

Evaluation of Beast Academy

A math program serving gifted, talented,
and historically marginalized students in
Mankato Area Public Schools

Thomas Torre Gibney
Juan Carlos Bojorquez

August 2022

© 2022 WestEd. All rights reserved.

Torre Gibney, T., & Bojorquez, J. C. (2022). *Evaluation of Beast Academy: A math program serving gifted, talented, and historically marginalized students in Mankato Area Public Schools*. WestEd.

The evaluation of Beast Academy in Mankato Area Public Schools and the development of this brief were funded by the Art of Problem Solving (AoPS). The content does not necessarily reflect the views or policies of AoPS.

WestEd is a nonpartisan, nonprofit agency that conducts and applies research, develops evidence-based solutions, and provides services and resources in the realms of education, human development, and related fields, with the end goal of improving outcomes and ensuring equity for individuals from infancy through adulthood. For more information, visit [WestEd.org](https://www.wested.org). For regular updates on research, free resources, solutions, and job postings from WestEd, subscribe to the E-Bulletin, our semimonthly e-newsletter, at [WestEd.org/subscribe](https://www.wested.org/subscribe).



Contents

Introduction	1
Evaluation Approach	2
Analytic Samples	3
Evaluation Findings	4
Student Achievement Results	4
Student Survey Results	6
Reflection on the Results	8
Appendix	11

Introduction

Beast Academy is a comics-based math program for students ages 6–13 that includes online and paper-based components and is used throughout the country. Beast Academy focuses on mastery of math skills instead of memorization, utilizing games and puzzles to develop students’ understanding and problem-solving skills in math. This brief reports on an independent evaluation of the efficacy of the Beast Academy program in Mankato Area Public Schools (MAPS), a school district serving approximately 8,500 K–12 students in Mankato, Minnesota.¹ Elementary and middle school MAPS classroom teachers used Beast Academy to provide instruction and enrichment to two groups: “Cluster” students, consisting of students the district has designated as gifted and talented, and “Rising Scholar” students, high-ability learners from populations that have been historically marginalized in programming for academically talented students, including those the district refers to as culturally diverse, economically diverse, linguistically diverse, and twice exceptional students.

WestEd’s evaluation focused on Beast Academy’s effect on math achievement among the population of MAPS students who began the study enrolled in grades 3, 4, and 5 in school year (SY) 2019/20 and remained enrolled in the district through grades 4, 5, and 6 in SY 2020/21. This primary analysis was complemented by survey analysis investigating students’ attitudes and perceptions toward math during the SY 2020/21. The evaluation team sought to answer the following questions:

1. Does 2 years’ participation in Beast Academy have a positive impact on students’ math achievement?
2. Does the impact of Beast Academy differ for Cluster and Rising Scholar students?
3. To what extent do students who used Beast Academy in SY 2020/21 report different attitudes and perceptions toward math compared with their peers who did not use Beast Academy?

Results from the evaluation indicated that the use of Beast Academy had a positive, statistically significant effect on math achievement after 2 years of exposure to the program. The magnitude of the effect varied depending on students’ Cluster or Rising Scholar status, with the largest effects seen for Rising Scholar students. Effects were also greater for students with

¹ The Art of Problem Solving (AoPS), developers of Beast Academy, contracted WestEd to conduct the independent evaluation.

higher levels of use on the Beast Academy Online platform, as measured by the number of lessons attempted over the 2-year intervention period. The higher the number of lessons attempted, the larger the gap in math achievement compared to students who had no exposure to Beast Academy Online. Additionally, the evaluation team found that students who reported using Beast Academy in SY 2020/21 had more positive attitudes and perceptions toward math on six attitudinal domains measured on a student survey.

This study occurred during the COVID-19 pandemic, and the pandemic's impact on students' learning environments and on the district's use of Beast Academy cannot be overstated. Although WestEd maintained a rigorous design and supplemented study-collected data about Beast Academy exposure with data from the district and AoPS to ensure fidelity of treatment/comparison designation, caution is advised in interpreting the results given that these data were collected during an unusual time in students' education.

The purpose of this brief is to provide an easily accessible summary of evaluation findings and study highlights. For more information about the technical details of the study, including design and methodology, matching procedures, treatment determination, analysis, and detailed results, please contact the authors of this brief.

Evaluation Approach

WestEd's evaluation estimated the impact of 2 years of Beast Academy programming on student achievement in math in the MAPS district. All students enrolled in grades 3, 4, and 5 in SY 2019/20 who remained enrolled in the district through grades 4, 5, and 6 in SY 2020/21 were eligible to participate in the study. Treatment was defined as having a record of using Beast Academy in each of the 2 school years, as measured by district records and total time spent on math learning activities on the online platform. Students with no record of Beast Academy use in either year were identified as potential comparison students. Because the intervention period covered 2 years, students with a partial record of Beast Academy use—that is, usage in only 1 of the 2 years—were excluded from the analyses given that they did not meet the definition of treatment *and* could not be selected as comparison students due to the potential for contamination. To estimate the impact of Beast Academy on achievement, WestEd analyzed baseline and outcome data from the NWEA Measures of Academic Progress (MAP) interim assessments in math, which were administered by the district.² The evaluation also supplemented the impact analysis by examining the math-based attitudinal outcomes associated with Beast Academy participation based on surveys in SY 2020/21 that captured students' attitudes and perceptions related to math, such as their perceived sense of self-efficacy to learn and do difficult math work.

² WestEd was originally contracted to evaluate the impact of Beast Academy over a 1-year period using state assessment data. Due to the disruption in statewide testing as a result of the COVID-19 pandemic, the evaluation plan was adjusted to accommodate a 2-year study period using the NWEA interim assessment data.

In each analysis, WestEd sought to determine the extent to which math performance or attitudinal patterns differed between students who used the Beast Academy program and their peers who were not exposed to the program. WestEd’s analytic methods included statistical modeling approaches that isolated the effect of Beast Academy usage on the outcomes while controlling for other factors that could be related to the results. Usage was defined in several ways. First, WestEd evaluators received anonymized programmatic data from the district that identified students as either Cluster or Rising Scholar Beast Academy participants. To obtain a more nuanced definition of treatment, the evaluation team also received detailed usage data from the district, which contained student-level information on the number of lessons attempted and time spent doing math-related work on the Beast Academy Online platform. Together, these data allowed WestEd to estimate the impact of Beast Academy on achievement for students who used the program in greater or lesser “dosages” over SY 2019/20 and SY 2020/21. WestEd estimated separate models to describe (a) the effect of any Beast Academy use on students’ math achievement, (b) the differential effect of Beast Academy use for Cluster students and Rising Scholar students, and (c) the effect of Beast Academy on students’ math achievement at different levels of Beast Academy use. For analysis of different levels of use, WestEd created several “use profiles” of treatment students based on the distribution of lessons attempted during the intervention period. These “use profiles” were based specifically on use data from students in the MAPS district. For the survey analysis, students were identified as Beast Academy users based on a self-reported measure of participation in Beast Academy over the 1-year period of SY 2020/21.

Analytic Samples

Achievement Analysis. Using SY 2018/19 as the baseline year (the year prior to the study intervention period), WestEd implemented a rigorous matched comparison group design to pair Beast Academy students enrolled in grades 2 through 4 with non-Beast Academy students who shared similar characteristics at the start of the study, including their baseline math achievement levels. The matching process resulted in a weighted sample of 1,094 students³ who were well balanced between treatment and control groups and equivalent at baseline on average achievement and demographic characteristics (treatment mean score of 212.11 on the spring NWEA MAP assessment in math, comparison mean score of 212.64; $p = 0.707$). In addition to meeting baseline equivalence on the pretest measure of achievement, the matched sample was also equivalent on all other demographic characteristics in the matching model, including race/ethnicity, gender, English learner status, special education status, and free-or-reduced-price lunch eligibility. The analysis compared the math outcomes of both groups

³ WestEd used one-to-one nearest-neighbor matching with replacement to identify the matched comparison group sample that was most balanced in terms of baseline achievement and demographic characteristics. This approach resulted in some comparison students being matched more than once with selected treatment students. The statistical models upon which all results are based included analytic weights to account for multiple-matched observations.

2 years later to determine the extent to which any observed differences between the two could be attributed to exposure to the Beast Academy program.

Student Survey Analysis. The analytic samples for the student survey analysis were limited to SY 2020/21. The survey was administered in spring 2021 to MAPS students in grades 4, 5, and 6 whose parents had previously consented to their participation in the study. Students were asked to confirm their assent prior to answering the survey questions. WestEd received a total of 1,061 usable responses to the survey. On each of the attitudinal domains examined, the analysis compared the average responses of students who reported using Beast Academy during SY 2020/21 with the average responses of those who said that they did not.

Evaluation Findings

Student Achievement Results

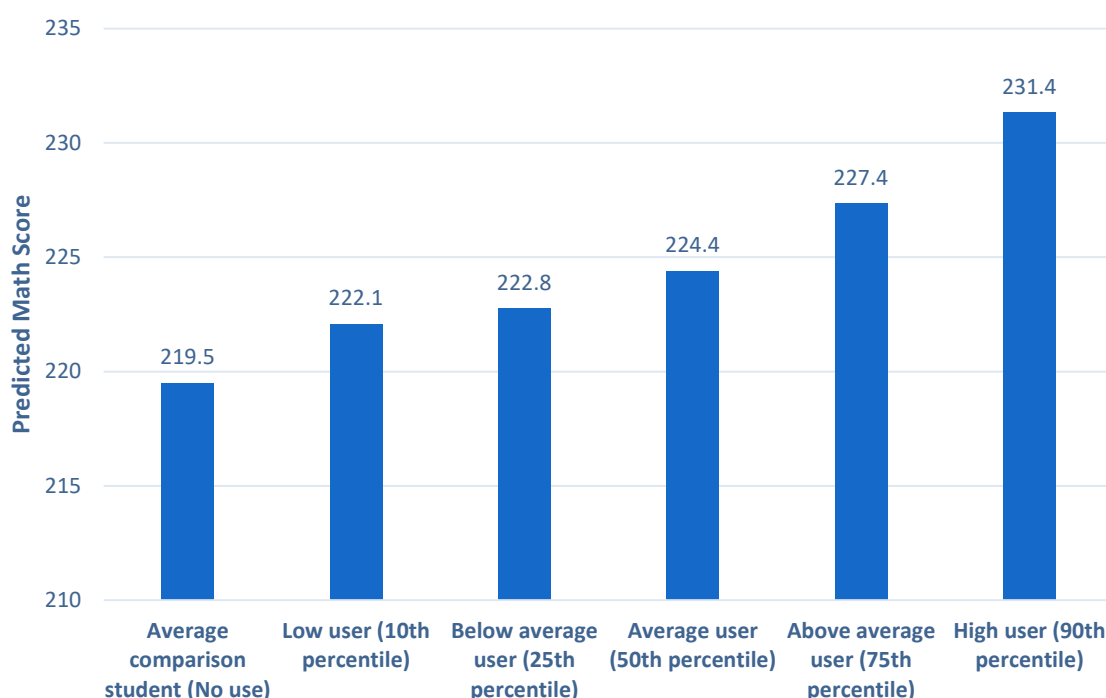
The evaluation team found that Beast Academy had statistically significant, positive effects on math achievement. Specifically, students who had a record of Beast Academy use in both years of the intervention period scored on average 8.78 scale points higher on the spring 2021 MAP assessment in math than did their matched comparison group peers who had no exposure to Beast Academy during that period (treatment mean = 228.28, comparison mean = 219.50, $p = .000$).

The impact of Beast Academy on math achievement varied by treatment group, with the largest effects seen for Rising Scholar students. Students identified in district administrative records as Rising Scholars (mean = 230.67) scored on average 11.17 scale points higher than did the comparison group (mean = 219.50). Students identified as Cluster students (mean = 229.64) scored 10.14 scale points higher than did the comparison group. A third treatment group, students who were not identified as Rising Scholar or Cluster students but who had a record of Beast Academy use in the 2 intervention years (mean = 226.24), scored on average 6.7 scale points higher than did the comparison group.

The impact also varied by the level of engagement on the Beast Academy Online platform, as measured by the number of lessons attempted over the 2-year intervention period. The higher the number of lessons attempted, the larger the gap in math achievement was compared to students who had no exposure to Beast Academy Online. For the average MAPS Beast Academy user—a student who attempted 120 lessons on the platform over 2 years—this effect was equivalent to a 4.9-point increase in math scale scores over the average score for the

comparison group. Specifically, attempting 120 lessons on the platform was associated with a math score of 224.40, whereas the comparison group's average was 219.50. WestEd estimated the predicted score for this and other use profiles based on how MAPS students used the platform during the intervention period. Figure 1 illustrates the upward trend in math scores associated with discrete levels of Beast Academy Online use. (See Tables A1 and A2 in the Appendix for complete tabled results of analyses.)

Figure 1. Higher Levels of Beast Academy Online Use Were Associated With Higher Math Achievement Scores



Note. Results are based on a weighted sample of 1,094 MAPS students enrolled in grades 4, 5, and 6 in SY 2020/21. "Predicted math score" refers to the expected MAP scale score associated with each level of student use on the online platform. Student use profiles are calculated based on the cumulative number of lessons attempted on the platform over SYs 2019/20 and 2020/21.

Student Survey Results

In addition to analyzing student achievement patterns, WestEd surveyed all MAPS students who were enrolled in grades 4 through 6 during SY 2020/21. The survey assessed students' attitudes and perceptions toward math in eight different domains using established measures validated through prior research: motivation to study, effort and perseverance, interest in math, cooperative learning, self-concept in math, growth mindset, self-efficacy in math, and beliefs about math as a learnable subject. Reliability tests indicated that all scales except growth mindset held together as expected, with reliability estimates above the generally accepted threshold (Cronbach's alpha ranged from 0.72 to 0.86; growth mindset = 0.55). Growth mindset outcomes were therefore not included in analyses.

To analyze the differences in math attitudes and perceptions, WestEd researchers regressed an indicator for treatment status on the scale composite scores for each of the seven attitudinal domains (excluding growth mindset). WestEd's estimation method compared Beast Academy students to their non-Beast Academy peers within the same school to control for unobserved differences in school settings that may have influenced their attitudes and perceptions toward math. Models calculated the average scale scores for each group within their schools and then averaged these estimates across all schools in the sample.

The evaluation team found that Beast Academy students had more positive attitudes and perceptions toward math on six of the domains measured. The largest positive association between Beast Academy use and math attitudes was found for the *self-concept in math* measure, followed by *self-efficacy in math*, *interest in math*, *effort and perseverance*, *beliefs about math as a learnable subject*, and *motivation to study*, in order of descending magnitude of differences between the two groups. These results were consistent when disaggregating the analyses by grade level. The findings listed below were statistically significant. WestEd's analysis revealed the following patterns:

- *Self-concept in math.* Beast Academy students had an average score of 3.50, compared to the non-Beast Academy average of 2.79, a difference of .71 on the 4-point scale for this measure, meaning Beast Academy students had higher appraisals of their own abilities in math.
- *Self-efficacy in math.* Beast Academy students had an average score of 2.94, compared to the non-Beast Academy average of 2.47, a difference of .47 points on the 5-point scale for this measure, meaning Beast Academy students expressed higher levels of confidence in their ability to learn and do difficult math work.
- *Interest in math.* Beast Academy students had an average score of 3.06, compared to the non-Beast Academy average of 2.76, a difference of .30 points on the 4-point scale for this measure, meaning Beast Academy students expressed higher levels of agreement with statements related to their interest in math.

- *Effort and perseverance.* Beast Academy students had an average score of 3.57, compared to the non-Beast Academy average of 3.28, a difference of .29 points on the 4-point scale for this measure, meaning Beast Academy students reported engaging and persevering in effortful math work at greater frequencies than non-Beast Academy students.
- *Beliefs about math as a learnable subject.* Beast Academy students had an average score of 5.33, compared to the non-Beast Academy average of 5.07, a difference of .26 points on the 6-point scale for this measure, meaning Beast Academy students were more likely to believe that math could be learned, as opposed to the belief that math is only accessible to high-ability students.
- *Motivation to study.* Beast Academy students had an average score of 3.18, compared to the non-Beast Academy average of 2.94, a difference of .24 points on the 4-point scale for this measure, meaning Beast Academy students expressed higher levels of agreement with statements related to their motivation to engage and persist in difficult math work.
- No significant differences were found between Beast Academy and non-Beast Academy students on the scale scores for *cooperative learning* and *growth mindset*.

Beast Academy students were also more likely than non-Beast Academy students to agree with positive statements related to their interest in and enjoyment of math in the current school year. Beast Academy students, compared to their non-Beast Academy peers who were enrolled in the same schools, had higher average responses to the items “*Math is interesting this year*” (3.97 versus 3.62, respectively) and “*I enjoy math this year*” (4.01 versus 3.45, respectively). Each of these items was measured on a 5-point scale, where 1 indicated *Totally untrue* and 5 indicated *Totally true*.

Additionally, WestEd examined open-ended responses to the questions “*What is your favorite thing about Beast Academy?*” and “*What is your least favorite thing about Beast Academy?*”

The predominant theme from responses was that Beast Academy was more challenging and often more engaging than students’ typical math work. Among the key themes observed by WestEd evaluators were the following:

- Students noted that they enjoyed being challenged by Beast Academy and that these challenges often made math work feel fun, fresh, or new. As one student put it, “I like Beast Academy because it is difficult and it teaches math in different ways.”
- Students liked Beast Academy’s collaborative nature, noting that they enjoyed working with their classmates to solve difficult problems. One student summed up this sentiment by saying, “My favorite thing about Beast Academy is working as a team and figuring out problems together.”
- Students appreciated the opportunity to try problems again if they did not get the answer right the first time; some further noted how the program encouraged

continued effort. As one student put it, “[Beast Academy] makes me want to keep trying and not give up in math even when it’s hard.” Some students wished they could have more opportunities to attempt a problem before seeing the answer or having to start over.

- Most answers to the “least favorite thing about Beast Academy” prompt had to do with how challenging the respondents felt the program was in a negative sense. While many students appreciated that the program pushed them to work through difficult problems, some students expressed frustration at the level of difficulty they encountered and, less frequently, a lack of clarity in the explanations to problems that they got wrong.

Reflection on the Results

Overall, the evaluation results provide evidence to suggest that Beast Academy positively influenced math achievement and math-related attitudes among the cohort of MAPS students who were enrolled in grades 4, 5, and 6 in SY 2020/21. Due to the unprecedented disruptions to learning conditions that occurred during the study period as a result of the COVID-19 pandemic, WestEd urges caution in extrapolating the results beyond this specific population of students. While the evaluation team made concerted efforts to preserve the rigor of the analyses, the study was conducted during an extraordinary time in the educational trajectories of students. Despite the evaluation team’s confidence in the findings, there are many potential unmeasured factors that could have affected students’ performance on and engagement in math-related activities during this time. For example, the way the district used Beast Academy changed over the course of the study period, and some students may not have received a full year of math instruction from the same teacher. Therefore, the results should be viewed as a snapshot of the relationship between Beast Academy use on the one hand and math achievement and attitudes on the other within the unique context of MAPS at this time.

Notwithstanding the need to interpret these results cautiously, the findings reflect a positive association between Beast Academy and student outcomes. The strength of the evidence found by the evaluation team is greatest for the impacts on math achievement. Because the student achievement analyses used a rigorous matched comparison group design, the significant differences in math scores between Beast Academy students and their matched peers can be reasonably attributed to participation in the Beast Academy program while acknowledging the potential confounding factors that may also have been at play. The evaluation team’s use of student-level usage data allowed for a more refined measure of student exposure to the

program—and, consequently, a more nuanced measure of treatment—which likely improved the precision of the estimates when compared to those generated using the district-provided indicators of Beast Academy exposure. On the other hand, since the usage data only captured students’ engagement with the online portion of the Beast Academy program, it may have missed other important ways in which students engaged with the program—for example, through direct teacher instruction and support.

When comparing the results based on the district-provided indicators of Beast Academy exposure with those based on the usage data, similar trends can be seen with respect to Beast Academy students’ higher math performance relative to their peers who did not use Beast Academy. With the district indicator data, the estimated average effect of Beast Academy exposure was 8.78 scale points higher than the average comparison student; with the usage data, the estimated average effect was a more conservative 4.94 scale points. It is beyond the scope of this evaluation study to unpack the differences between the district indicator of Beast Academy exposure and the usage data gathered through the online platform. Nonetheless, both the district indicator data and the usage data reveal a positive association between Beast Academy participation and student performance. Given that the usage data provide a more nuanced glimpse into students’ actual levels of engagement on the online platform, there is an opportunity to further explore these data to understand how Cluster and Rising Scholar students differ from each other in terms of their usage and achievement patterns. From an implementation perspective, more can also be learned about how teachers use Beast Academy to support the unique learning needs of Cluster and Rising Scholar students.

The student survey results can be interpreted as measures of associations between students’ average attitudes and perceptions toward math at the end of SY 2020/21 and their self-reported usage of Beast Academy at any time during SY 2020/21. As with the student achievement analyses, all results were positive and highly statistically significant, meaning it is unlikely that they could have been obtained due to chance. Results from the open-ended survey items largely corroborated the trends from the scale analyses.

WestEd’s survey estimation approach controlled for unobserved school-level characteristics, ensuring that no two students were compared across different school contexts. However, the absence of linked pretest measures of the outcomes meant the evaluation team could not control for preexisting differences in students’ attitudes and perceptions toward math. Consequently, results of the survey analysis cannot be exclusively attributed to participation in Beast Academy. The treatment status indicator for the student survey analyses also lacked the nuance of its usage counterpart in the achievement analyses, as it was a self-reported dichotomous indicator of Beast Academy use. Unlike the usage data, this measure could only indicate whether students had used Beast Academy at any point in the prior school year, even if the extent of that use was limited to a single session or lesson; it did not capture richer patterns related to dosage or length of exposure. Nevertheless, the data show clear patterns of differences between Beast Academy and non-Beast Academy students. Future research could

probe the potential causes for these differences and the mechanisms by which Beast Academy may help promote feelings of self-concept, self-efficacy, interest, effort, beliefs, and motivation in math. An in-depth qualitative examination of students' experiences using the program, for example, might ask students to share, in their own words, what specific features of the Beast Academy program contribute to such feelings. By the same token, a more rigorous quantitative study, such as one that randomly assigned students to different levels of exposure to the program, could help bring to light the extent to which the differences in outcomes observed in WestEd's survey are driven by specific components of the program or by other unmeasured characteristics of the students themselves.

An equally important question, especially in the context of the COVID-19 disruptions to students' learning environments, concerns which learning conditions are most conducive to maximizing the potential benefits of Beast Academy. The results from the student achievement analyses suggest that higher levels of use on the online platform—at least as measured by the number of lessons attempted—predicted higher math scores among the population of students examined for this study. It may be reasonable to ask what a realistic number of lessons attempted could be in a given school year for a given student. To aid in these kinds of discussions, the study identified several student use profiles and detailed the predicted math score associated with discrete levels of engagement on the platform (i.e., the number of lessons attempted). Future research can build upon the results of this study by further exploring the conditions under which Beast Academy can be most effective in meeting the unique learning needs of each student.

Appendix

Table A1. Model Results for the Student Achievement Analyses, by Beast Academy Status and Treatment Group

Treatment Group	Mean Math Score, Beast Academy Students	Mean Math Score, Non-Beast Academy Students	Mean Difference	Standard Error	p-Value	Treatment N
Any Beast Academy designation	228.28	219.50	8.78***	1.052	0.000	547
Cluster students	229.64	219.50	10.14***	1.382	0.000	211
Rising Scholar students	230.67	219.50	11.17***	1.437	0.000	86
Beast Academy user, neither Cluster nor Rising Scholar	226.24	219.50	6.74***	1.057	0.000	250

Note. Weighted N = 1,094. All results were statistically significant at the $p < 0.001$ level (***).

Table A2. Model Results for the Student Achievement Analyses, by Levels of Beast Academy Use

Use Profile	Usage Level (in lessons attempted)	Corresponding Percentile Rank	Marginal Effect	Predicted Math Score	Standard Error	p-Value
Low user	13	10th	2.63***	222.08	0.705	0.000
Below average user	45	25th	3.32***	222.78	0.635	0.000
Average Beast Academy user	120	50th	4.94***	224.40	0.499	0.000
Above average user	257	75th	7.90***	227.36	0.441	0.000
High user	442	90th	11.89***	231.35	0.765	0.000

Note. Weighted N = 1,094. All results were statistically significant at the $p < 0.001$ level (***). The “Average Beast Academy user” is defined as a student with the median number of lessons attempted during the intervention period. Corresponding percentile rank refers to where a given student fell on the distribution of lessons attempted. The marginal effect refers to the predicted change in the coefficient on math performance associated with the corresponding number of lessons attempted, and the predicted math score is the regression-adjusted score associated with that number of lessons attempted. The model includes all matched treatment and comparison students, without censoring the data to remove students with atypical use patterns.

