effort and chance of passing.

The various measures of amount of effort devoted to a course, demonstrated to have strong positive relationships to passing, collectively measure magnitude, early initiation of activity, and persistence and evenness of activity over the duration of a course. Reinforced by the predictive strength of number of problems done, the research team infers that required assignments or any other activity that must be completed in a timely manner during a course may provide the highest rate of learning and consequently of passing. Assignments required to be turned in throughout a course increase, for the motivated students, amount of effort and early initiation, persistence and evenness of activity.

The fact that the regression analysis did not find a positive relationship between use of online support and positive outcomes should not be interpreted to mean that online support is not an important factor to increase student engagement and success. As students, OSPs and faculty members explained, several factors complicated students' ability to fully use the support services, including their limited online experience, their lack of awareness that these services were available and the difficulties they experienced interacting with some aspects of the online platform. It is thus the advice of the research team that additional investigations be conducted into the role that online and other support can play in the delivery of AOLE courses once the initial technical and other complications have been addressed.

3. What do key stakeholders (students, faculty, online support services, coordinators, and leaders) tell us they have learned?

Much has been learned and improvements are already in progress in the second AOLE iteration. Perhaps most importantly, the faculty members who taught these courses, although they had to contend with major difficulties along the way, all believe that the content that has been developed has tremendous potential to advance students' critical thinking and problem solving abilities. The fact that content is continuously updated in response to student input is a major pedagogical advantage. One faculty member noted: "If you want to teach algebra as a language, visualization is important—this can be achieved in the online format along with a degree of contextualization that is often missing from basic skills and introductory courses." Another instructor echoed this sentiment, commenting: "The courses are much more contextualized and cover much more material than the regular courses. It is an exciting way to engage students." A colleague noted: "It was exciting to contextualize the material. Udacity has brought to the table ways to make it more inquiry-based and added real life context."

Appendix A. Methods and Methodology

Data Preparation

Records from the different data sources were matched by student ID.

There are seven basic categories of data, by content, including new variables created from the original data:

- Student characteristics (e.g., matriculation, gender, age, Pell status, underrepresented minority status, partner high school student or not)
- Course characteristics (units)
- Case relevance for difference levels of analysis (e.g., students who participated in a course vs. students who did not or who withdrew)
- Student performance (pass/fail, grade)
- Student effort and persistence (e.g., log in time, weeks in which more than a minimum time was logged in, relationship between login time and expected conventional in-class time by course, problems done)
- Student exposure to and use of support (e.g., outreach phone calls to them, contacts with OSPs, time spend with OSPs, keyboard input by students and OSPs)
- Qualitative data from the three surveys

Recording errors were identified, largely through tests of reasonableness.

Identification of missing cases by variable, and decisions on which could potentially yield useful information (qualitatively, or quantitatively descriptively if integrated into statistical tests). We removed from the analyses students who either withdrew from their course or who received an Incomplete, leaving 213 who completed their courses.

- Of the 36 withdrawals, 4 were matriculated
- Mean sessions: 39 vs. 84 for completing students
- Mean weeks active for at least 30 minutes: 4.8 vs. 10.4 for completing students
- Mean age: 28 vs. 23 for completing students

The three surveys were tested for representativeness against student characteristics and were found to be representative only sporadically by variable.

Pass/fail is used as the dependent variable.

Quantitative Methods

1. Basic descriptions of students by 34 independent variables in cross tabulations answer questions about who took what courses and how they did.

2. Tests of interactions between the independent variables and pass/fail, using Pearson chi square

3. The Benjamini-Hochberg False-Discovery Rate (FDR) was applied to the list of contingency table results to adjust for the multiple comparisons significance problem (with 35 tests we expect 1.75 to have apparent significance by random chance). This dropped two tables below significance relative to their apparent 0.05 significance. (see Tukey 1977 and Benjamini *et al*, 1995)

Models of relationships between pass/fail and the independent variables

Relationships offer the richest potential for gaining insight into the AOLE project, if the data are clean, the variables are well defined, the number of cases is sufficient, and there actually are relationships to be found. The models address the question: what student background and characteristics and use of online material and supports are associated with success and failure?

Pass/fail, being a binary variable, is more challenging to study than continuous dependent variables. We use logit models to derive theoretical functions between the probability of passing and various independent variables representing student characteristics, effort, persistence, and use of services.

The logit model estimated using logistic regression circumvents the invalid attempt to model binary functions with ordinary least squares (OLS) into a valid approach which models a continuous function of the odds of passing under the influence of the independent variables, expressed as natural logs of the odds, iteratively maximizing the log likelihood rather than minimizing sums of squares. We then translate the predicted log odds into probabilities of pass/fail as functions of the significant independent variables.

Logit model as natural log of odds (p/1-p), where p = probability of passing:

a) $Ln[p/(1-p)] = \alpha + \beta_1 X_i + \beta_2 X_i + \ldots$

b) The coefficients are estimated using iterative maximum likelihood estimation, rather than ordinary least squares. They are partial, i.e., their values are controlling for the other variables.

c) Each odds ratio of passing (within context of independent variable) = e^{β} For example, the ratio of the odds of a female passing as a multiple of the odds of a male passing.

d) Logit model as probability: p(Passing) = $e^{(\alpha + \beta_1 X_i + \beta_2 X_i + \ldots)} \Big/ 1 + e^{(\alpha + \beta_1 X_i + \beta_2 X_i + \ldots)}$

Independent variables were screened with combinations of a comprehensive correlation matrix and with three measures of variance inflation attributable to multicollinearity: tolerance, VIF and $VIF^{½}$. The latter represents the multiple of variance inflation for each independent variable.

The following tables display the values of the three multicollinearity statistics.

Table A.1. Multicollinearity Indicators

Independent Variables	Collinearity Statistics: All Independent Variables Examined			Collinearity Statistics: Only Model-Significant Variables		
	Tolerance	VIF	$VIF^{½}$	Tolerance	VIF	$VIF^{½}$
1st Week Logged In ≥ 0.5 Hours	0.755	1.324	1.151			
1st Week Spoke with Support	0.701	1.426	1.194			
Age	0.579	1.726	1.314			
First Generation	0.761	1.314	1.146			
Gender	0.860	1.163	1.079			
Math 6L03	0.491	2.036	1.427			
Math 801	0.538	1.860	1.364			
Math 802	0.357	2.805	1.675			
Meaningful Phone Calls	0.686	1.458	1.207			
Partner High School	0.357	2.801	1.674			
Ratio Logged Time to F2F Class Time	0.279	3.589	1.894			
Stat 9501	0.457	2.186	1.479			
Stat 9502	0.355	2.817	1.678			
Support Staff Characters Typed	0.636	1.573	1.254	0.806	1.240	1.114
Number of Sessions Logged In	0.277	3.616	1.901	0.365	2.743	1.656
Total Problems Done	0.212	4.708	2.170	0.434	2.303	1.518
Video Time	0.256	3.902	1.975	0.440	2.273	1.508
Weeks Active ≥ 1/2 Hour	0.162	6.182	2.486	0.327	3.060	1.749

All of the intercorrelations of the selected variables in the final models are reduced relative to pre-inclusion. Technically VIF and its transforms are not considered appropriate for logistic regression, so we conducted a test to see if the variables they indicated as troublesome would produce inconsistent results, which they did; especially for weeks active at different minimum hours per week. This proved helpful in trimming the models for consistency.

We used backward stepwise selection of independent variables. Backward stepwise is recommended by some statisticians after a reduced variable set has been decided, while forward stepwise is recommended for more exploratory phases of a project. There is no universal agreement on which is preferable. (To be prudent we conducted both and obtained the same results.) Each step needs to be examined.

Our models are evaluated according to the following criteria

- Statistical validity: We believe the logit model is the correct method for the specific data available.
- Statistical significance of results: The probability that the coefficients are the result of a random process, rather than representing a repeatable association
- Strength of relationship: Logistic regression cannot provide the conventional R^2 seen in ordinary least squares regression (OLS), but the appropriate statistics are applied and evaluated in this approach.
- Direction of relationship: Do odds of passing increase or decrease as the independent variable(s) increase? The log odds β_i coefficients can be positive, zero or negative, as in OLS.
- Form of relationship: The manner in which odds of passing functions as each independent variable changes is revealed in graphs of the logit model as probability: p(Passing). One positive attribute of a form is to see what would happen to it at higher and higher values of the input (independent) variable. The logistic forms in this analysis either approach an asymptotic maximum or would if the independent variable's values were extended sufficiently. In contrast, poor forms attain absurd highs or lows if extended. (See graphing method below.)
- Consistency of relationship: Confidence in possible insights for an independent variable increases if it behaves similarly in modified circumstances, rather than jumping in value.
- Substantive meaningfulness: If the predicted odds ratios equal or are close to 1:1 relative to the variable's scale, then the independent variable is not associated with substantive change in pass/fail. If the confidence interval for the log odds ratio lies entirely below 1.0, then we are confident (we are using 95%) that the independent variable lowers the probability of the student passing as its value increases, and vice versa. Values that straddle 1.0 do not provide clear insight as they are both positive and negative. We do not include these in our conclusions as substantively meaningful.

Required logistic regression assumptions are met by the six models

- Binary or ordinal dependent variable—pass/fail is binary
- The dependent variable's desired outcome, passing, is coded = 1
- Independent error terms
- Independent variables are linearly related to the log odds. (This does not mean the non-log dependent and independent variables to be related linearly.) (see Garson, 2012)
- Limited multicollinearity (VIF < 10, or $VIF^{½}$ < 3.16) (see Kutner *et al*, 2004)

- Parsimony: not fitted with many independent variables. The six models have a maximum of three.
- Relatively large sample sizes

Graphing method

Model 6 in Figure 6 is straightforward: the one independent variable, Video Time, is graphed as the x_i value against p(Passing) as $\hat{y}_i$. A comprehensive graph of multiple independent variables on p(Passing) would have to be multidimensional, which is difficult to interpret. The multi-independent variable relationships in this report are graphed one independent variable at a time, but with the effect of the other independent variables included at their mean values for the relevant student group. This represents the cross section of the comprehensive multidimensional graph sliced at the combination of the values of the graphed independent variable and those mean values of the other independent variables, thus holding them constant graphically. For example, the equation for graphing the effect of Support Characters Typed on p(Passing) is:

$$\text{p(Passing)} = e^{(\alpha + \beta_1 S_i + \beta_2 \overline{V} + \beta_3 \overline{W})} \Big/ 1 + e^{(\alpha + \beta_1 S_i + \beta_2 \overline{V} + \beta_3 \overline{W})},$$

Si: individual values of Staff Characters
$\overline{V}$: mean of Video Time
$\overline{W}$: mean of Weeks Active

Comparison of regression results across student groups

The log odds (β) is the measure used to compare logistic regressions with each other to test for significance of differences. The compared coefficients must come from regressions with identical sets of independent variables. Only two of the regressions of the six have the identical configuration, and attempting to force more to be identical would be futile, as most of the rejected independent variables are extremely insignificant. One test was done to gain insight into strength of the differences: Problems Done for all students vs. matriculated students. The Wald χ^2 test is highly significant: (< 0.0001).

Appendix B. Detailed Results Tables

Table B.1. Contingency Table Coefficients (In rank order of significance)

Group	Independent Variable Tested	p-value
All	Mean Problems Done	0.0000
All	Pass/Fail: +/- Mean Wks Active 1/2 Hr	0.0000
All	Ratio to F2F Class Time	0.0000
Math 8 Non-matriculated	Ratio to F2F Class Time	0.0000
All	Ratio to F2F Class Time (+/- 1 SD)	0.0000
Partner HS	Ratio to F2F Class Time	0.0000
All	By Course	0.0000
Non-matriculated	By Course	0.0000
Stat 95 Non-matriculated	Ratio to F2F Class Time	0.0000
Math 6 Matriculated	Mean Problems Done	0.0010
All	Spoke w OSP <= Wk5	0.0020
Stat 95 Matriculated	Pass/Fail: +/- Mean Wks Active 1/2 Hr	0.0050
Math 6 Matriculated	Pass/Fail: +/- Mean Wks Active 1/2 Hr	0.0050
Stat 95 Matriculated	Mean Problems Done	0.0080
Math 6 Non-matriculated	Pass/Fail: +/- Mean Wks Active 1/2 Hr	0.0110
Math 8 Non-matriculated	Mean Problems Done	0.0130
Math 6 Non-matriculated	Mean Problems Done	0.0140
Math 8 Non-matriculated	Pass/Fail: +/- Mean Wks Active 1/2 Hr	0.0280
Stat 95 Matriculated	Ratio to F2F Class Time	0.0500
All	Age (2 groups)	0.0600
Matriculated	By Course	0.0650
All	Gender	0.0990
Math 6 Matriculated	Ratio to F2F Class Time	0.1180
Stat 95 Non-matriculated	Mean Problems Done	0.1690
Partner HS	Pass/Fail: +/- Mean Wks Active 1/2 Hr	0.1850
Partner HS	By Course	0.3220
Non-matriculated	Age (2 groups)	0.3290
Partner HS	Mean Problems Done	0.4000
Matriculated	Gender	0.4690
Partner HS	Age (2 groups)	0.4750
Math 6 Non-matriculated	Ratio to F2F Class Time	0.5330
Partner HS	Gender	0.5870
Math 8 Matriculated	Mean Problems Done	0.5900
Math 8 Matriculated	Ratio to F2F Class Time	0.6140
Partner HS	Age, Actual (5 groups, 15-19)	0.9030

Tables below line deemed insignificant by Benjamini-Hochberg FDR, using Q = 0.05.

Setting FDR Q = α = 0.05 is conservative for this procedure. A higher Q would allow more tables to be deemed significant. (Higher Q is common in exploratory analysis, especially when monetary consequences, as in preliminary tests of drug substances, would prematurely eliminate promising cases.)

Table B.2. Six Logistic Regression Models: Coefficients for Significant Relationships

Notes		Model Statistics	Model 1 All Students	Model 2 Matriculated Students	Model 3 Non-matriculated	Model 4 Math 6L	Model 5 Math 8	Model 6 Stat 95
Sample N is smaller for each subgroup.		**Sample N (number of students)**	213	98	115	81	58	74
All models highly significant. There is no least-squares comparable R^2 for logistic regression. Easily interpretable measures include cases correctly predicted and significance.		**Model Chi-Square**	88.724	35.422	56.306	26.712	32.696	53.643
		Significance	0.000000	0.000000	0.000000	0.000023	0.000000	0.000000
		df	2	2	5	4	3	2
		Percent Cases Correctly Predicted						
		Passed (≥ C, or NC for Math 6L)	80.3	82.9	73.3	50.0	76.9	89.5
		Failed (< C, or NC for Math 6L)	85.2	68.4	92.9	90.2	93.3	86.1
		Overall	83.6	74.5	87.8	80.2	89.7	87.8
Regression constant		**Constant of Model (B_0)**	-5.035970	-5.063511	-6.443615	-5.38333	-6.39604	-8.55363
See discussion & graphs for interpretation		**Significant Variables**						
Odds Ratios for All, Matriculated, Math 6, & Math 8 students: odds passing/odds failing per added problem done, while the other variables remain unchanged. Is most meaningful if upper & lower % limits are all > or < 1.0. Log Odds is regression coefficient; the Odds Ratio is more interpretable.	EFFORT	**Total Problems Done**						
		Odds Ratio (e^B)	1.305382	1.366689		1.359496	1.399837	
		Lower 95% Limit	1.169934	1.159301		1.103739	1.034251	
		Upper 95% Limit	1.456513	1.611177	not sig.	1.674518	1.894650	not sig.
		Significance (Wald chi-sq 1 df)	0.000002	0.000008		0.003875	0.029403	
		Log Odds Ratio B	0.266496	0.312391		0.307114	0.336356	
Odds Ratios for All, Matriculated, Math 6, & Stat 95 students: ratio of odds passing to odds failing per unit video time, while the other variables remain unchanged.		**Video Time**						
		Odds Ratio (e^B)	1.000097	1.000075	1.000134			1.000638
		Lower 95% Limit	1.000045	1.000008	1.000034			1.000359
		Upper 95% Limit	1.000149	1.000143	1.000233	not sig.	not sig.	1.000918
		Significance (Wald chi-sq 1 df)	0.000274	0.000274	0.008757			0.000008
		Log Odds Ratio B	0.000097	0.000075	0.000134			0.000638
Odds Ratio for Non-matriculated students: ratio of odds passing to odds failing per added week of at 1/2 hour time logged in, while Video Time & Support Staff Characters Typed remain unchanged.		**Weeks Active for ≥ 1/2 Hour**						
		Odds Ratio (e^B)			1.342327			
		Lower 95% Limit			1.060320			
		Upper 95% Limit	not sig.	not sig.	1.699338	not sig.	not sig.	not sig.
		Significance (Wald chi-sq 1 df)			0.014416			
		Log Odds Ratio B			0.294405			
Odds Ratio for Math 8 students: ratio of odds passing to odds failing per added session logged in, while Total Problems Done remains unchanged.		**Number of Sessions Logged In**						
		Odds Ratio (e^B)					1.032909	
		Lower 95% Limit					1.005240	
		Upper 95% Limit	not sig.	not sig.	not sig.	not sig.	1.061339	not sig.
		Significance (Wald chi-sq 1 df)					0.019427	
		Log Odds Ratio B					0.032379	

continued on next page

Table B.2. (continued)

Notes		Model Statistics	Model 1 All Students	Model 2 Matriculated Students	Model 3 Non-matriculated	Model 4 Math 6L	Model 5 Math 8	Model 6 Stat 95
Odds Ratio for Non-matriculated students: ratio of odds passing to odds failing per added character typed for the student, while Weeks Active & Video Time remain unchanged. The relationship is negative (odds < 1), possibly reflecting greater support time for struggling students.	- SUPPORT -	**Support Staff Characters Typed**						
		Odds Ratio (e^B)			0.999819			
		Lower 95% Limit			0.999666			
		Upper 95% Limit	not sig.	not sig.	0.999971	not sig.	not sig.	not sig.
		Significance (Wald chi-sq 1 df)			0.019699			
		Log Odds Ratio B			-0.000181			
These measures have problematic interpretations in general, and are reported here for thoroughness. Significance in the Hosmer-Lemeshow test is reversed by definition. The Nagelkirke is a log likelihood pseudo-R^2 that may understate the fit. [see Garson (2012) p. 92]		**Other Model-Quality Measures**						
		Hosmer & Lemeshow Test						
		Chi-Square (8 df)	20.318	11.276	11.848	13.639	1.937	10.002
		Significance	0.009	0.187	0.158	0.092	0.983	0.265
		-2 Log Likelihood	182.431	97.811	75.706	63.833	29.027	48.888
		Nagelkerke (= Cox & Snell normed)	0.473	0.408	0.567	0.417	0.658	0.688

Table B.3. Six Logistic Regression Models: Coefficients for Insignificant Relationships (from shaded cells in previous table) [from last step before elimination]

Model Statistics	Model 1 All Students	Model 2 Matriculated Students	Model 3 Non-matriculated	Model 4 Math 6L	Model 5 Math 8	Model 6 Stat 95
Total Problems Done						
Odds Ratio (e^B)			1.127023			0.861521
Lower 95% Limit			0.953596			0.486355
Upper 95% Limit			1.331992			1.526085
Significance (Wald chi-sq 1 df)			0.160729			0.609383
Log Odds Ratio B			0.119580			-0.149056
Video Time						
Odds Ratio (e^B)				1.000152	0.999966	
Lower 95% Limit				0.999975	0.999844	
Upper 95% Limit				1.000330	1.000088	
Significance (Wald chi-sq 1 df)				0.092254	0.581737	
Log Odds Ratio B				0.000152	-0.000034	
Weeks Active for ≥ 1/2 Hour						
Odds Ratio (e^B)	1.077056	1.078410		1.032639	0.841410	0.979721
Lower 95% Limit	0.893662	0.828763		0.662594	0.372591	0.528929
Upper 95% Limit	1.298085	1.403257		1.609348	1.900130	1.814711
Significance (Wald chi-sq 1 df)	0.435716	0.574184		0.887181	0.677800	0.948060
Log Odds Ratio B	0.074231	0.075488		0.032118	-0.172676	-0.020488
Number of Sessions Logged In						
Odds Ratio (e^B)	1.002185	1.008481	0.995279	1.016687		0.981418
Lower 95% Limit	0.990424	0.990384	0.981689	0.998838		0.963141
Upper 95% Limit	1.014085	1.026909	1.009056	1.034855		1.000042
Significance (Wald chi-sq 1 df)	0.717076	0.360677	0.499881	0.067056		0.050518
Log Odds Ratio B	0.002182	0.008445	-0.004732	0.016549		-0.018756
Support Staff Characters Typed						
Odds Ratio (e^B)	0.999992	0.999985		0.999958	0.999954	1.000018
Lower 95% Limit	0.999973	0.999956		0.999883	0.999898	0.999956
Upper 95% Limit	1.000011	1.000015		1.000032	1.000011	1.000081
Significance (Wald chi-sq 1 df)	0.399418	0.333039		0.265938	0.110791	0.565401
Log Odds Ratio B	-0.000008	-0.000015		-0.000042	-0.000046	0.000018

References

Benjamini, Y., Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. J. *Roy. Statist. Soc.* B, 57, pp. 289–300

Buckingham Shum, Simon. (2012). Learning Analytics *Policy Brief*: UNESCO Institute for Information Technologies in Education.

Cox, Rebecca. (2009) The College Fear Factor: How Students and Professors Misunderstand One Another. Harvard University Press

Challenges. UK: Knowledge Media Institute, The Open University.

Fritz, John. (2011). Classroom Walls that Talk: Using Online Course Activity Data of Successful Students to Raise Self-awareness of Underperforming Peers. *The Internet and Higher Education, 14*(2), 89-97. doi: 10.1016/j.iheduc.2010.07.007

Garson, G. David. (2012). *Logistic Binomial Regression.* Raleigh, NC: Statistical Associates Publishing.

Jaggars, Shanna Smith, & Bailey, Thomas. (2010). Effectiveness of Fully Online Courses for College Students: Response to a Department of Education Meta-Analysis. In C. C. R. Center (Ed.). New York, New York: Teacher's College, Columbia University.

Kutner M.H., Nachtsheim C.J., Neter J. (2004). *Applied Linear Regression Models*, 4th ed, New York: McGraw-Hill Irwin.

Macfadyen, Leah P., & Dawson, Shane. (2010). Mining LMS Data to Develop an "Early Warning System" for Educators: A Proof of Concept. *Computers & Education*(54), 11.

Ryabov, Igor. (2012). The Effect of Time Online on Grades in Online Sociology Courses. *MERLOT Journal of Online Learning and Teaching, 8*(1).

Siemens, George. (2011). Learning and Academic Analytics. Retrieved from http://www.learninganalytics.net/

Tukey, J.W. (1977). Some Thoughts on Clinical Trials, Especially Problems of Multiplicity. *Science*, Vol 198, No. 4318, pp. 679-684.

U.S. Department of Education Office of Planning, Evaluation, and Policy Development. (2009). Evaluation of Evidence-Based Practices in Online Learning A Meta-Analysis and Review of Online Learning Studies (p. 93). Washington, D.C.

Whitmer, J., Fernandes, Kathy, & Allen, Bill. (2012). Analytics in Progress: Technology Use, Student Characteristics, and Student Achievement. *Educause Review Online*(July 2012).

Xu, Di, & Jaggars, Shanna Smith. (2011). The Effectiveness of Distance Education Across Virginia's Community Colleges: Evidence from Introductory College-Level Math and English Courses. *Educational Evaluation and Policy Analysis, 33*(3), 360-377.