

Goblins Carolinas Pilot Analysis

Anika Alam, Eesha Iyer, Jing Liu¹

**The Center for Educational Data Science and
Innovation (EDSI)**

University of Maryland, College Park

August 14, 2026

¹ Corresponding author (jliu28@umd.edu).

Executive Summary

This report presents a descriptive analysis of a spring 2026 pilot of Goblins in two North Carolina school districts: **Vance County Schools** and **Cabarrus County Schools**. Due to data limitations and the short period of implementation, the goal of this report is not to provide causal estimates of the impact of Goblins on students; rather, this report complements our evaluation in Deer Valley Public Schools to provide a comprehensive picture of usage and early efficacy of Goblins across a range of district settings. The two districts we study differed in the grade levels served by Goblins, implementation details, and specific tests used to measure math achievement. We thus present our findings separately for each district.

Specifically, the analysis of Vance County Schools examines Goblins usage for grades 6 to 8 math students at three school sites. With access to data beyond the implementation schools, we compare the growth of students' iReady math assessment scores from winter to spring between Goblins users and their peers in the same grade level and district who were not offered Goblins. We find that usage of Goblins is associated with a 0.18 standard deviation increase in iReady math score. Limited qualitative data from teacher interviews and teacher and student surveys suggest favorable perception of Goblins in effectively differentiating instruction and meeting individual student learning needs.

The analysis of Cabarrus County Schools examines only Math 1 students with an activated Goblins license at one school site. As we do not have access to data beyond this one implementation school, we are only able to conduct within-sample analyses for Math 1 students. Higher usage of Goblins (at least 2 problems solved per week) is associated with a 0.15 standard deviation improvement in End-of-Course assessment scores. However, due to the small sample size, this estimate is not statistically significant. Cabarrus did not approve our request to conduct surveys or interviews with teachers or students so we do not have data on teacher or student perceptions.

Across the two districts, despite our limited ability to conduct causal analyses, we find directionally consistent results on the potential positive effects of Goblins on student learning. These findings are in line with our evaluation in Deer Valley Public Schools, with coefficients range between 0.11 and 0.18 standard deviations across the three sites. We also find large variability in adoption and usage across school site and grades. These results help inform and benchmark future studies to rigorously evaluate the impact of Goblins using a rigorously designed randomized controlled trial.

Part 1: Vance County

Goblins Implementation in Vance County

Goblins was offered to grades 6-8 math classrooms across three schools in Vance County: STEM Early High School, AdVance Academy, and Vance Virtual Village Academy.

These schools were selected because each school has math teachers who instructs students across all more than one grade level, making them well suited for implementing a common supplemental instructional resource. Informational interviews with district math coaches shared that the primary goal of incorporating Goblins was to supplement core math instruction by reducing instructional planning demands on teachers while providing differentiated learning opportunities for students requiring additional support.

Although the schools shared this common implementation model, they serve substantially different student populations and operate under distinct educational models.

STEM Early High School is an accelerated STEM program serving students in grades 6 to 8.

AdVance Academy is an alternative school serving students in grades 6 to 12 and provides personalized instructional pathways for learners whose needs are not optimally met in traditional school settings.

Vance Virtual Village Academy is a fully virtual K-12 school that frequently incorporates asynchronous learning opportunities.

These differences in school mission, instructional delivery, and student population provide important context for interpreting implementation and outcomes across sites. A total of **185** students participated in the pilot. The implementation period extended from **March 4 through May 13 2026**, spanning approximately **eight weeks of instruction**.

Data

The data provided by Vance County Schools consist of student-level records for all grades 6 to 8 students, including a comparison sample of grades 6 to 8 students at Vance County Middle School who were not offered Goblins. The records include students' iReady math assessment scores and demographic characteristics, including gender and race/ethnicity. The data also include teacher-level information such as teacher education level and years of experience.

Goblins usage data consist of detailed student-level records that capture engagement with the platform. The data include measures of time spent using Goblins, the number of problems and assignments attempted and completed, and the instructional modules students accessed during the implementation period.

The only mutual information in both sources that enables linking is student first and last names. Because students self-entered their names when creating their Goblins account, Goblins' platform data contained inconsistencies in name spelling, abbreviations, nicknames, and typographical errors relative to the district administrative records. We used a fuzzy matching procedure that successfully matched **87% of Goblins users ($N = 141$) identified by the platform as belonging to Vance County Schools.**

Outcome Measures

We measure math achievement using iReady scores. iReady assessments are administered three times a year and can place students at an overall grade level that can be above or below their current grade level and provides performance results within each of the four math domains: *number and operations*, *algebra and algebraic thinking*, *data and measurement*, and *geometry*.

We focus on the spring 2026 iReady assessment that was taken on April 22. We standardize the overall scale score by grade.

Goblins Engagement Metrics

Following our Deer Valley evaluation (Alam, Dominguez-Mel, & Liu, 2026), we characterize student engagement on Goblins using three complementary measures:

Active minutes captures the total time students spend on the platform, including interacting with the AI tutor, working on assignments or independent practice problems, and navigating other features within Goblins.

Problems attempted refers to the number of problems that students open, regardless of whether they ultimately submit a response.

Problems solved represents the number of problems students answer correctly. Because the AI tutor requires students to solve the current problem before progressing to the next one, this measure is equivalent to the total number of correctly completed problems. For all engagement measures, we aggregate activity across both teacher-assigned and independent practice modes and do not distinguish between these two channels of platform use.

Again, consistent with our Deer Valley evaluation, we use problems solved as our primary measure of platform engagement. While active minutes and problems attempted are highly correlated with problems solved, each has important limitations. Active minutes include idle time spent on Goblins' platform, while problems attempted captures only problems opened by the student and does not indicate whether the student actually provided an answer.

Goblins Usage Statistics

Figure 1 provides descriptive statistics on Goblins usage using the 141 students we successfully identified. Across the eight-week pilot, students spent an average of 36 minutes on the platform, attempted 9 problems, and completed 7 problems.

Grade 6 students had the highest usage and spent an average of 52 minutes on Goblins, nearly 44% higher than the sample average. Grade 6 students attempted and completed 13 and 11 problems, respectively, which translates to 44% to 36% higher usage than the sample average. In contrast, grades 7 and 8 students have a usage level at roughly half to 2/3 of what their grade 6 peers have. While the specific mechanisms are not clear, we observe a similarly higher engagement level for grade 6 students than other middle school grades in Deer Valley.

It is also important to note the substantial variation in Goblins usage across the three school sites. Student engagement was particularly low at Vance Virtual Village Academy, where nearly 74% of Goblins users recorded no activity during the pilot and students solved, on average, only one problem over the eight-week implementation period. In comparison, students at AdVance Academy solved an average of three problems, while students at STEM Early High solved an average of nine problems.

Sample Characteristics

The analytic sample consists of 680 students in grades 6 to 8 from Vance County Schools. Of these, 141 students (21%) had a linked Goblins account and comprise the treatment group. The remaining 539 students (79%) attended Vance County Middle School, where Goblins was not offered, and serve as the comparison group.

Table 1 reports the sample characteristics and disaggregates characteristics of students with a Goblins account to those in the comparison school. In the analytic sample, 67% of students are Black, 20% are Hispanic, and 10% are White. We do not have information about disability status or English Language Learner (ELL) designation. Nearly half of the teachers are novice and have fewer than 5 years of experience, and over 30% of the teachers have veteran status with over 15 years of experience.

Importantly, compared to their peers, students with a Goblins account had much higher winter iReady achievement scores (0.68 SD), 22 percentage points more likely to be above grade level, and 12 percentage point less likely to be Black students. Thus, Goblins users in Vance appear to be much higher achieving and more advantaged than non-Goblins users. We control for all these baseline and demographic differences in our regression analysis.

Analytic Approach

The quantitative analysis compares students with a Goblins account to students who were not offered Goblins. We utilize ordinary least squares regressions where Model 1 does not include any controls and Model 2 controls for gender, race/ethnicity, and winter iReady math performance.

The winter iReady assessment was administered at the end of January 2026 and serves as the most complete approximation for math ability at the beginning of the Goblins pilot launch. We do not explore variation in Goblins usage because there was limited variation in student-level engagement. Usage was particularly low at Vance County, where the average Goblins user solved just 7 problems over the eight-week pilot. As a result, it was not possible to meaningfully categorize levels of engagement. More than half of Goblins users solved fewer than 3 problems in total, while only 10% solved more than 16 problems over the pilot—an average of 2 problems per week. For these reasons, we do not categorize levels of engagement that was used in the analysis for Cabarrus County.

We complement the quantitative analyses with surveys and interviews of participating teachers to examine implementation and identify challenges associated with the pilot launch. All six math teachers at treated school sites were offered a \$50 gift card for completing a teacher survey and interview and administering student surveys in their classes. In accordance with district policy, students' families were required to provide consent for students to participate in the survey. Overall participation was limited: three teachers and ten students completed the surveys, and two teachers participated in interviews.

Quantitative Findings

Table 2 presents the regression results. Model 1 does not include student characteristics and compares average spring iReady scores across the two groups. On average, Goblins users had spring iReady scores that were 0.53 standard deviations higher than those of non-Goblins users ($p < 0.001$). After adjusting for student gender, race/ethnicity, and pre-implementation math achievement, the estimated difference is reduced to 0.18 standard

deviations ($p < 0.05$). Notably, this estimate is larger than the effect observed in Deer Valley (0.11 SD increase in state standardized test scores), which may reflect differences in implementation, the assessment used to measure student achievement, or empirical methods used. In particular, given that Goblins users are higher achieving and overall more advantaged than non-Goblin users (Table 1), there might be important omitted variables (e.g., poverty, motivation) in our model that bias the estimate upwardly.

Qualitative Findings

The **Appendix Tables and Figures** present responses from **the student and teacher surveys**, as well as selected teacher testimonials describing their experiences with Goblins during the pilot.

All teacher respondents reported using the pre-made assignments available on the Goblins platform, the teacher avatar, and complementary instructional tools such as IXL and Desmos to supplement their instruction.

To better understand perceptions of Goblins, teachers and students were asked to rate a series of positive statements about the platform's usefulness, the quality of its feedback, and its ability to support differentiated instruction. Responses were measured using a **five-point Likert scale** ranging from 1 (strongly disagree) to 5 (strongly agree).

Teacher responses were uniformly **positive**. All teachers either slightly agreed or strongly agreed with every statement. All teachers strongly agreed that Goblins supports differentiated instruction by providing multiple ways for students to engage with content (e.g., visual, audio, and interactive resources, as well as multiple language options), offers useful dashboard data to monitor student progress, and supports independent, self-paced learning.

Student responses were more varied but remained **generally favorable**. Across survey items, between 40% and 100% of respondents agreed or strongly agreed with the positive statements about Goblins. Students expressed the strongest agreement that Goblins allowed them to work at their own pace and that they enjoyed using the platform to learn math.

Agreement was comparatively lower on statements that Goblins helped them understand math better, provided clear and easy-to-understand instructions and explanations, and explained topics in multiple ways.

Teacher interviews revealed optimism about Goblins' potential to provide personalized feedback tailored to individual student needs. At the same time, teachers identified several implementation challenges that may have limited students' engagement with the tool. In particular, students experienced difficulties accessing Goblins at home because of limited internet bandwidth. Teachers also noted that Goblins was introduced relatively late in the spring semester, which limited the amount of time available for students and teachers to become familiar with the tool and integrate it into instruction.

Discussion

This analysis may underestimate the association between Goblins usage and math learning because the outcome measure captured student performance only five to six weeks after the pilot began, rather than at the conclusion of the full implementation period. As a result, the assessment may not fully reflect the cumulative effects of Goblins usage on student learning. This limitation underscores the importance of aligning future implementation timelines with benchmark assessment schedules so that outcome measures capture the complete period of exposure to the intervention.

Part 2: Cabarrus County

Goblins Implementation in Cabarrus County

In Cabarrus County Schools, Goblins was offered at a single site, **Mount Pleasant High School**, across Math 1, Math 2, and Math 3, and Math 4 courses. These courses are part of Cabarrus County's required high school math sequence for graduation.

Math 1 generally corresponds to a traditional Algebra I course and serves as the entry point into the high school sequence.

Math 2 emphasizes geometric concepts while continuing to build algebraic reasoning.

Math 3 is broadly equivalent to a traditional Algebra 2 course.

Math 4 is equivalent to pre-calculus. Together, these courses provide the foundational mathematics instruction required for all students pursuing a North Carolina high school diploma.

A total of **382** students in these math courses activated a Goblins license. Similar to Vance County, the implementation period extended from **March 4 through mid-May 2026**, spanning approximately **eight weeks of instruction**. This timeframe provided students with sustained access to Goblins during regular math instruction, allowing for an examination of implementation patterns and student engagement over the course of the pilot.

Data

The data provided by Cabarrus County Schools consist of student-level records for students enrolled at Mount Pleasant High School. These records include students' Spring 2026 North Carolina End-of-Course (EOC) math assessment results, Mastery View Predictive Assessment (MVPA) assessment results, and demographic characteristics, including gender, race/ethnicity, disability status, and English learner (EL) designation. MVPA assessments are interim assessments to assess a subset of course standards that have been taught. Cabarrus uses MVPA to assess content mastery for Math 1 and Math 3 and was administered in October and March of the 2025-26 school year. EOC math assessment scores are described in detail in the following subsection.

Goblins usage data consist of detailed student-level records that capture engagement with the platform. The data include measures of time spent using Goblins, the number of

problems and assignments attempted and completed, and the instructional modules students accessed during the implementation period.

Cabarrus County Schools worked with Goblins to link district administrative records with Goblins usage data using individual student names available in both data sources. This matching process successfully linked approximately **100% of students identified as Goblins users within Cabarrus County Schools**, resulting in a nearly complete integrated dataset for analysis.

Outcome measures

Student math achievement was measured using North Carolina EOC assessments, which are administered to all students enrolled in Math 1 and Math 3 at the end of the academic year (typically in May). The state requires that all students enrolled in these courses complete the corresponding EOC assessment and that the assessment accounts for at least 20% of the student's final course grade. Because North Carolina does not administer a state-mandated EOC assessment for Math 2 and Math 4, comparable achievement data are not available for students enrolled in that course. We standardize the EOC score by grade level and create a binary indicator if a student passed their EOC.

Goblins usage data indicate that there were only nine Math 3 students, all of whom had activated a Goblins license but had no activity on the platform. Because there was no observed Goblins engagement among Math 3 students, we cannot meaningfully examine patterns of Goblins usage or assess the relationship between Goblins implementation and student math achievement for this group. As a result, analyses examining the relationship between Goblins implementation and student math achievement are limited to students enrolled in Math 1.

Goblins Usage Statistics

The usage period was from March 4 to May 13, approximately spanning 8 instructional weeks. **Figure 2** provides descriptive statistics on how Goblins was used in the overall district sample and by math course (i.e., Math 1, 2, and 4). Across the eight-week pilot at Mount Pleasant High, students spent an average of 110 minutes on the platform, attempted 31 problems, and completed 28 problems. Math 2 students had the highest usage and spent an average of 158 minutes on Goblins, nearly 43% higher than the sample average; Math 2 students attempted and completed 51 and 48 problems, respectively, which translates to about 66% higher usage than the sample average. Math 4 students spent about the same number of minutes on Goblins and solved as many problems as Math 1 students.

To better distinguish levels of Goblins usage, we categorize students into two levels of platform engagement:

- Students who averaged *fewer than two problems solved per week* during the pilot were classified as **low engagement users** and served as a pseudo-comparison group.
- Students who averaged *two or more problems solved per week* were classified as **high engagement users**.

Our within-sample comparison approach examines whether greater engagement with Goblins between these two groups of students was associated with differences in math achievement.

Sample Characteristics

The analysis focuses on the 167 students enrolled in Math 1 at Mount Pleasant High School, representing approximately 43% of all students who used Goblins during the pilot. This analytic sample was selected because students enrolled in Math 2 and Math 4 do not participate in North Carolina's state-mandated math assessments, which served as the measure of student achievement in this study. Given the small size of this analytic sample, we are limited in statistical power to detect small coefficient size.

Table 3 provides detailed usage statistics on Goblins' engagement for the analytic sample. 63% of student users solved fewer than 2 problems per week, on average. The second largest group of users solved between 5 to 9 problems per week on average and spent 10 to 14 minutes weekly on the platform.

Table 4 reports the sample characteristics and disaggregates characteristics of students above and below the two problems solved per week threshold. In the analytic sample, 63% of students are White, 22% are Hispanic, 11% have an IEP, and 11% are designed as English Language Learners. There is no statistical difference in the composition of students above and below the usage threshold ($p > 0.05$)

Analytic Approach

As discussed above, because student-level records were not available for schools besides Mount Pleasant High School, we are not able to construct a comparison group for examining end-of-year math achievement. Consequently, the analysis relies on variation in Goblins engagement among Goblins users.

The analysis uses a single definition of Goblins usage as the independent variable that dichotomizes Goblins users at Mount Pleasant High School based on an average

threshold of two or more problems solved per week.

The analysis runs two ordinary least squares regressions. The first regression does not include any controls and the second regression includes controls for gender, race/ethnicity, disability status, English Language Learner (ELL) status, and whether the student passed the spring MVPA assessment. The spring MVPA assessment, administered in March 2026, serves as the closest available proxy for students' math ability prior to the Goblins pilot launch. Despite this rich set of control variables, the interpretation of the results should remain correlational rather than causal given threats of potential omitted variable bias, such as student motivation, teacher effects, and other unobserved factors.

Findings

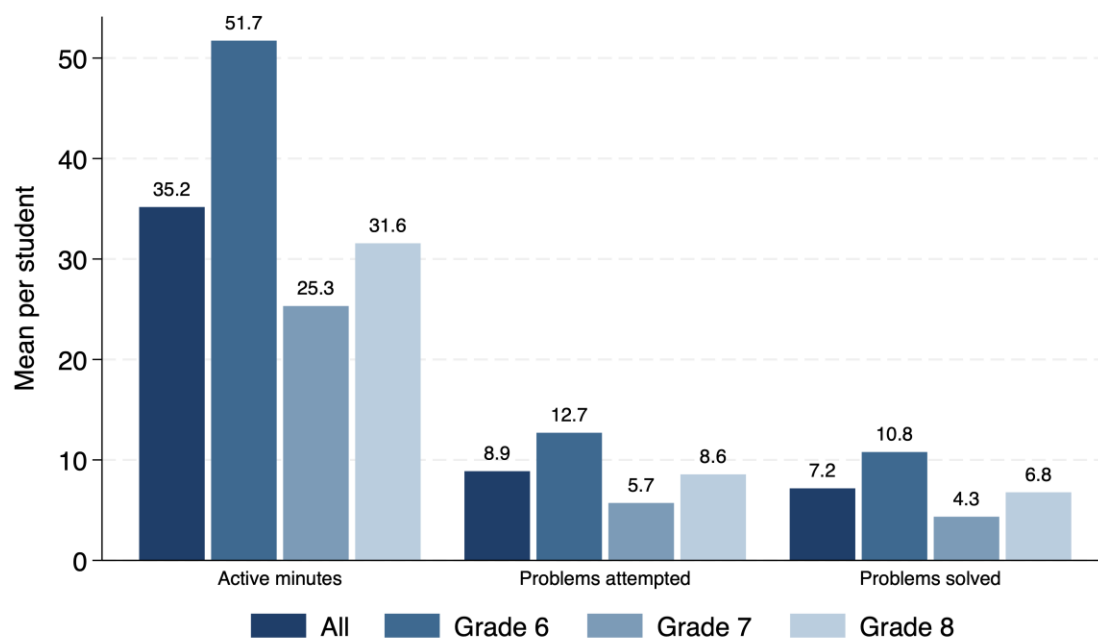
Table 5 presents the regression results using the dichotomous measure of Goblins usage. Model 1 is a parsimonious specification that does not include student characteristics and compares average EOC scores across the two usage groups. Although the difference in standardized Math 1 EOC scores between the groups is sizable (0.27 SD), it is not statistically significant ($p > 0.1$), potentially reflecting our small sample size. Because this is an unadjusted comparison, the observed difference could also reflect pre-existing differences in achievement, with higher achieving students potentially engaging more with Goblins. Model 2 adjusts for student characteristics. The results remain consistent, with no statistically significant difference in standardized Math 1 EOC scores between the higher- and lower-usage groups ($p > 0.05$).

Discussion

Although Goblins was utilized by 382 students at Mount Pleasant High, we could not include more than half of the students in the analysis because outcome data were unavailable for the remaining students. This substantial amount of missing outcome data reduced the analytic sample, limiting our ability to detect relationships while also affecting the generalizability of the findings. Without a comparison group, our analyses compare Goblins users based on their usage level. Taking as a whole, our findings suggest that a higher usage might associate with a higher likelihood of passing the EOC assessment.

Tables and Figures

Figure 1: Vance Usage Statistics



Notes: N = 141 and captures 37 students in grade 6, 38 students in grade 7, and 66 students in grade 8.

Table 1: Vance Sample Characteristics

	All Grades 6-8 Students	Goblins Users	VCMS Students (Without Goblins)	Difference
Grade 6	0.30	0.26	0.31	-0.05
Grade 7	0.32	0.27	0.34	-0.07
Grade 8	0.37	0.47	0.35	0.12*
Panel A: Student Characteristics				
Female	0.43	0.43	0.43	0.00
Asian	0.00	0.01	0.00	0.01
Black	0.67	0.57	0.69	-0.12**
Hispanic	0.20	0.16	0.21	-0.05
White	0.10	0.23	0.06	0.17***
Other race/ ethnicity	0.03	0.03	0.04	-0.01
Mid-year iReady score (SD)	-0.00	0.54	-0.15	0.68***
At grade level or higher mid-year	0.14	0.31	0.09	0.22***
Student <i>N</i>	660	141	519	660
Panel B: Teacher Characteristics				
Has masters	0.00	0.00	0.00	0.00
Experience < 5	0.48	0.44	0.49	-0.05
Experience 5-15	0.2	0.11	0.21	-0.10
Experience > 15	0.31	0.44	0.30	0.14
Missing license data	0.15	0.25	0.17	-0.08**
Teacher <i>N</i>	48	6	42	48

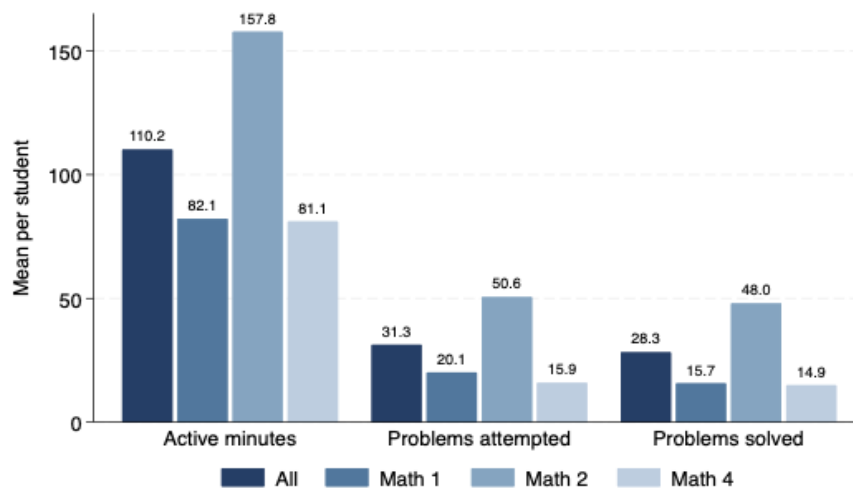
Notes: T-test statistical significance levels are reported in the final column: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2: Vance Regression Results

Outcome	Model 1 (no controls)	Model 2 (with controls)
Spring iReady assessment score (Z)	0.523*** (0.093)	0.181* (0.083)
Lagged test scores and demographics		X

Notes: Standard errors are reported in parentheses. Statistical significance levels: $+p < 0.10$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Figure 2: Cabarrus Usage Statistics



Notes: $N = 382$.

Table 3: Cabarrus Detailed Goblins Statistics

<i>N</i> and % of Sample	
Total Problems Attempted	
Less than 2	80 (47.9%)
2 to less than 3	22 (13.2%)
3 to less than 4	21 (12.6%)
4 to less than 5	17 (10.2%)
5 to less than 10	22 (13.2%)
10 to 17	5 (3.0%)
Total Problems Solved	
Less than 2	105 (62.9%)
2 to less than 3	14 (8.4%)
3 to less than 4	15 (9.0%)
4 to less than 5	11 (6.6%)
5 to less than 10	18 (10.8%)
10 to 17	4 (2.4%)
Total Active Minutes	
Less than 5	67 (40.1%)
5 to less than 10	19 (11.4%)
10 to less than 15	30 (18.0%)
15 to less than 20	20 (12.0%)
20 to less than 30	19 (11.4%)
30 or more	12 (7.2%)
<i>N</i>	167

Table 4: Cabarrus Analytic Sample Characteristics

	Overall Sample	High Users	Low Users	Difference
Passed Spring MVPA	0.410	0.468	0.431	0.037
Female	0.467	0.500	0.479	0.021
Has an IEP	0.105	0.032	0.078	-0.046
English Language Learner	0.105	0.081	0.096	-0.015
White	0.629	0.661	0.641	0.020
Hispanic	0.219	0.177	0.204	-0.027
Multiple Races	0.038	0.016	0.030	-0.014
Black	0.048	0.065	0.054	0.011
Asian	0.029	0.065	0.042	0.023
<i>N</i>	167	105	62	167

Notes: T-test statistical significance levels are reported in the final column: $+p < 0.10$, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

Table 5: Cabarrus Regression results

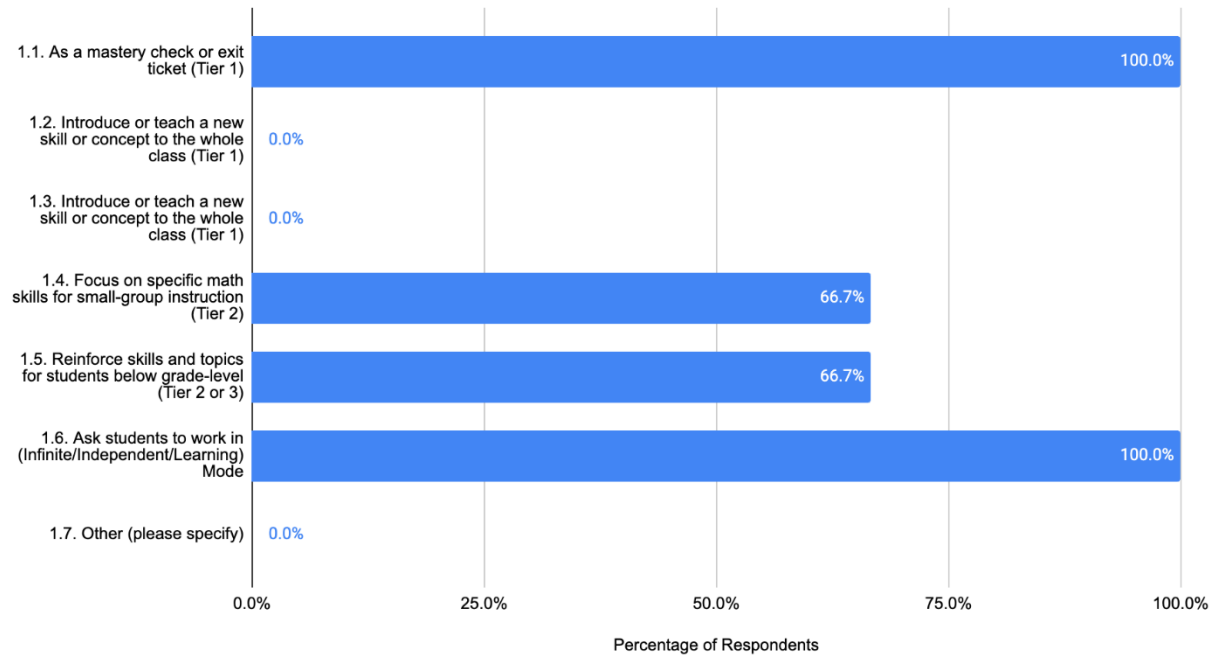
Outcome	Model 1: Comparing high Goblins users to low Goblins users (no controls)	Model 2: Comparing high Goblins users to low Goblins users (with controls)
Math 1 EOC assessment (Z)	0.262 (0.159)	0.154 (0.138)

Notes: Standard errors are reported in parentheses. Statistical significance levels: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

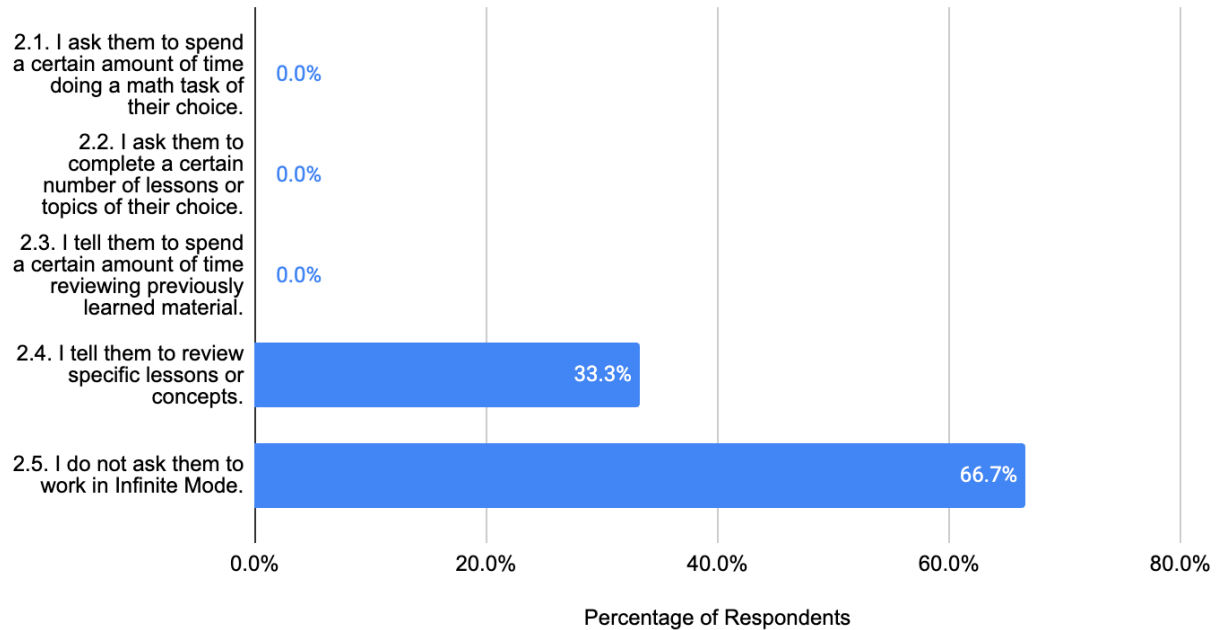
Appendix Tables and Figures

Figure A1: Vance Teacher Survey Responses

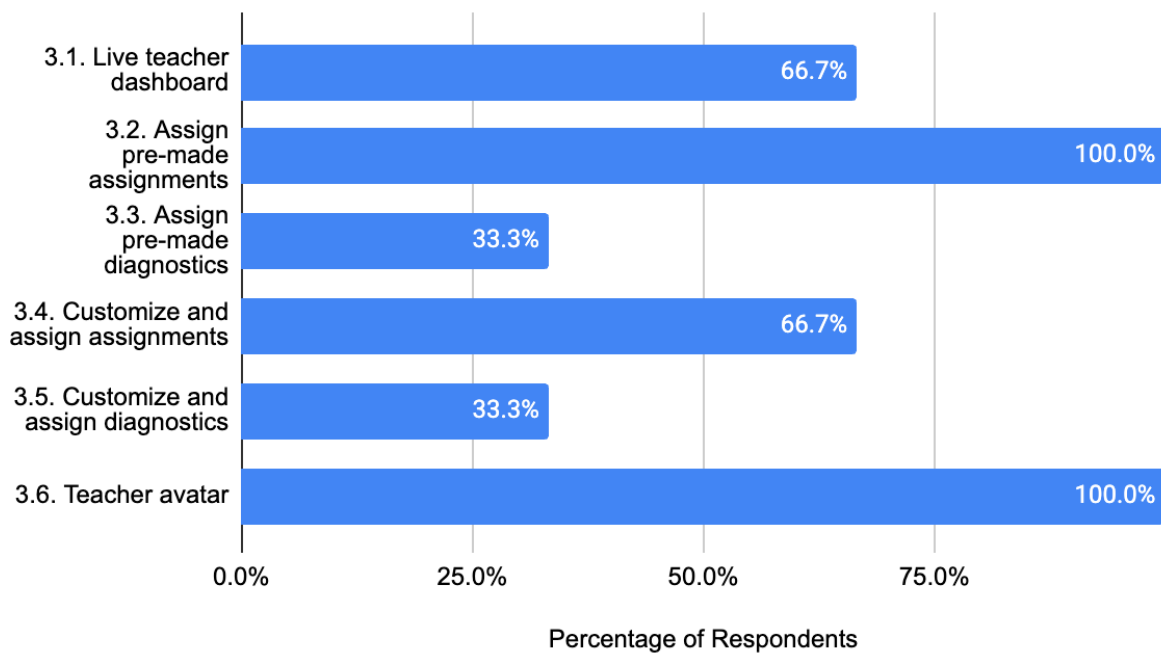
Question 1. How do you incorporate Goblins in your math instruction. Select all that apply.



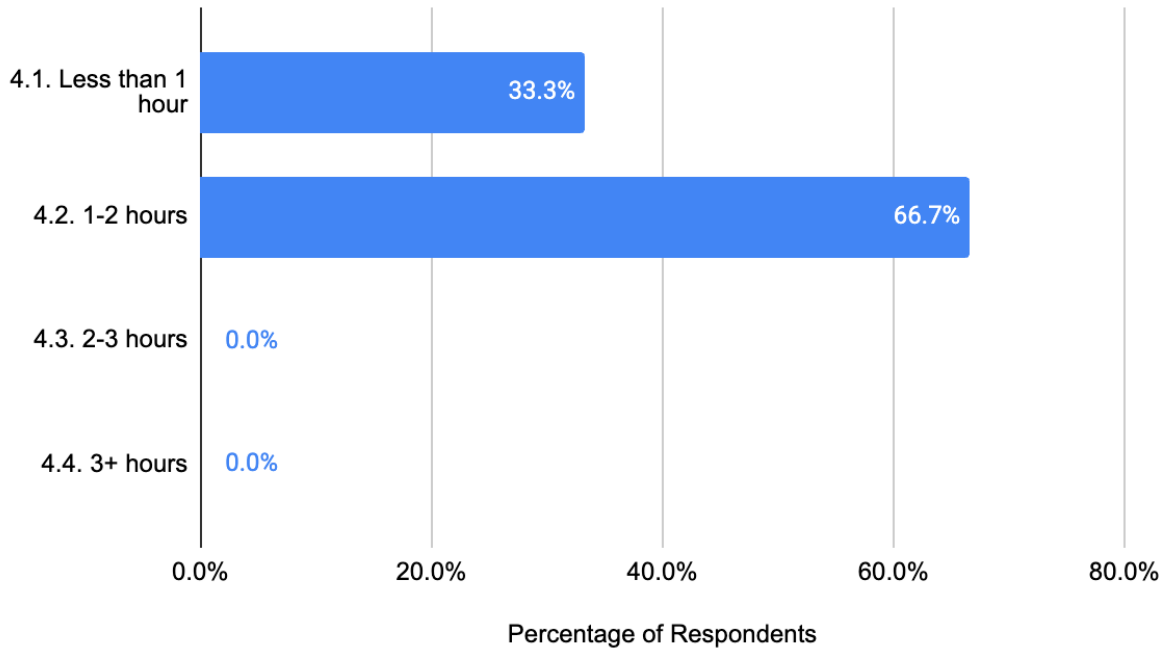
Question 2. If you ask students to work in Goblins' Infinite Mode where students can work independently, how do you incorporate and monitor it? Select all that apply.



Question 3. What Goblins features do you use? Select all that apply.



Question 4. How much time did you spend learning to use Goblins?



Question 5. What other digital learning tools do you use in your math instruction? Select all that apply.

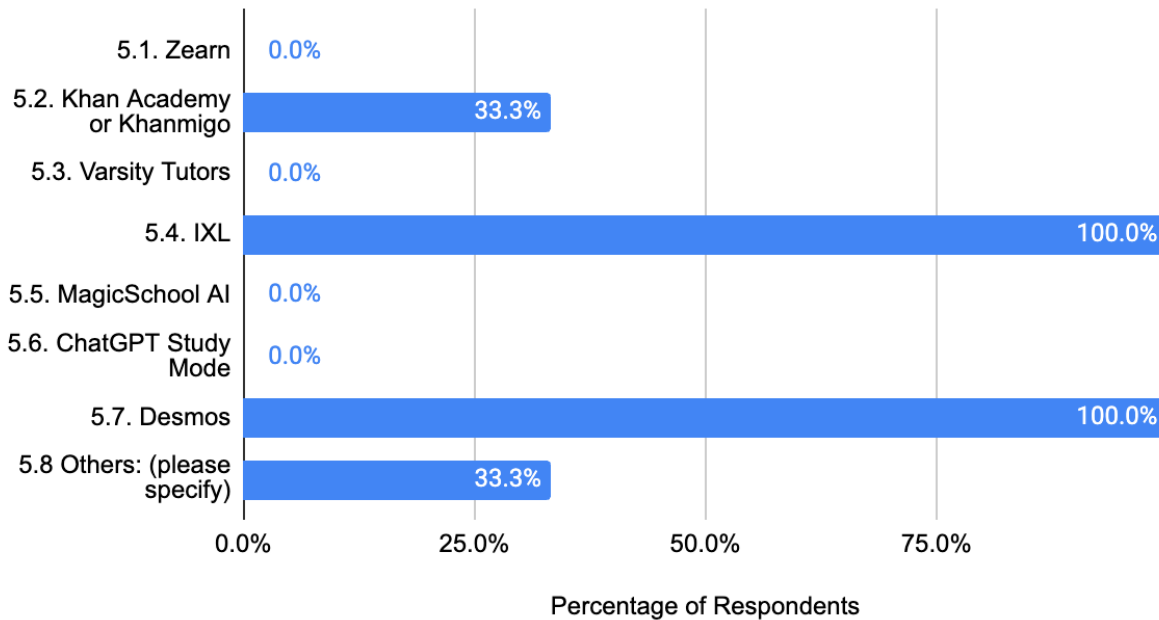


Table A1: Teacher Survey Responses

On a scale of 1 to 5, please rate the following statements based on how you feel about the effectiveness of Goblins in your classroom: (1 = strongly disagree; 5 = strongly agree)

	Slightly Agree	Strongly Agree
This tool helps my students learn mathematics effectively.	2	1
The feedback Goblins provides helps my students understand their mistakes.	1	2
The support the tool provides promotes long-term learning and retention of math concepts.	2	1
Goblins supports differentiated instruction by offering multiple ways for students to engage with content (e.g., multimodal resources such as visuals, audio, and interactive tasks, and availability of multiple language options).	0	3
Goblins supports differentiated instruction by adapting difficulty levels to push the student forward.	1	2
Using this tool saves me time in instructional planning and preparation (e.g., creating assignments, grading, analyzing student performance, and adjusting instruction based on data).	1	2
The tool provides useful data and insights that help me monitor student progress.	0	3
The tool allows me to easily identify and address individual student learning gaps.	1	2
My students are engaged and motivated when using this tool.	1	2
The tool is easy for both me and my students to use (user-friendly interface).	1	2
The tool aligns well with my curriculum and learning standards.	1	2
The tool supports independent learning and student self-paced progress.	0	3
I would recommend this tool to other educators.	1	2
Overall, this is an effective tool for mathematics learning in my classroom.	1	2

Table A2: Vance Teacher Testimonials and Feedback

*"With Goblins, you immediately get live feedback.
That's the plus it has over any other platform I've used."*

*"If you just pop an answer in, it's going to ask you to give it step-by-step, which I love.
I adore that ability."*

*"We've dealt with it enough to know that it can be a promising tool, but to really expedite its features
and make it be as useful as it should, we needed more time."*

"I love that kids are writing, not typing, you know what I mean? It gets them to interact."

*"For it to flow properly in the classroom and be as effective as it should be, or it's intended to be...
it should be introduced in the beginning (of the school year)"*

— Compacted Math Teacher, VCS

*"In our alternative school, where the students' gaps in their learning are too far - I think Goblins helps
fill in the gaps."*

*"I can look back at student conversations when the AI tutor is questioning them. It's great to see how
they respond to the questions and the immediate feedback that they get."*

*"One of the disadvantages is that it is not synced to CLEVER IDs so, they can't just sign in on one
platform."*

— Alternative Middle School Math Teacher, VCS

Table A3: Vance Student Survey Responses

On a scale of 1 to 5, please rate the following statements based on how you feel about the effectiveness of Goblins in your classroom: (1 = strongly disagree; 5 = strongly agree)

	Strongly Disagree	Slightly Disagree	Neither Agree nor Disagree	Slightly Agree	Strongly Agree
Question 1. This tool helps me understand math better.	0	2	2	3	3
Question 2. The feedback I get from this tool helps me learn from my mistakes.	0	2	0	6	2
Question 3. The tool helps me figure out next steps when I am stuck on a math problem.	0	0	2	4	4
Question 4. The instructions and explanations in this tool are clear and easy to understand.	1	3	0	3	3

Question 5. This tool complements the learning I get from my math teacher.	1	0	1	4	4
Question 6. This tool gives me different ways to learn (such as videos, visuals, interactive problems, or step-by-step explanations) that help me understand math.	2	0	2	3	3
Question 7. I can work at my own pace when using this tool.	0	0	0	3	7
Question 8. This tool helps me feel more confident in math.	0	1	5	3	1
Question 9. This tool helps me try more challenging math problems.	1	0	4	3	0
Question 10. I stay focused and engaged when using this tool.	1	0	1	2	4
Question 11. I like using this tool to learn math.	1	0	0	1	8
Question 12. I would like to keep using this tool in the future.	1	1	2	3	3