A randomized controlled trial to evaluate and compare the teaching effectiveness of three online education platforms: Quantic, edX, and Khan Academy

August 3, 2017

Melissa L. Constantine, PhD, MPAff
Research Fellow, Division of Health Policy & Management
University of Minnesota

President, Health Measurement and Survey Research Consultants

Table of Contents

Executive Summary

Purpose

Student enrollment in distance education courses—courses in which the instructional content is delivered exclusively via the Internet—has been growing steadily in the United States. According to the U.S. Department of Education's most recent statistical report, the number of higher education students enrolled in at least one distance education course increased by a half million between 2012 and 2015—from 5.5 million to 6 million (NCES, 2017). Over the same period, enrollment in non-degreed online programs—including MOOC's (Massive Online Open Courses)—increased worldwide from zero in 2012 to 23 million in 2015 (Shah, 2016). That prompts the question: how *effective* are these online platforms relative to learner knowledge acquisition and retention? And which one is the most effective? In response to such concerns, the purpose of the current study is to evaluate and compare the teaching effectiveness of three online statistics courses offered by Quantic, edX, and Khan Academy, respectively.

Methods

We conducted a Web-based research study using an experimental between-group design to compare the effects of three on-line educational statistics courses on study participants' understanding of basic concepts in probability and statistics. Eligible respondents were randomly assigned to one of three statistics courses, offered on three different on-line education platforms (Quantic, edX, and Khan Academy). A randomized design allowed us to control for potential confounders and enabled us to draw causal inferences about the effects of the teaching approaches of the courses on knowledge acquisition (Shadish, Cook & Campbell, 2002).

Results

A significantly greater proportion of learners who started the Quantic statistics course completed it (79.6%) compared to those who started the Khan Academy course (66.6%) ($p \leq 0.05$). Quantic learners demonstrate better knowledge acquisition through Final Exam scores than either edX or Khan Academy learners ($p \leq 0.05$). A greater proportion of Quantic learners (78.4%) scored 70% or higher on their statistics course than either edX learners (61%) or Khan Academy learners (62.8%). Overall, Quantic learners were more satisfied with their assigned course than edX or Khan Academy learners, with Quantic learners also reporting their course was better able to hold their attention than edX or Khan Academy learners ($p < 0.0001$). Nevertheless, manual counts of the distribution of final exam scores indicate that a much higher proportion of learners in the Quantic course had a final exam score above 70% (78.4%) compared to the edX (61.0%; p=.0243) and Khan Academy (62.8%; p=.0367) courses. The difference between Quantic and edX was 17 percentage points, while the difference between Quantic and Khan Academy was 16 percentage points. Notably, in the study sample, the percentage of edX and Khan Academy students who scored below 50% correct on the exam was over almost 5 times greater than the percentage of Quantic students who scored below that mark.

Discussion

Study results demonstrate that Quantic delivers better outcomes, more consistently, and with greater student satisfaction than two of the most celebrated Massive Open Online Course (MOOC) providers, edX and Khan Academy. Additionally, the Quantic course completion time is significantly lower, at 2.5 hours, compared to the 4.75 hours to complete the edX course and the almost 9 hours to complete the Khan Academy course.

Introduction

Background

Student enrollment in distance education courses—courses in which the instructional content is delivered exclusively via the Internet—has been growing steadily in the United States. According to the U.S. Department of Education's most recent statistical report, the number of higher education students enrolled in at least one distance education course increased by a half million between 2012 and 2015—from 5.5 million to 6 million (NCES, 2017). Over the same period, enrollment in non-degreed online programs—including MOOC's (Massive Online Open Courses)—increased worldwide from zero in 2012 to 23 million in 2015 (Shah, 2016). That prompts the question: how *effective* are these online platforms relative to learner knowledge acquisition and retention? And which one is the most effective? In response to such concerns, the purpose of the current study is to evaluate and compare the teaching effectiveness of three online statistics courses offered by Quantic, edX, and Khan Academy, respectively.

In response to the growing demand for distance education, various educational technology firms have worked to translate what are considered best practices in classroom learning to the online experience. One of these best practices is *active learning*, an approach that engages students more vigorously in the learning process (Tintle et al). Instead of listening to lectures or watching videos, students in active learning environments acquire new concepts by participating in exercises and discussions that require them to analyze problems and think creatively of solutions. For instance, a statistics course with an active learning approach may require students to virtually meet with other students to discuss a problem and use computer technology to solve it (Aliaga et al. 2005).

Findings from various research studies indicate that the active learning approach increases student retention in both classroom and online environments (Lockwood, Parr & Smith, Kvam 2000; Phan et al, 2015). Of the three online learning modules evaluated in this study, the Quantic course best exemplifies the active learning approach because it does not use videos to teach statistics but engages students in interactive exercises and provides them immediate feedback in response to every answer they submit. By contrast, the edX course presents it content on videos and does not offer interactive exercises. The Khan Academy course also uses video instruction but includes some interactive exercises.

A second best practice of online learning is the use of static or dynamic digital screen displays. Static displays are those on which the symbols and graphics do not change automatically and are not responsive to viewer interaction. Dynamic displays are those on which the language symbols change automatically as a normal part of user experience. A dynamic teaching approach incorporates ongoing, individualized feedback to learners as they engage in activities. (Rieber, 2016).

A secondary factor in learner success is learner motivation. Research findings consistently suggest that learners with a purpose, such as achieving course credit or parental reward associated with good grades, a form of extrinsic motivation, out-perform students without this motivation (Phan et al, 2016). Along the continuum of motivation, intrinsic motivation is modelled as the most self-determined, an autonomous and high quality motivation whereas extrinsic motivation is considered less self-determined and more controlled than intrinsic motivation (Barak et al, 2016, Vallerand et al, 1992). While high performing students may possess a greater degree of intrinsic motivation, their more moderately performing counterparts are apt to be more extrinsically motivated. Extrinsic motivation is initiated and regulated by external contingencies, including the promise of reward or punishment. i.e., passing or failing a class, or failing to advance to the next grade level, or desire to avoid parental reaction to a failing course grade. Therefore, the role of extrinsic motivation in student performance has broader applicability across the spectrum of learners. Within the typology of extrinsic motivation, the major influencers in an educational environment are likely to be identified motivation and externally regulated motivation. Identified motivation is based on the perceived usefulness of the behavior, i.e., a student identifies with the value of the course and willingly accepts responsibility for regulating their behavior in completing the course (Utvaer and Haugan, 2016). While many on-line learning platforms are integral components of degree-granting curricula, MOOCs do not confer credit and do not require learners meet any admission criteria. Therefore, it is reasonable to expect that MOOC learners have lower extrinsic-identified motivation to complete a course than learners using other on-line platforms, with the associated potential for differential learning or knowledge acquisition outcomes.

Methods

Study Design

The general method used to compare the effectiveness of the Quantic, edX, and Khan Academy online education platforms for teaching statistics was an experimental randomized control design.

Students who qualified for the study (see selection criteria below) were randomly assigned to one of the three online statistics courses. The randomized design controls the distribution of potential confounding variables across the three test groups, enabling us to validly infer that differences in student test scores are caused by differences in the teaching methods (Shadish, Cook & Campbell).

Course Selection

The primary goal of this study is the evaluation of teaching effectiveness of Quantic statistics courses compared to close competitors. Multiple on-line learning platforms offering introductory statistics courses were evaluated for equivalence to the Quantic statistics course across learning methodology (dynamic-static and active-passive), course content, audience reach and enrollment. Additional evaluative criteria for selection of statistics courses were open and free access to course. After thoroughly reviewing seven potential platforms, the edX and Khan Academy statistics courses were determined to be the most comparable to the Quantic course in respect of their course content and their intended audiences.

Course Profiles

As noted in Table 1, each course employs a mix of learning methodologies and the approximate time to complete each course ranges from 2.5 hours for Quantic to almost 9 hours for Khan Academy. Full links to the course content and completion times are provided in Appendix Table B. Below are brief narrative descriptions of the Quantic, edX, and Khan Academy online education platforms and their statistics courses.

Quantic

Quantic, founded in 2014, teaches new concepts to learners by engaging them in interactive exercises and providing immediate feedback in response to their inputs. The interactive platform displays affirmative messages when learners enter correct answers and explanations when they submit incorrect responses. This method of instruction is more effective than lectures because it allows students to check their new understanding of concepts as they're learning them. Quantic focuses on creating engaging, conversational, and humorous lessons to make the content memorable to the student. Instead of just presenting facts—the traditional mode of instruction—Quantic integrates stories and real-world examples into mental models that help students retain what they learn.

Quantic's "Introduction to Statistics" course includes four modules: Data Collection, One-Variable Statistics, Two-Variable Statistics, and Probability Fundamentals. Some topics covered in Quantic's Two-Variable Statistics module were not covered in the edX and Khan Academy courses. For that reason, this module was not a required module for this study. The total minimum required completion time for the three Quantic modules (Data Collection, One-Variable Statistics, and Probability Fundamentals) is 2 hours and 30 minutes This includes four graded, untimed "SmartCase" lessons that assess how well students learned the concepts taught in the course.

edX

The edX open-source online education platform was created by Harvard University and MIT in 2012 for hosting Massive Open Online Courses (MOOC's). Most MOOCs are developed by faculty of higher education institutions, but offered free of charge to the general public. Typically, no credit towards a degree or certificate is given for completion of a MOOC.

The edX "Introduction to Statistics" online course evaluated in this study was developed by two faculty members in the Statistics Department at the University of California, Berkeley, Its three modules—Descriptive Statistics, Probability, and Inference—are taught using a video lecture format, without any practice exercises. Their total completion time is 4 hours and 45 minutes. The course is available for free on YouTube and at edx.org.

Khan Academy

Khan Academy offers online instructional videos in a variety of subjects, including math modules for learners of all ages, from kindergarten to postsecondary level. Its Statistics and Probability sequence covers 13 general topic areas. Eight of the more advanced topic areas—including advanced regression, significance tests and ANOVA—did not map to those in the Quantic or edX courses. Thus, students in the Khan Academy test group were required to study only five of the thirteen topics.

As noted above, Khan Academy presents its content using a video-based approach. But it also engages in students in practice exercises after they view the videos. The modules assigned to students in this study required at minimum 8 hours and 49 minutes to complete. The course is available for free on YouTube and at khanacademy.org.

Table 1. Course Content Comparison

Platform/Course	Learning Methodology	Time to Complete (hours: minutes)	Base Compensation
Quantic	Dynamic, Active	2:30	$45
edX	Static, Passive	4:45	$70
Khan Academy	Dynamic, Passive	8:49	$95

Recruitment Details

Study participants were recruited using informational postings on social media sites, such as MBA and undergraduate student groups. (See Table A in the Appendix for a list of all the sources used for recruitment.) The informational postings included a link to a web page that described what participants are required to do in the study and the terms of their compensation. The webpage was hosted on the survey platform Qualtrics, which this study used for survey administration. The time commitment required from the participants varied greatly depending on the course assigned—i.e., the minimum course completion times were 2.5 hours for Quantic, 4.75 hours for edX, and 9 hours for Khan Academy. Respondents to the informational postings were informed that, to adjust for these differences, those assigned to Quantic course would receive a base amount of $45 for completing it, those assigned to the edX course would receive $70, and those assigned to the Khan Academy would receive $95.

Sample Composition

Those interested in entering the study were instructed to answer an online questionnaire to determine their eligibility. To qualify for the study, students were required to be 16 years of age or older, reside in the US, use English as the primary language spoken in home or self-report “good” or “very good” English language reading and comprehension skills, and have no prior coursework in statistics. Respondents were also asked to submit a valid email address and verify that they had an US IP address.

Sample Sizes

The target sample size was 75 study participants completing each course using sequential enrollment. This number was based on power calculations assuming t-tests to detect a difference in means of final exam scores of 7 percentage points with power=.80 alpha=.05. Assumption of sample means and standard deviation for power analysis was based on national norms for the final exam questions.

Study Administration and Management

Qualtrics, an on-line survey platform, was used for data collection. After completing the questionnaire, all qualified respondents meeting all the screen inclusion criteria were randomly assigned to one of the three courses. Then, each study participant received an email with a unique link for accessing the course. The emails and course access links were distributed through the Qualtrics survey platform, a system that enabled the researchers to track the amount of time students spent on each course component and require/enforce that they were devoting sufficient time to each.

Achievement Incentives

Following course completion, participants were asked to complete a 35-question final exam. Respondents who completed the final exam were provided a base compensation commensurate with the required course time to completion. To motivate students in all three test groups to strive for the highest scores they were capable of, we offered additional compensation to those who achieved certain performance benchmarks. Specifically, respondents who achieved a score of 30% correct or higher on the exam received a $10 bonus and respondents with a score in the top tenth percentile within their course group received an extra $20.

Measures

Effectiveness

In this study, the primary measure of educational effectiveness was student performance on questions drawn from the Comprehensive Assessment of Outcomes for a First Course in Statistics 4 (CAOS 4). This assessment was developed by the ARTIST (Assessment Resource Tools for Improving Statistical Thinking) project, an initiative funded by the National Science Foundation (NSF), and it covers the topics of data collection and design, graphical representations, variability, sampling variability, and tests of significance and bivariate data. The COAS 4 is a validated test designed to “assess students’ statistical reasoning after completion of any first course in statistics”. All questions use a forced choice format. The CAOS 4 demonstrates excellent reliability in a large national sample (n=23,6445) of undergraduate students, with a Cronbachs alpha of .78 (delMas et al., 2007; Garfield et al., 2006). The final exam for this study consisted of 35 question items from the CAOS test. The exam scores were normed to a 0-100 scale score.

Learner Satisfaction and Motivation

In addition to the final exam, a questionnaire for gauging learner satisfaction and motivations for learning was administered to all students in the study.

All students were asked to answer the question “How likely is it that you would recommend this course to a friend or colleague?” on a scale of 0 to 10, where 0 indicated “not at all likely” and 10 indicated “extremely likely”. Learners were also asked to rate the difficulty of the course, how well the course held their attention, and whether taking the course had changed their interest level in statistics (increased interest, decreased interest or no change in interest). In addition, learners were asked to rate the course on the following dimensions: its teaching effectiveness, how entertaining it was, how seriousness it was, how engaging it was, how boring it was, and the extent to which it provided good examples for learning statistics.

Learner motivation was assessed using sub-scales of the Academic Motivation Scale, College Version (AMS-C 28). The AMS-C was developed for use in research on education. The AMS-C demonstrates very good psychometric properties with overall internal consistency Cronbach alpha=.81 and test-retest reliability correlation=.79. The seven-factor structure has been confirmed using confirmatory factor analysis. The full 28 item scale assesses intrinsic motivation (3 sub-scales, 4 items each) and extrinsic motivation (3 sub-scales, 4 items each) and amotivation (1 sub-scale, 4 items) towards education (Utvaer at al. 2016, Vallerand et al, 1992). Two of the extrinsic motivation sub-scales were included in the survey: Extrinsic Motivation-Identified (EM-I) and Extrinsic Motivation-External Regulation (EM-ER). EM-I is based on the perceived usefulness of the behavior and is an assessment of the respondents desire to engage in an educational task to gain a sense of importance and personal value. EM-ER is the least autonomous form of motivation. EM-ER is initiated and regulated by external contingencies, such as avoiding negative consequences or to achieve rewards. This is particularly salient given the compensation structure for study participants.

In a validation study, The EM-I and EM-ER were found to be the most important forms of motivation to students. Both the EM-I and EM-ER demonstrate good internal consistency (Cronbach alpha=.62 and .83, respectively) and reliability (test-retest correlation EM-I=.71 and EM-ER=.83). Response categories for each of the items use a 5-point ranking, from 1=Does Not Correspond at All to 5=Corresponds Exactly. The 8 sub-scale items are presented in Appendix C. Learners were also asked to report the extent to which financial compensation was a reason they completed the study.

Statistical analysis

The analyses included data from all learners who completed surveys and final exams. Bivariate analysis, including within and across-course differences in demographics, completion rate and course pass rate, used ANOVA and cell chi-square. T-tests of difference of means were used to compare knowledge scores across courses. Univariate and multivariate regression models (OLS) were used to evaluate predictors of knowledge acquisition among all learners. All statistical analyses were conducted using SAS© 9.4.

Results

Study Enrollment and Course Completion

Data for this analysis was collected between April 29, 2017 and July 11, 2017. A total of 3,238 respondents clicked on the screening/ study information link, 1,620 completed the questionnaire for determining eligibility, and 924 of those who completed the questionnaire met the criteria for inclusion in the study. Individualized emails with the study link were sent to the 924 eligible respondents. 436 of the potential learners entered the study site, an open rate of 47.2%. Figure 1 indicates the number of learners randomly assigned to and completing each course. 234 of those who entered the study site completed their assigned course and the final exam within the 2 week interval allowed. Thus, the overall study completion rate was 49.2%.

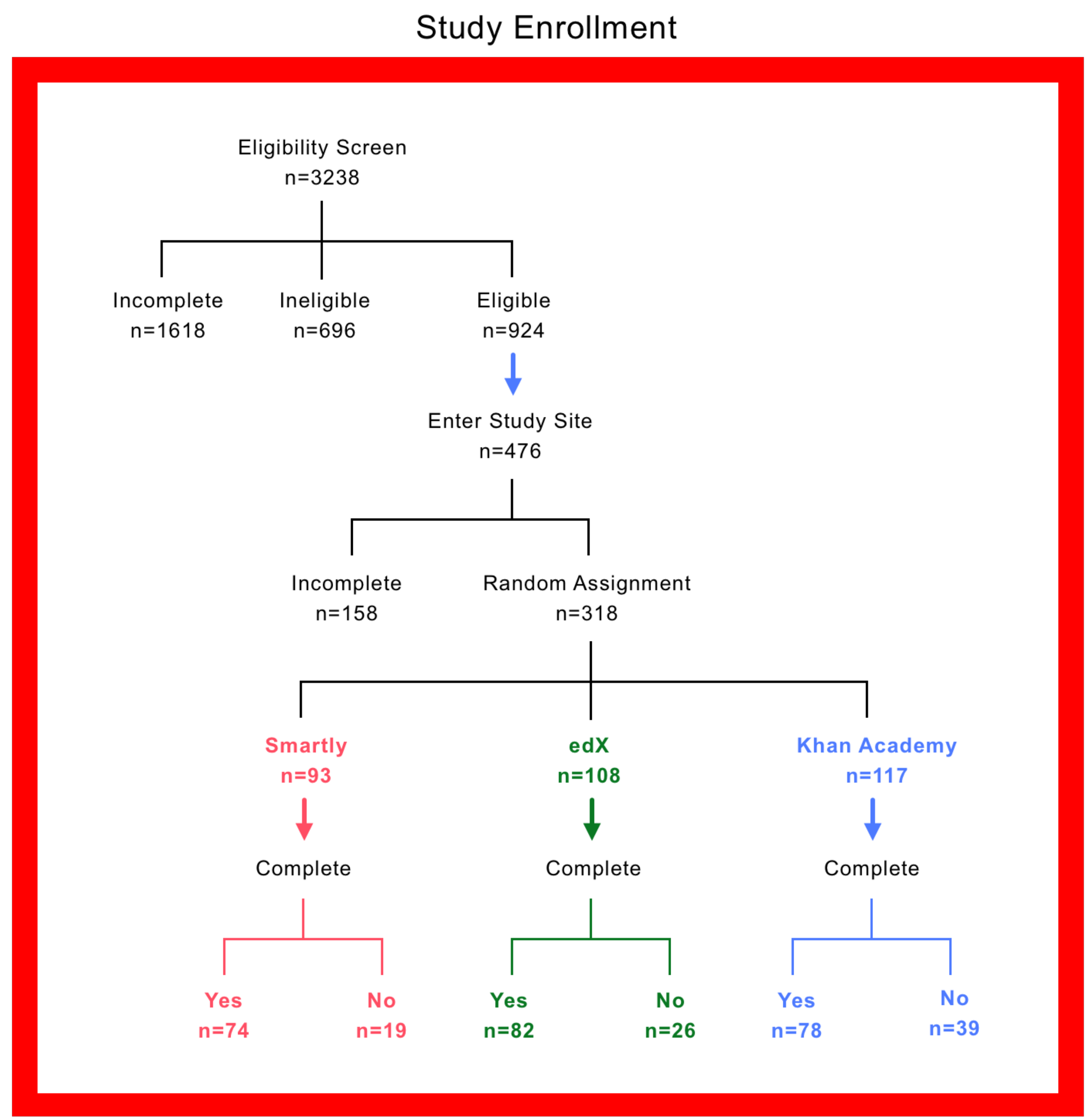

Figure 1. Study Enrollment

As shown in Table 2 below, the study completion rate varied across the test groups. Quantic learners had a significantly higher completion rate than learners assigned to either edX learners or Khan Academy learners. The first screen learners are directed to following randomization informs the learner of the length of time required to complete the assigned course as well as the compensation amount they will receive after course completion. Despite the higher compensation offered for completing the Khan Academy course, the completion rate for that course (66.6%) is significantly lower than the completion rate for the Quantic course (79.6%).

Regarding those who dropped out of the study, the amount of time spent on the study site before dropping out did not vary significantly across test groups, with most of the attrition occurring soon after learners started the course. Considering the higher compensation offered for completing the Khan Academy course, the low completion rate for Khan Academy learners may be an indicator of learners' unwillingness to engage in a course that requires almost 9 hours to complete. As an indicator of potential students' initial appraisal of the value or appeal of course completion, it is also possible that Khan Academy's low completion rate is an indication of learners' dissatisfaction with Khan Academy's visual and teaching format.

Table 2. Completion Rate by Course

	Quantic n (%)	edX n (%)	Khan Academy n (%)	Total n (%)
Course Completion after RA	74 (79.6)	82 (75.9)	78 (66.6)*	234 (73.6)
*Chi-square p≤0.05				

Demographic Differences Across Groups

230 respondents completed one of the statistics courses as well as the final exam within the allotted two-week window for course and exam completion. As displayed in Table 3 below, the proportion of learners who self-report as Latino is significantly higher in the edX group compared to the other two groups.

The percentage of learners with a high school education or less also differed significantly across groups: 39% of Quantic learners had a high school education or less, compared with 24.4% of edX learners and 26.9% of Khan Academy learners. The Quantic group also had a higher proportion of learners who are currently high school students.

Table 3. Study Participant Demographics (n (%))

	Quantic 74 (31.3)	edX 81 (34.8)	Khan Academy 78 (33.9)	Total n=230
Age (mean, std)	21.7 (6.9)	23.3 (7.1)	22.6 (6.2)	22.5 (6.3)
Gender (female)	47 (63.5)	54 (65.9)	48 (61.5)	149 (63.7)
Race				
White	37 (50)	41 (50)	39 (50)	117 (50)
African American	5 (6.8)	5 (6.1)	7 (9.0)	17 (7.3)
Asian	30 (40.5)	27 (32.9)	31 (39.7)	88 (37.6)
Other	3 (4.1)	2 (2.4)	1 (1.3)	6 (2.6)
Hispanic/Latino	4 (5.4)	12 (14.8)*	8 (10.3)	24 (10.3)
Education				
HS or less	29 (39.2)	20 (24.4)*	21 (26.9)^	70 (29.9)
Some college	17 (23.0)	22 (26.8)	26 (33.3)	65 (27.8)
Associates	4 (5.4)	3 (3.7)	3 (3.9)	10 (4.3)
Bachelors	22 (29.7)	28 (34.2)	20 (25.6)	70 (29.9)
Masters or Doctoral	2 (2.7)	9 (10.0)	8 (10.3)	19 (8.1)
Current student	41 (55.4)	43 (52.4)	36 (46.2)	120 (51.3)
HS^	5 (12.2)	1 (2.3)	0	6 (5.0)
College	31 (75.6)	32 (74.4)	30 (85.7)	93 (78.2)
Graduate/Professional	4 (9.8)	9 (20.9)	4 (11.4)	17 (14.3)
Other	1 (2.4)	1 (2.3)	1 (2.9)	3 (2.5)
On-line course experience	49 (66.2)	53 (64.6)	54 (69.2)	156 (66.7)

Chi-sq ^p≤0.10;*p≤0.05;**p≤0.01

Due to rounding columns may not equal 100%

Multiple categories allowed, total may exceed 100%

Knowledge Acquisition

Quantic Learners Had Higher Mean Exam Scores

Quantic learners demonstrated greater knowledge acquisition than both edX and Khan Academy learners did. Quantic learners had a mean final exam score of over 79% correct, at least 5 points higher than either edX or Khan Academy learners. Table 4 shows the mean exam scores (standardized) and standard deviations. T-tests comparing the final exam means between Quantic students and edX students found a statistically significant difference of 5.7 percentage points (p=0.0185). A comparison of Quantic students to Khan Academy students found a statistically significant difference of 4.9 percentage points (p=0.0319) in the final exam means. The lowest final exam score in the Quantic group was 42.9. The corresponding figures for the edX and Khan Academy groups were 28.6 and 34.3, respectively.

Table 4. Comparison of Quantic Learner Final Exam Scores to edX and Khan Academy Learner Final Exam Scores (t-test)

Course platform	Final Exam completes (n)	Mean Final Exam (std)	p-value	Median Final Exam	Min-Max Score
Quantic	74	77.8 (11.7)		80	42.9-97.1
edX	82	72.1 (16.0)	0.0185	74.2	28.6-100
Khan Academy	78	72.9 (15.6)	0.0319	77.1	31.4-100
t-test difference of means: Quantic to edX and Quantic to Khan Academy					

Quantic Exam Scores Have a Lower Standard Deviation

The standard deviation, a primary indicator of exam score variance, is significantly lower for Quantic learners than edX or Khan Academy learners. With a perfectly normal distribution, the smaller variance of Quantic final exam scores would indicate that a smaller proportion of Quantic learners score below the 70% threshold compared to edX or Khan Academy learners. However, tests for normality indicate that the final exam scores are not normally distributed for any of the three courses (F=0.0367), nor are the final exam score group variances equal. The final exam score distributions for all three courses are negatively skewed—i.e., the mean is less than the median, as shown in Figure 2. Test statistic values are provided below the graph.

Table 5. Test for normality of final exam distributions across courses

Course platform	Shapiro-Wilk W statistic (p)	Kolmogorov-Srminov D statistic (p)	Skewness
Quantic	0.9492 (0.0049)	0.1284 (0.0100)	-0.8157
edX	0.9559 (0.0067)	.0924 (0.0833)	-0.6964
Khan Academy	0.9630 (0.0230)	0.1578 (0.0100)	-0.5849

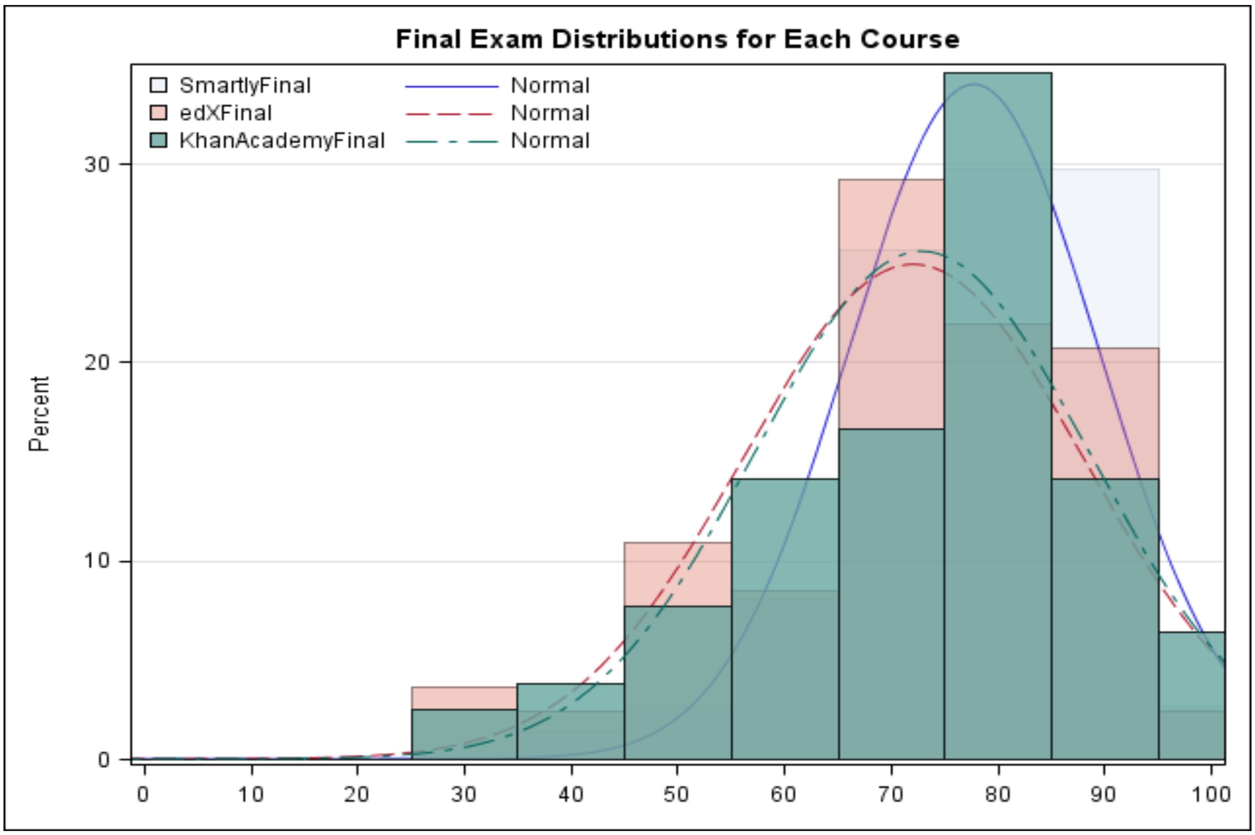

Figure 2. Final Exam distributions

Higher proportion of Quantic Learners with Exam Scores Over 70%

Despite the skewness and lack of normality evident in final exam score distributions, t-tests are considered robust to assumptions of normality, allowing us to have confidence in our findings. The lack of strict conformance to assumptions of normality precludes the ability to use final exam score standard deviations to estimate the proportion of learners in each course who scored 70% or more in a general scenario. Nevertheless, manual counts of the distribution of final exam scores indicate that a significantly higher proportion of learners in the Quantic course had a final exam score over 70% (78.4%) compared to the edX (61.0%; p=.0243) and Khan Academy (62.8%; p=.0367) courses. The difference between Quantic and edX was 17 percentage points, while the difference between Quantic and Khan Academy was 16 percentage points. Notably, in the study sample, the percentage of edX and Khan Academy students who scored below 50% correct on the exam was almost five times greater than the percentage of Quantic students who scored below that mark.

Table 6. Learners with score=70% or higher

Course	% Learners over 70%	p	% Learners scoring below 50%	p
Quantic	78.4		2.7	
edX	61.0	0.0243	12.2	0.0417
Khan Academy	62.8	0.0357	10.3	0.0604
Chi-sq. comparison of Quantic to edX and Quantic to Khan Academy				

Factors associated with Final Exam scores – Regression Analysis

We also ran regressions to quantify the effects of various factors—including the statistics course completed—on final exam scores.

Course as Dichotomous Variable

In our first regression model, we grouped edX and Khan learners into a single category of non-Quantic learners and coded *Course* as a dichotomous variable with the possible values "Quantic" or "Not Quantic". As shown in Table 7 below, the final exam scores of Quantic learners are over 5 percentage points higher than those of non-Quantic learners. Being a Quantic learner is the factor that most strongly influences the final exam score—its standardized estimate of 1.17 is significantly greater than that of the other variables.

Table 7. Unadjusted Regression Models: Predictors of Final Exam Score

	Parameter Estimate (se)	p	Standardized Estimate
Course	5.28 (2.05)	0.0108	1.17
Education (dichotomized as HS or less)	-1.97 (1.01)	0.0520	-0.12
Gender	5.40 (1.97)	0.0066	0.18
Importance of financial compensation	2.83 (1.04)	0.0071	0.18
Motivation: External regulation	0.09 (0.98)	0.93	0.01
Motivation: Identified	0.92 (1.00)	0.36	0.06

Surprisingly, learners above the higher school level do not perform as well as learners with a high school education or less. But, this unadjusted estimate may be inflated due to the greater proportion of Quantic learners who are current high school students.

Another finding from the analysis is that gender has a significant effect on final exam score, with females scoring over 5 percentage points higher than their male counterparts. This finding is consistent with the test results from the national sample of college undergraduates who comprised the sample to validate the CAOS test.

Financial compensation for completing the course also has a significant effect on final exam performance, but neither of the Extrinsic Motivation measures were significant factors.

Multivariate Regression Model – Control for Course

Next, we controlled for the *Course* variable and ran a multivariate regression using the variables Education, Gender, Financial Compensation, Motivation-External regulation, and Motivation-Identified. In this model, gender is still a significant driver of final exam performance (standardize estimate = 0.19) but financial compensation is the strongest factor influencing exam scores (standardized estimate = 0.25). See Table 8 below.

Table 8. Multivariate Adjusted Regression: Predictors of Final Exam (Controlling for Course)

	Parameter Estimate (se)	p	Standardized Estimate
Education (HS or less)	-3.13 (2.05)	.1296	-0.10
Gender	5.79 (1.98)	0.0039	0.19
Importance of financial compensation	3.91 (1.07)	0.0003	0.25
Motivation: External regulation	-2.03 (1.46)	0.1658	-0.14
Motivation: Identified	3.10 (1.53)	.04370	0.21
$F=4.83$ ($p<0.0001$) $R^2=0.1188$ Adjusted $R^2=0.0850$			

Notably, the offer of financial compensation (up to $125 to complete a course) was a necessary component of our study design to ensure that enough subjects participated in the study, but it does not have such direct relevance in a formal educational setting.

Learner Satisfaction

No Differences in Perceptions of Course Difficulty

Learner perceptions of course difficulty did not vary significantly across test groups. As shown in Figure 3, the majority of learners in all three courses rate the difficulty level of the course to be “about right”.

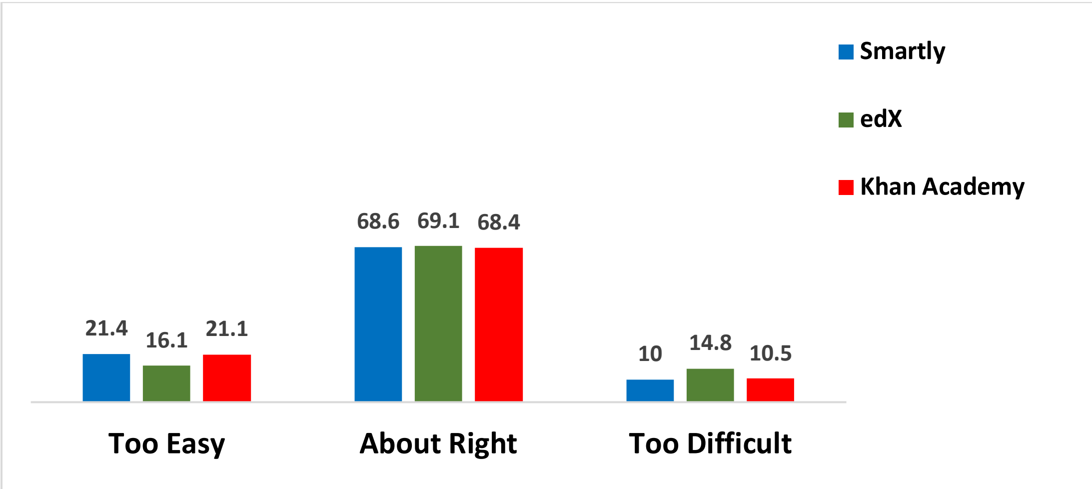

Figure 3. Difficulty Level of Course (p=0.8016)

Quantic Learners More Likely to Recommend the Course to Others

All course participants were asked to respond (at the conclusion of the study) to a standard satisfaction question, “How likely are you to recommend Quantic to a friend or colleague?” with a response scale of 0 = “would not recommend” to 10 = “definitely would recommend”. Taking a conservative approach, we consider tail scores of 0, 1, or 2 as strongly not recommending, and tail scores of 8, 9, 10 as strongly recommending.

Table 9. Learners Recommendation of Course

Course platform	Would NOT Recommend	Would Recommend
Quantic	4.1%	32.4%
edX	12.2%	15.9%
Khan Academy	10.3%	23.1%

* chi-sq. p≤0.10

The difference in course satisfaction is evident in the tails of the response distribution to the recommendation question. As noted in Table 8 above, almost 4 times higher percentage of edX learners than Quantic learners would not recommend the course an almost 3 times greater percentage of Khan Academy learners than Quantic learners would not recommend the course, while over twice the percentage of learners recommend the Quantic course over the edX and Khan Academy courses.

Quantic Course Better at Holding Student Attention

Course completers were also asked to indicate whether the course was able to hold their attention well. A significantly higher proportion of Quantic students felt the course held their attention well (82.4%) than did edX (57.6%) or Khan Academy (63.2%) students (p=0.0039). Figure 4 presents student comparative course ratings along this dimension. An interesting, though not necessarily surprising, finding is that learner ratings of how well a course held their attention is not associated with learner final exam scores.

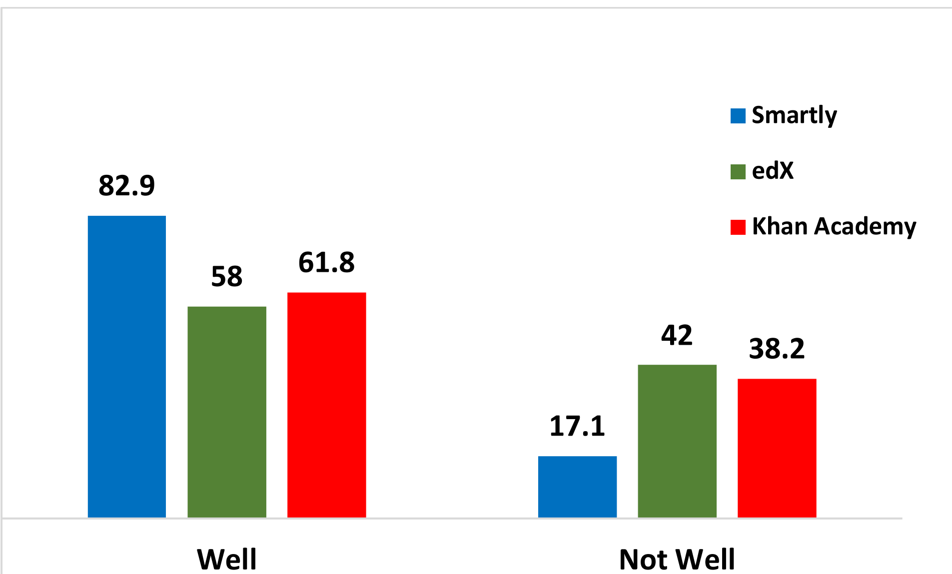

Figure 4. How well did Course Hold Attention (p=0.0020)

Quantic Perceived as More Entertaining

A large majority of Quantic learners (82.4%) rate the course as “entertaining” (82.4%) and “engaging” (88.6%). In contrast, only 38.8% of edX learners and 48% of Khan Academy learners rated their courses as entertaining and engaging (edX=56.8%, Khan Academy=79%). The differences are statistically significant (p<.0001).

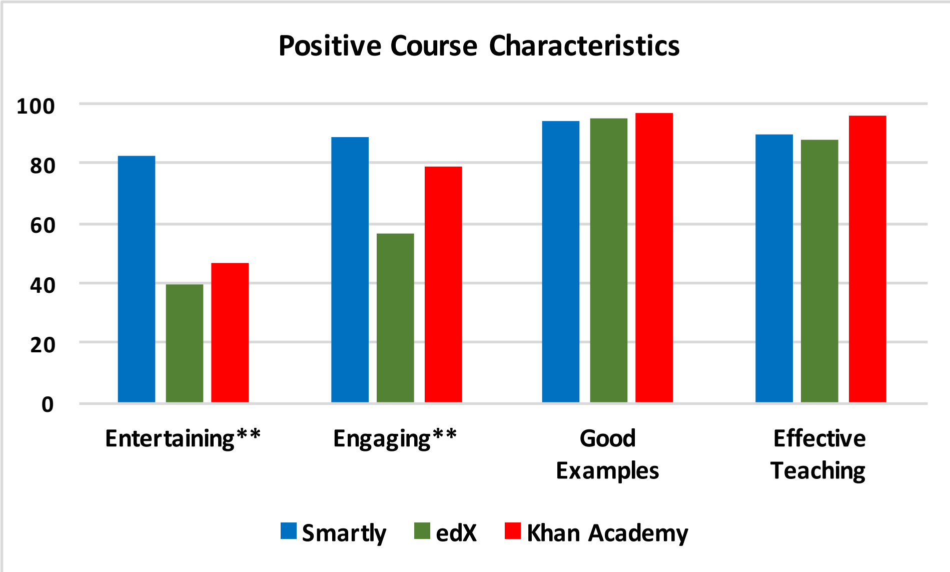

Figure 5. Positive Course Characteristics

Quantic Perceived as Less Boring

Conversely, 73.8% of edX learners and 59.2% of Khan Academy learners rated their courses as “boring”—while only 28.4% of Quantic learners rated it so. These differences are also statistically significant (p<.0001). In addition, less than 3% of Quantic learners perceived the course as “overly serious” in tone, while 38.3% of edX learners and 18.4% of Khan Academy learners perceived their courses as “overly serious”.

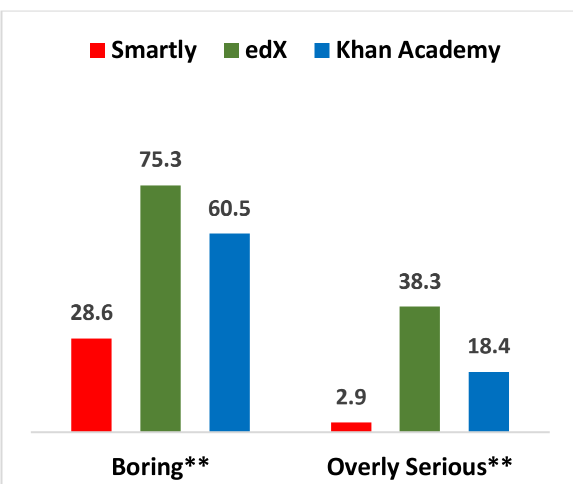

Figure 6. Negative Course Characteristics (p≤0.01)

Learner Motivation

Students' responses to the survey questions for measuring their extrinsic motivations for learning did not correlate strongly with their course completion rates. This result is not surprising considering that the learners were not offered any academic credit for completing their assigned courses.

Since a very high percentage of learners across all courses indicated that the offer of financial compensation (78.2%) was a strong impetus for completing the course, it may be unreasonable to expect other forms of motivation to be significant factors in this scenario.

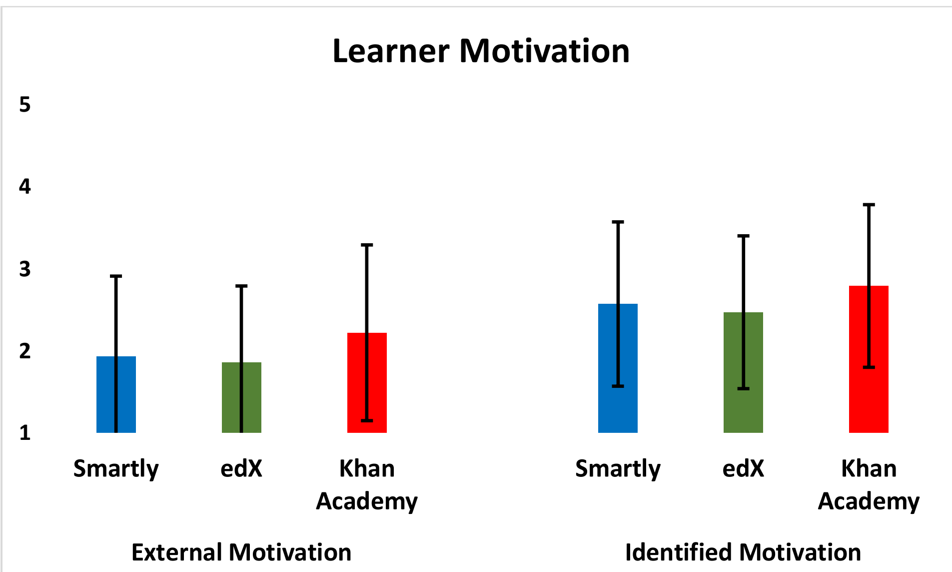

Figure 7. Learner Motivation (External Motivation, p=0.0597; Identified Motivation, p=0.1141)

Discussion

This study uncovered strong evidence that the Quantic online education platform is superior to the edX and Khan Academy online platforms for learner knowledge acquisition and learner satisfaction. Considering that students can complete the Quantic statistics course in half the time of the edX course and in one-fourth the time of the Khan Academy course, the findings should prompt students and educators to consider Quantic as an alternative to these more popular courses.

This superior performance of the Quantic learners compared to the edX learners may be attributable to the teaching method used in the edX course: the presentation of recorded videos without any interactive exercises to engage the learner—an approach that is typical of most MOOCs. The drawbacks of this static and passive methodology are clearly evident in the words of an edX student who had a score of 49% on the final exam:

> *I felt pretty confident with the course til I got to the final. I wished it would have had some more interactive examples do it, I feel if I had a change to do some problems as practice it would not have crushed my spirit as much after taking the final.*

Frustration with receiving too much information at once, without the opportunity to learn each statistical concept incrementally through practice exercises, was a recurring theme in the comments of edX learners:

> *I find that learning one concept at a time and then testing my competence leads to a good understanding of the topic at hand. The lack of small areas of study followed my testing my new knowledge made it difficult for me to follow with the presenters.*

The study data indicate that the course completion rate for the Quantic online statistics course (79.6%) was significantly higher than the course completion rate for the Khan Academy online statistics course (66.6%)—even though the offered compensation for completing the Khan Academy course was 111% greater than that for the Quantic course. One Khan Academy learner noted:

While Khan Academy is great for reviewing certain concepts, I personally don't think it should be used as standalone courses because the similar videos quickly lose the engagement factor and the exercises do not require creativity.

Based on the last comment and others similar to it, as well as findings from other studies (Rieber, 2016), we infer that the shorter amount of time required to complete the Quantic course is a primary reason for the 13% higher course completion rate compared to the Khan Academy course.

Comparison with National CAOS Sample Test Results

Interestingly, the mean final exam score for learners in the study was significantly higher than the mean scores for the large sample of undergraduate students who participated in the validation study of the CAOS. Specifically, the mean final exam score for that study, administered to a sample of students who were enrolled in an undergraduate statistics course, was 55.8 with a standard deviation of 16.1.

(A subsequent study to confirm the factor structure of CAOS employed a sample of 23,645 students, comprised of undergraduates who had taken college statistics courses and high school students enrolled in Advanced Placement statistics courses. The mean exam scores collected in that nine-year longitudinal study were stable over time, staying close to the 50% correct level.)

The large disparity in mean final exam scores between the students in our study and the students in the original CAOS study has several possible explanations. The most glaring is the learning environment. Students in the CAOS study were all enrolled in statistics courses taught in a structured classroom setting, whereas the learners in our study used self-pace online learning solutions.

A second possible explanation is age: the students in our study were on average older than the students in the CAOS study. However, age is most likely not a determining factor because students in our study with a high school education or less actually outperformed their counterparts with higher levels of education. Further, in unadjusted regression models, learner age was not found to be a significant predictor of success on final exam score.

A third possible explanation is the large variance in CAOS mean exam scores across sub-topics. For instance, CAOS mean scores for the data collection and sample variability sub-modules are in the range of 38%-45% correct while the mean scores for the variability and the bivariate data sub-modules are in the range of 66%-70% correct. Yet, this factor also does not explain the disparity between our student test scores and the CAOS student test scores because the bivariate data sub-topic was not part of our test curriculum, while data collection and sample variability were in the curriculum.

Study Limitations

The implications of this study are limited to a certain extent by the sample. While over half of the learners in this study were students in a diploma or degree program, none were offered a course credit that can be applied towards earning an academic credential. This factor precludes our ability to generalize to the population of interest for online courses within degree programs.

Also, since participants in the study were offered financial compensation to complete their courses, we cannot infer what the course completion rates would be in a more typical scenario—i.e., one in which students pay to take courses and are not paid to complete them.

References

Allen, I. E., & Seaman, J. (2016). Online Report Card: Tracking Online Education in the United States. Babson Survey Research Group.

Aliaga, M., Cuff, C., Garfield, J., Lock, R., Utts, J., & Witmer, J. (2005). Guidelines for assessment and instruction in statistics education (GAISE): College report. Alexandria, VA: American Statistical Association.

Barak, M., Watted, A., & Haick, H. (2016). Motivation to learn in massive open online courses: Examining aspects of language and social engagement. Computers & Education, 94, 49-60.

delMas, R. (July, 2014). Trends in students' conceptual understanding of statistics. In K. Makar, B. de Sousa, & R. Gould (Eds.), Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA. Voorburg, The Netherlands: International Statistical Institute.

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. Statistics Education Research Journal, 6(2), 28-58.

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. Statistics Education Research Journal, 6(2), 28-58.

Garfield, J. B., et al. "Comprehensive Assessment of Outcomes in a first course in Statistics (CAOS)." Retrieved October 20 (2006): 2007.

Kvam, P. H. (2000). The effect of active learning methods on student retention in engineering statistics. The American Statistician, 54(2), 136–140.

Lockwood, C. A., Ng, P., & Pinto, J. (2007). An interpretive business statistics course encompassing diverse teaching and learning styles. Academy of Educational Leadership Journal,11(1), 11–23.

National Center for Education Statistics' Integrated Postsecondary Education Data System (IPEDS), *accessed 7/30/2017.*

Parr, W. C., & Smith, M. A. (1998). Developing case-based business statistics courses. The American Statistician, 52(4), 330–337.

Phan, T., McNeil, S. G., & Robin, B. R. (2016). Students' patterns of engagement and course performance in a Massive Open Online Course. Computers & Education, 95, 36-44.

Rieber, L. P. (2016). Participation patterns in a massive open online course (MOOC) about statistics. British Journal of Educational Technology.

William R.. Shadish, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Wadsworth Cengage learning; 2002.

Tintle, N., Topliff, K., VanderStoep, J., Holmes, V. & Swanson, T. (2012). Retention of statistical concepts in a preliminary randomization-based introductory statistics curriculum. Statistics Education Research Journal, 11(1), 21-40.

Utvær, Britt Karin Støen, and Gørill Haugan (2016). "The Academic Motivation Scale: Dimensionality, Reliability, and Construct Validity Among Vocational Students." Nordic Journal of Vocational Education and Training 6.2: 17.

Vallerand, Robert J., et al. (1992). "The Academic Motivation Scale: A measure of intrinsic, extrinsic, and amotivation in education." Educational and psychological measurement 52.4: 1003-1017.

Appendix A: Website URLs used for recruitment

Group / Site	Source Appended
Survey Circle	surveycircle
Facebook Business School Groups	FBBschool
Facebook Medical School Groups	FBMedschool
Facebook Law School Groups	FBlaw
Facebook Pre-Med Groups	FBpremed
Facebook Undergraduate Groups	UGschool
Handshake (Job Board	handshake
LinkedIn	linkedin
Craigslist	craigslist
Facebook Post Doc and Phd Groups	FBphdpstd
FB Job Boards	fbjob
Facebook College Alumni	alumn
Facebook Market & Housing	fbmark
Reddit – Participants Board	Redt
Georgetown University student board	GTU
Harvard University Careers Portal	HVD
Facebook Engineering Groups	fbeng
Facebook Air Force Group	fbaf
Shared interest – medical health problems	fbhlth
Facebook – broad interests	fbbroad
Call for Participants Website	cfp

Appendix B: Course modules and Time to complete

Course Modules	Completion Time (seconds)	URL
Quantic 1	1800	https://smart.ly/research-study17/join/account?target=/course/learn-data-collection--Quantic/d79316c4-a419-41ff-ae42-be14fe70af44
Quantic 2	2700	https://smart.ly/research-study17/join/account?target=/course/learn-one-variable-statistics--Quantic/8d720fb5-0320-4bf4-8c40-50808267a1be
Quantic 3	2700	https://smart.ly/research-study17/join/account?target=/course/learn-probability-fundamentals--Quantic/e08c8202-595b-462a-a161-534e40ab7448
EdX 1	5216	https://www.youtube.com/playlist?list=PL2fCZiDqOYYUHoIH4AWVHuwyEet3_xPZ7
EdX 2	6600	https://www.youtube.com/playlist?list=PL2fCZiDqOYYWHVVeHGTnIK_C3pB3W-o-M
EdX3	5264	https://www.youtube.com/watch?v=_UmyvpSnZOc&list=PL2fCZiDqOYYWpx4cgPlm-cMs4i4XdMlBK
Khan 1	1020	https://www.khanacademy.org/math/statistics-probability/displaying-describing-data/statistics-overview/v/understanding-statistical-questions
Khan 2	1440	https://www.khanacademy.org/math/statistics-probability/displaying-describing-data/categorical-data-displays/v/reading-bar-graphs
Khan 3	1560	https://www.khanacademy.org/math/statistics-probability/displaying-describing-data/two-way-tables-for-categorical-data/v/two-way-frequency-tables-and-venn-diagrams
Khan 4	900	https://www.khanacademy.org/math/statistics-probability/displaying-describing-data/dot-plots-frequency-tables/v/ways-to-represent-data
Khan 5	660	https://www.khanacademy.org/math/statistics-probability/displaying-describing-data/histogram/v/histograms-intro
Khan 6	1260	https://www.khanacademy.org/math/statistics-probability/displaying-describing-data/comparing-features-distributions/v/shapes-of-distributions
Khan 7	1260	https://www.khanacademy.org/math/statistics-probability/displaying-describing-data/mean-median-basics/v/statistics-the-average
Khan 8	1740	https://www.khanacademy.org/math/statistics-probability/displaying-describing-data/more-mean-median/e/calculating-the-mean-from-various-data-displays
Khan 9	1020	https://www.khanacademy.org/math/statistics-probability/displaying-describing-data/range-iqr-mad/v/range-and-mid-range
Khan 10	1500	https://www.khanacademy.org/math/statistics-probability/displaying-describing-data/box-whisker-plots/v/box-and-whisker-plot-exercise-example

Khan 11	3180	https://www.khanacademy.org/math/statistics-probability/displaying-describing-data/pop-variance-standard-deviation/v/range-variance-and-standard-deviation-as-measures-of-dispersion
Khan 12	2200	https://www.khanacademy.org/math/statistics-probability/probability-library/basic-theoretical-probability/v/basic-probability
Khan 13	900	https://www.khanacademy.org/math/statistics-probability/probability-library/probability-sample-spaces/v/events-and-outcomes-3
Khan 14	1080	https://www.khanacademy.org/math/statistics-probability/probability-library/experimental-probability-lib/v/experimental-probability
Khan 15	1860	https://www.khanacademy.org/math/statistics-probability/probability-library/basic-set-ops/v/intersection-and-union-of-sets
Khan 16	1200	https://www.khanacademy.org/math/statistics-probability/probability-library/addition-rule-lib/v/probability-with-playing-cards-and-venn-diagrams
Khan 17	3480	https://www.khanacademy.org/math/statistics-probability/probability-library/multiplication-rule-independent/v/compound-sample-spaces
Khan 18	1980	https://www.khanacademy.org/math/statistics-probability/probability-library/multiplication-rule-dependent/v/introduction-to-dependent-probability
Khan 19	660	https://www.khanacademy.org/math/statistics-probability/probability-library/conditional-probability-independence/v/calculating-conditional-probability
Khan 20	1260	https://www.khanacademy.org/math/statistics-probability/designing-studies/sampling-and-surveys/v/reasonable-samples
Khan 21	1560	https://www.khanacademy.org/math/statistics-probability/designing-studies/experiments-stats-library/v/correlation-and-causality

Appendix C. AMS-C Items

External Motivation	mean (std)
Because without some statistics I would not find a high paying job later on	1.94 (1.03)
Because I think that a statistics course will help me better prepare for the career I have chosen	2.69 (1.19)
Because eventually it will enable me to enter the job market in a field that I like	2.28 (1.17)
Because I want to have "the good life" later on.	2.04 (1.20)
Because this will help me make a better choice regarding my career orientation.	2.26 (1.23)
In order to have a better salary later on.	2.08 (1.91)
Because I believe that a little additional time on statistics will improve my competence as a worker.	3.22 (1.21)
The financial compensation to complete the study	4.13 (0.92)
Response values range from 1=Does Not Correspond at All to 5=Corresponds Exactly	

Appendix D: Distributions

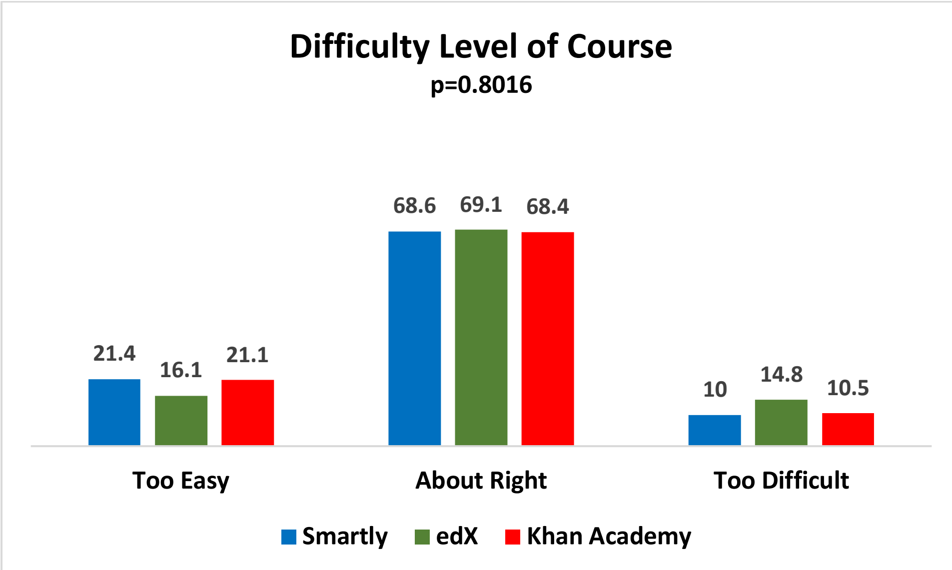

Figure 1

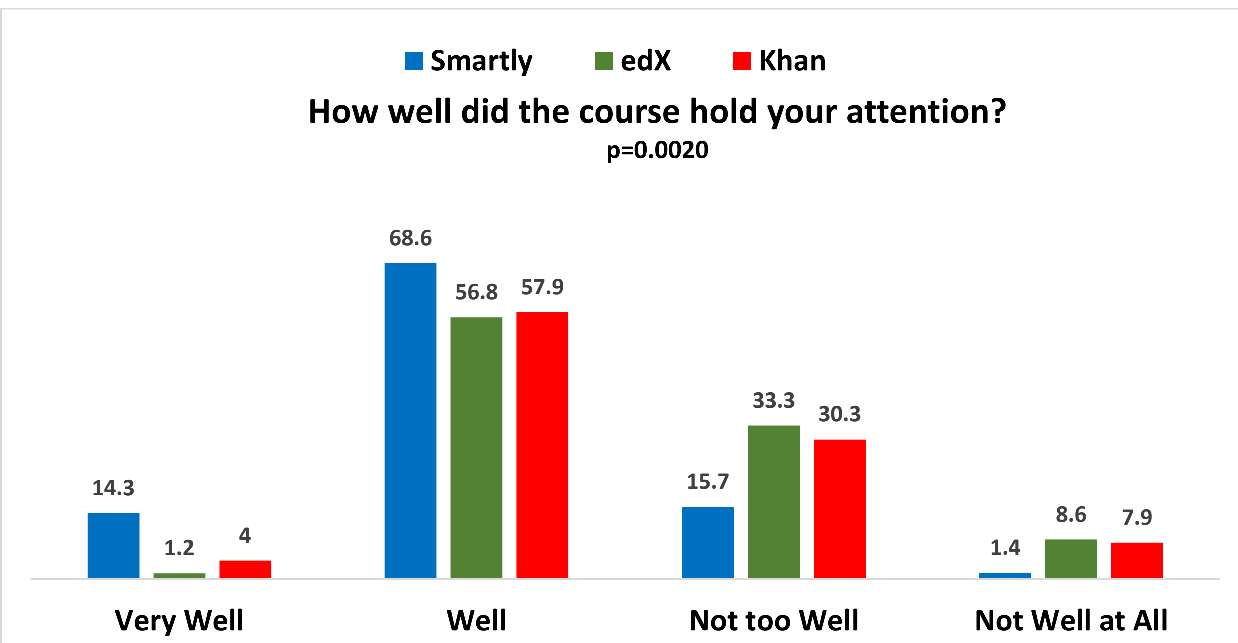

Figure 2

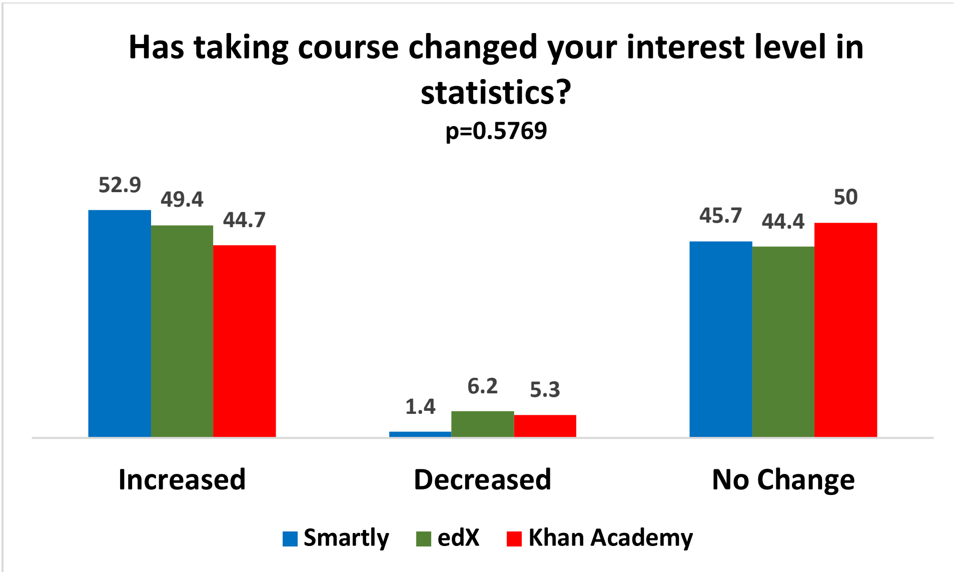

Figure 3

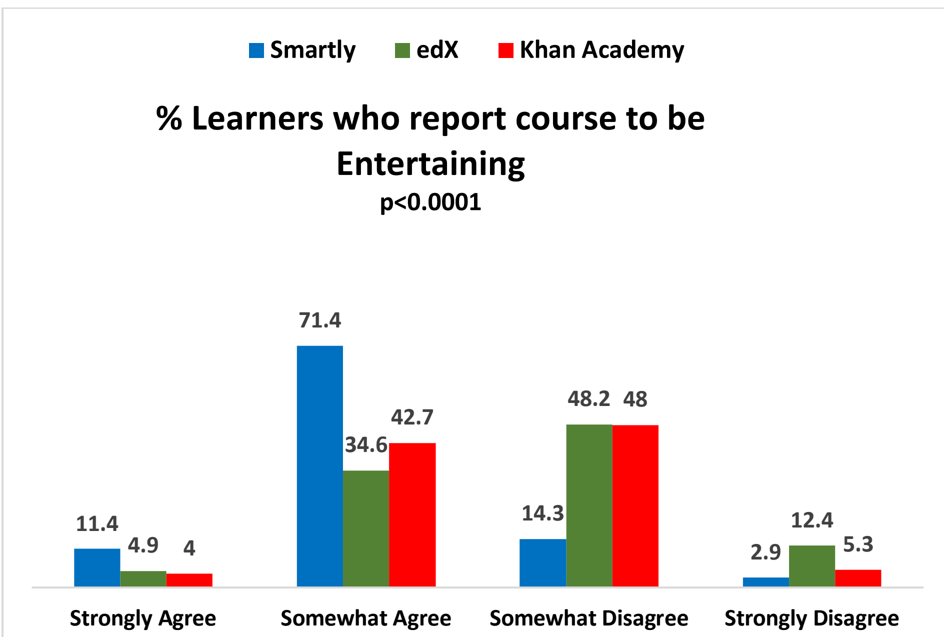

Figure 4

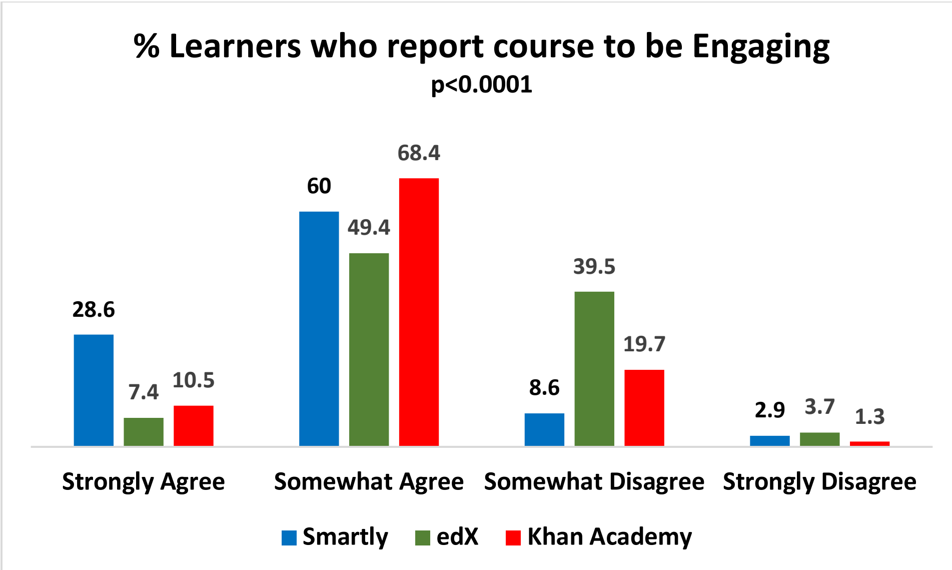

Figure 5

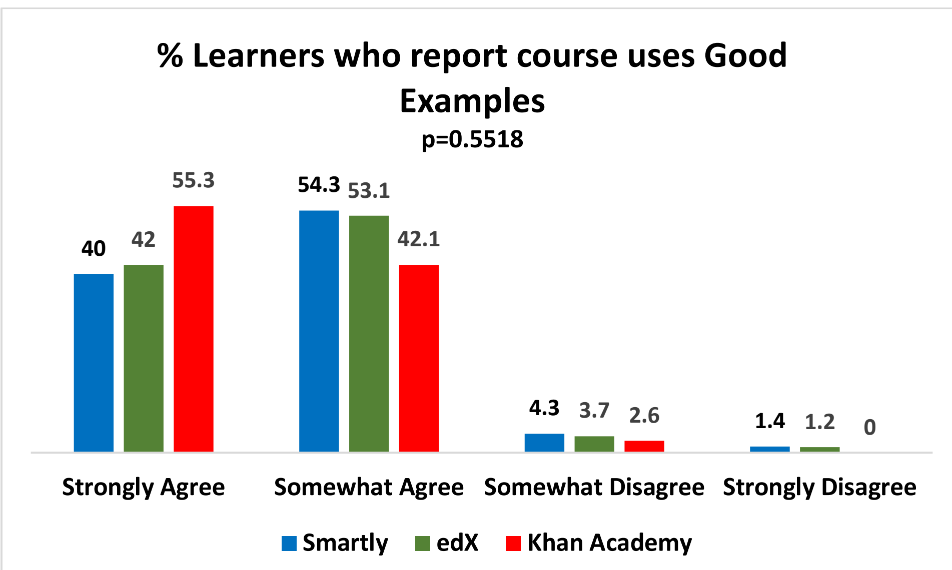

Figure 6

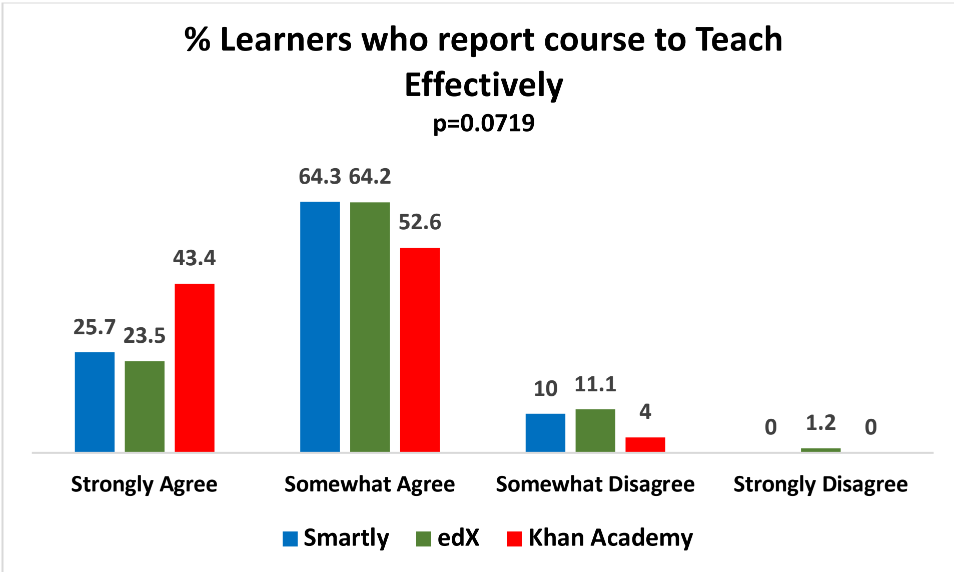

Figure 7

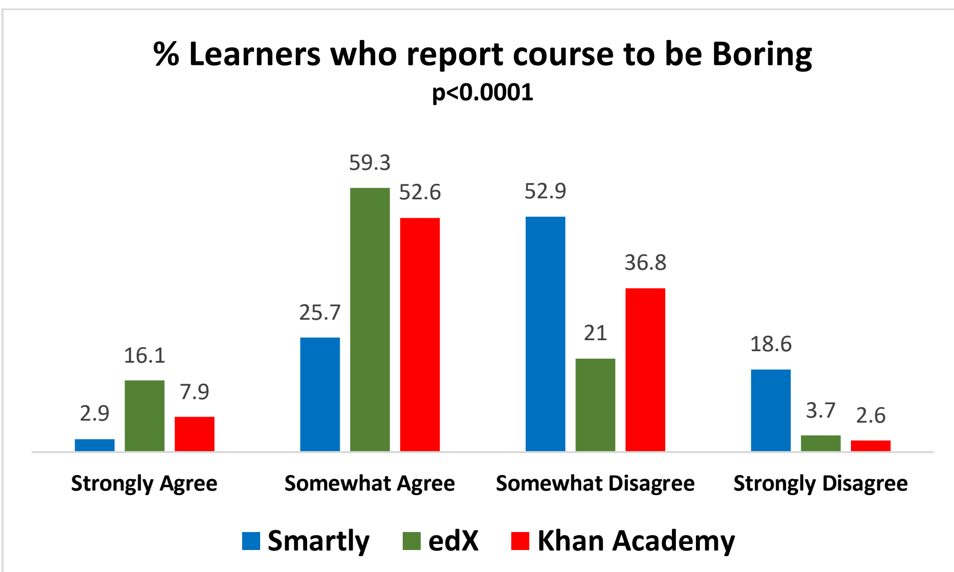

Figure 8

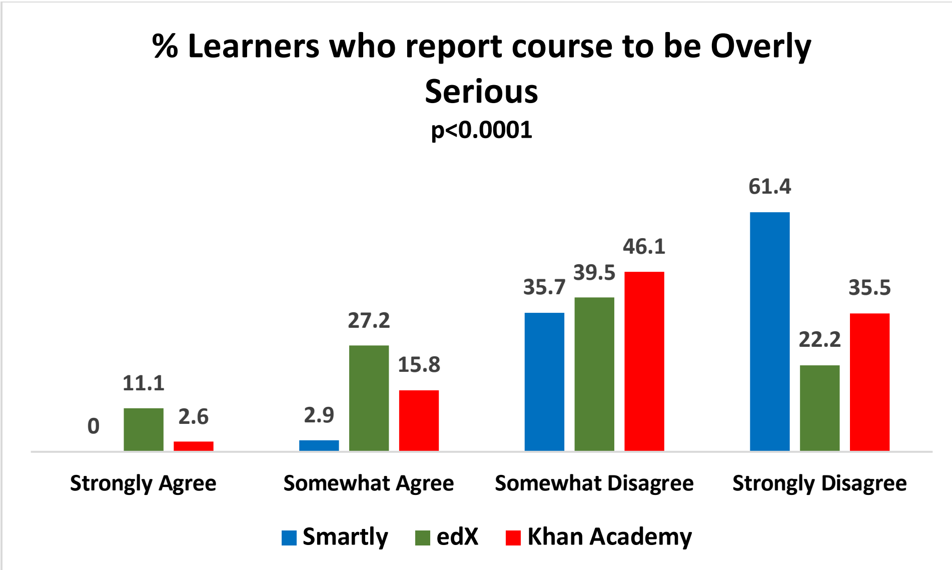

Figure 9

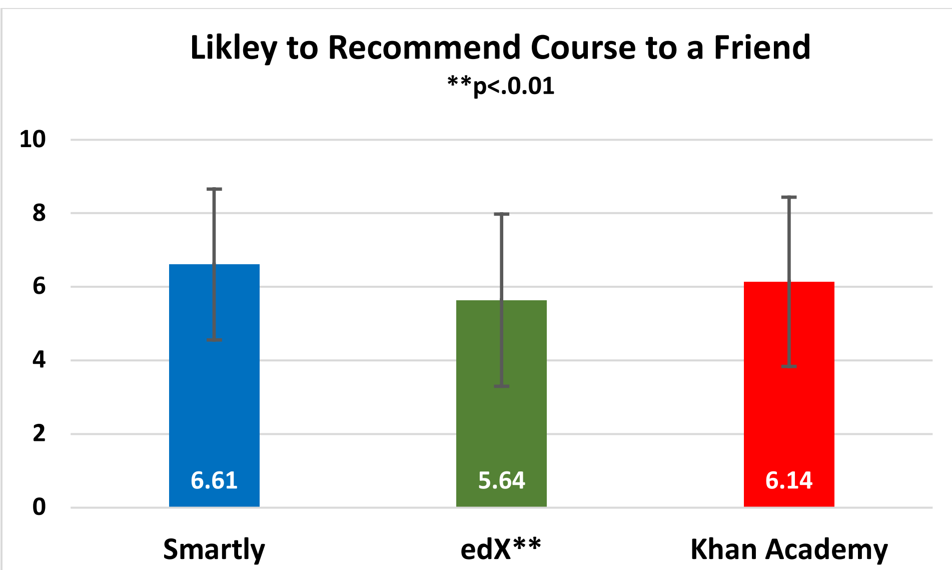

Figure 10