Speak Agent

Effect of Speak Agent on Math and English Language Proficiency Scores at Middle Grades in Prince George's County Public Schools

ABOUT LEANLAB EDUCATION

Leanlab Education is a nonprofit research organization that helps schools understand and solve their biggest problems. We listen to parents, educators, and students to help them pinpoint specific problems that hinder student achievement. Then, we match those schools with education entrepreneurs. Together, they develop the next generation classroom tools through research and direct feedback.

AUTHOR

Erin Huebert, Ph.D
Leanlab Education

OVERVIEW

The goal of this research report is to evaluate the effect of Speak Agent Math+LanguageSM student usage on student outcomes in Math (measured with the Math Benchmark assessment) and English Language Proficiency (measured with the WIDA ACCESS for ELLs assessment) across grades 6 to 8 within Prince George's County Public Schools (PGCPS).

Speak Agent Math+Language is an academic language learning technology program that strives to improve math communication, vocabulary, content knowledge, modeling, and reasoning skills.

Through a quasi-experimental data analysis, this study finds that there is a positive, independent effect of Speak Agent Math+Language activity completions on Math Benchmark and WIDA ACCESS scores.

DATA & METHODS

To estimate the effect of Speak Agent Math+Language usage on students' scores in Math and English Language Proficiency, multivariable regression analysis was performed. This section will describe the data used in the analysis as well as provide a brief explanation of how to interpret regression results.

A. DATA DESCRIPTION

The key outcome variables of interest in this study are Math and English Language Proficiency. The key explanatory variable is Speak Agent student use.

MATH SCORES

To measure math scores, the Math Benchmark quarterly assessments are used. The assessment consists of a mix of selected response and machine scored text entry response items. The Benchmark tests occurred directly after the Q1, Q2, and Q3 marking periods and should, therefore, correspond to the content covered in the classroom curriculum (and in Speak Agent content) during said periods. The third quarter assessment was optional so the sample size for the third quarter is significantly smaller. The measure used in the analysis is the Percent Correct a student received on the assessment, so the valid range is 0-100.

ENGLISH LANGUAGE PROFICIENCY

To measure English Language Proficiency, the Scale Scores from the WIDA's ACCESS for ELLs assessment are used - a summative English language proficiency assessment for multilingual learners. ACCESS for ELLs is a standards-referenced test, which means that student performance is compared to English language development standards WIDA has defined. Any student can achieve any score, and students are not ranked against each other or against the expected performance of monolingual English speakers. Thus, this study only evaluates the impact of Speak Agent on English language proficiency for multilingual learners in the middle grades and not all learners.

The Scale Scores from the ACCESS assessment are used. Scale scores allow us to compare student performance across grades and within each domain. Scores across domains cannot be compared since different methods are used to score the different domain tests. All scores range from 100-600. See this link for more information on understanding scores. The assessment was taken primarily in quarter 3.

SPEAK AGENT USAGE

There are a number of ways that Speak Agent use can be measured, including Time Spent on Speak Agent (in seconds/hours), Number of Activities Completed, Number of Correct Answers Achieved, among others. The most consistent and valid measure is Number of Activities completed since it measures simply whether a student completed an assigned lesson or not. Time Spent on Speak Agent is not as accurate as it cannot account for possible idle time and Correct Answers is not as valid as it measures a student's performance rather than use. Nonetheless, both measures are included in the regression models as controls.

SUMMARY STATISTICS OF KEY VARIABLES

Table 1 shows the summary statistics for each key variable described above. It is noteworthy to point out that a large proportion of students have no use on Speak Agent, thus the variable is significantly skewed towards zero[1].

1. In the regression analysis, a logged transformation of the Speak Agent usage variable is not used despite the non-normal distribution.

Table 1. Summary Statistics of Key Variables

Quarter	Total Observations	Mean	Standard Deviation	Min	Max
BENCHMARK: PERCENT CORRECT					
Q1	20,166	17.45%	0.16	0	100%
Q2	21,604	19.02%	0.15	0	96%
Q3	5,159	14.57%	0.13	0	83%
ACCESS: SCALE SCORE OVERALL (POSSIBLE RANGE 100-600)					
Q3	5,361	329.91	35.54	220	432
SPEAK AGENT: NUMBER OF ACTIVITIES COMPLETED (POSSIBLE RANGE 0-13)					
Q1	20,166	0.75	1.48	0 (14,762 w/0)	13
Q2	21,604	0.88	2.04	0 (16,972 w/0)	13
Q3	9,481	0.72	1.76	0 (7,664 w/0)	13

Quarter	Total Observations	Mean	Standard Deviation	Min	Max
SPEAK AGENT: TIME ON TASK (IN HOURS)					
Q1	20,166	0.764	2.32	0 (10,658 w/ 0)	84.4
Q2	21,604	0.594	2.62	0 (14,813 w/ 0)	172.5
Q3	9,481	0.54	1.62	0 (6,132 w/ 0)	43.1

B. SAMPLE

The sample used in this analysis comes from 86 middle schools, K-8 academies, and 6th-grade elementary schools within PGCPS. There are 1,597 classrooms represented in the sample accounting for 24,180 unique students. Table 2 breaks down the student sample by type of assessment and by quarter.

Table 2: Student Sample Size

	Total Observations	Q1	Q2	Q3	Q4
Benchmark	46,929	20,166	21,604	5,159	0
ACCESS	5,361	0	0	5,361	0

C. HOW TO INTERPRET REGRESSION RESULTS

Regression allows us to predict a value for y (outcome variable) from x (explanatory variable) that tells us the relationship (positive or negative) and the effect size between the two variables. The computed value - the regression coefficient - tells us how much the outcome variable changes when the explanatory variable changes by one unit. If the regression coefficient is statistically significant, we can be confident that our prediction is accurate in the population at large and that our results did not occur by chance.

For example, in our analysis, Math Scores (y) are being predicted by Speak Agent Activities Completed (x). If the coefficient of (x) is 1.3, it means that for every one additional activity completed, math scores are expected to be 1.3 points higher. And if the coefficient is statistically significant ($p<0.05$), it means that there is less than a 5% chance that our result is a result of random chance and, therefore, there is strong evidence that the effect will be observed at the population level.

We use two types of regression in this analysis: cross-sectional and longitudinal.

In cross-sectional analysis, we predict what the score will be of a student randomly selected from the population at one point in time. Time is held constant and samples may change from one period of time to the next (i.e. students in Q1 sample may be different from students in Q2 sample).

Both samples are pooled in the analysis. We do not know or account for where students started (their baseline scores) and only know and account for where they ended at each quarter. We are assessing the difference in score at a "societal" level, not of any one particular student. We are examining how characteristics of groups/individuals (i.e. gender, time spent on Speak Agent) impact levels of scores at any time. For example, girls get higher scores than boys, students who use Speak Agent more get higher scores than students who use Speak Agent less. Concretely, we could say if we randomly select student A who has completed 1 activity, we can predict that her/his score will be 330 and if we randomly select student B who has completed 2 activities, we can predict that she/he will have a score of 331.73.

In longitudinal analysis, we predict how much a score will change from one quarter to the next for a student randomly selected from the population. The sample is the exact same set of individuals from one period of time to the next (i.e. Q1 to Q2, etc.). We know and account for where they started (their baseline scores) and know and account for their increases/decreases from one quarter to the next. We are evaluating the change

in score at the individual level rather than the societal level. We are examining how characteristics of groups/individuals (i.e. gender, time spent on Speak Agent) impact change in scores from one time to another time. The coefficient in a time series model tells us what to expect if Student A were to change her Speak Agent usage, how much would her own score change. Concretely, if Student A were to increase the number of activities he/she completed from quarter 1 to quarter 2 by 3 activities, for example, we can predict that Student A's score would increase by 5 points.

Statistically Significant Results

Statistical significance is determined by the p-value - a statistical measure used to determine the likelihood that an observed outcome is the result of chance. Therefore, a smaller p-value means that there is stronger evidence that the result did NOT happen by chance. A p-value of 0.05 or lower is generally considered statistically significant. P-values are often measured in increments of $p<0.05$, $p<0.01$, and $p<0.001$. With the latter, we are 99.99% certain that the result did not happen by chance.

FINDINGS

What is the effect of Speak Agent on student scores in Math and English Language Proficiency?

MATH BENCHMARK SCORES

The cross-sectional analysis below presents the results of the effect of Speak Agent usage on Benchmark scores at the student level at a constant point in time. That is, controlling for other possible explanatory variables of Math Benchmark scores, the results here show the independent effect of Speak Agent on Math Benchmark scores.

Finding 1.

The primary finding is that the Number of Speak Agent activities completed is positively associated with the overall Percent Correct on the Math Benchmark Assessment.

The results in Table 3 show that the regression coefficient is 0.010 and statistically significant at $p<0.001$, which means that, on average, every additional activity completed on Speak Agent corresponds with a 1% higher score on the Math Benchmark Assessment. Said differently, and illustrated in Figure 1, a student who completes one activity on Speak Agent is expected to get about 17.8% correct, a different student that completes two activities is expected to get 18.8% correct, while a student who completes 10 activities on Speak Agent is predicted to get about 27.1% correct.

Table 3: Regression Results for Math Benchmark Scores

Speak Agent Explanatory Variables	Regression Coefficients
# Activities Completed on Speak Agent	0.010***
Time Spent on Speak Agent (hours)	0.002**
Correct Answers on Speak Agent	0.000
STUDENT-LEVEL CONTROL VARIABLES	
Grade	-0.005
SCHOOL-LEVEL CONTROL VARIABLES	
School Title	0.007*
Total Enrollment	0.000
Attendance Rate	0.014*
% Proficient in ELA	0.001*

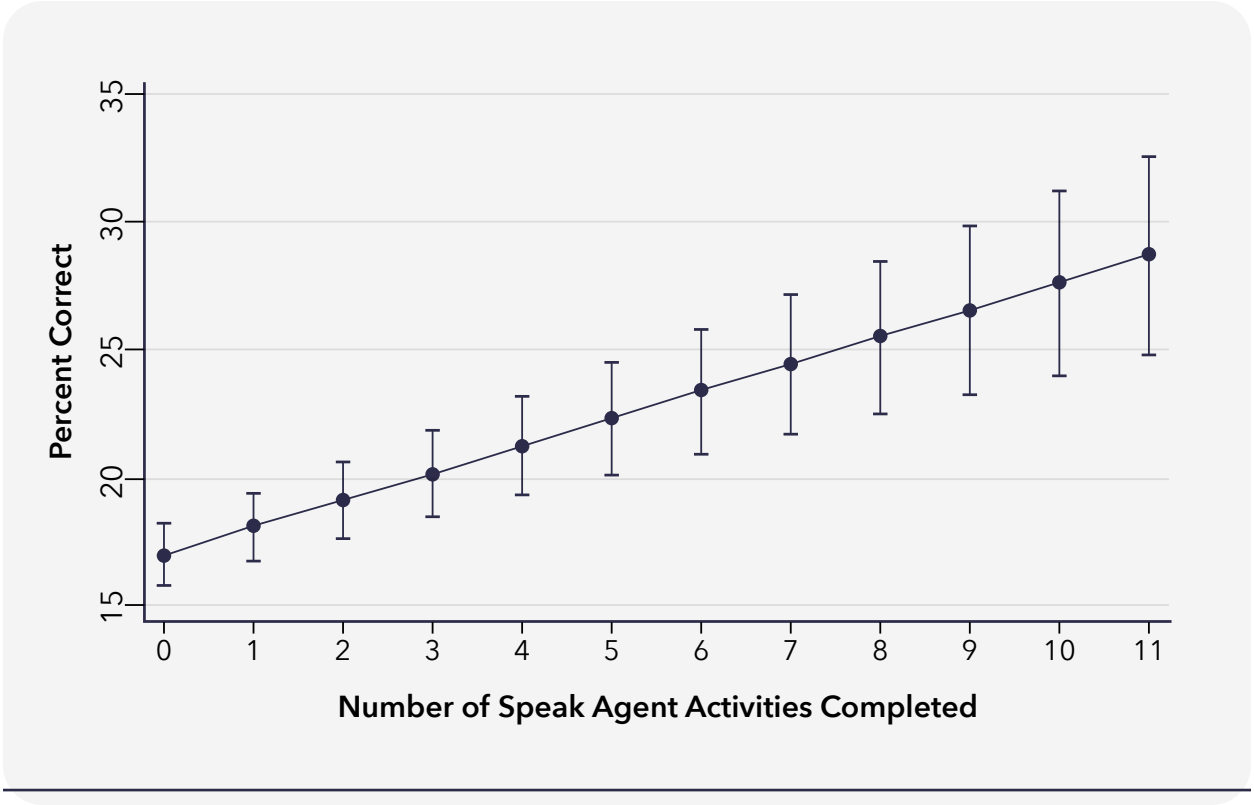
Key

- N = 45,594
- *** p<0.001
- ** p<0.01
- * p<0.05
- R^2 = 0.0459

Notes

- *OLS Estimation with clustered standard errors at school level (not shown). Multi-level models yield the same results.*

Figure 1. Predicted Benchmark Scores



Note: On graphs such as Figure 1, the way to tell that the relationship is statistically significant is to check where the confidence intervals don't overlap.

For example, in Figure 1, the confidence interval at 11 and the confidence interval at 0 do not overlap, which means the plotted relationship is statistically significant. There are, therefore, meaningful differences among Speak Agent users dependent on the level of use illustrated with the line.

Finding 2.

The effect of Speak Agent on Math Benchmark score is among those students with a proficiency level in the 3.0 to 3.9 and 4.0 to 4.9 range, as presented in Table 4. There is no detectable variance by race/ethnicity on Math Benchmark scores, as the sample only includes a small selection of multilingual learners that took the Benchmark assessment in quarter 3.

Table 4. Effect sizes of Speak Agent on Math Benchmark Scores, by Proficiency Level

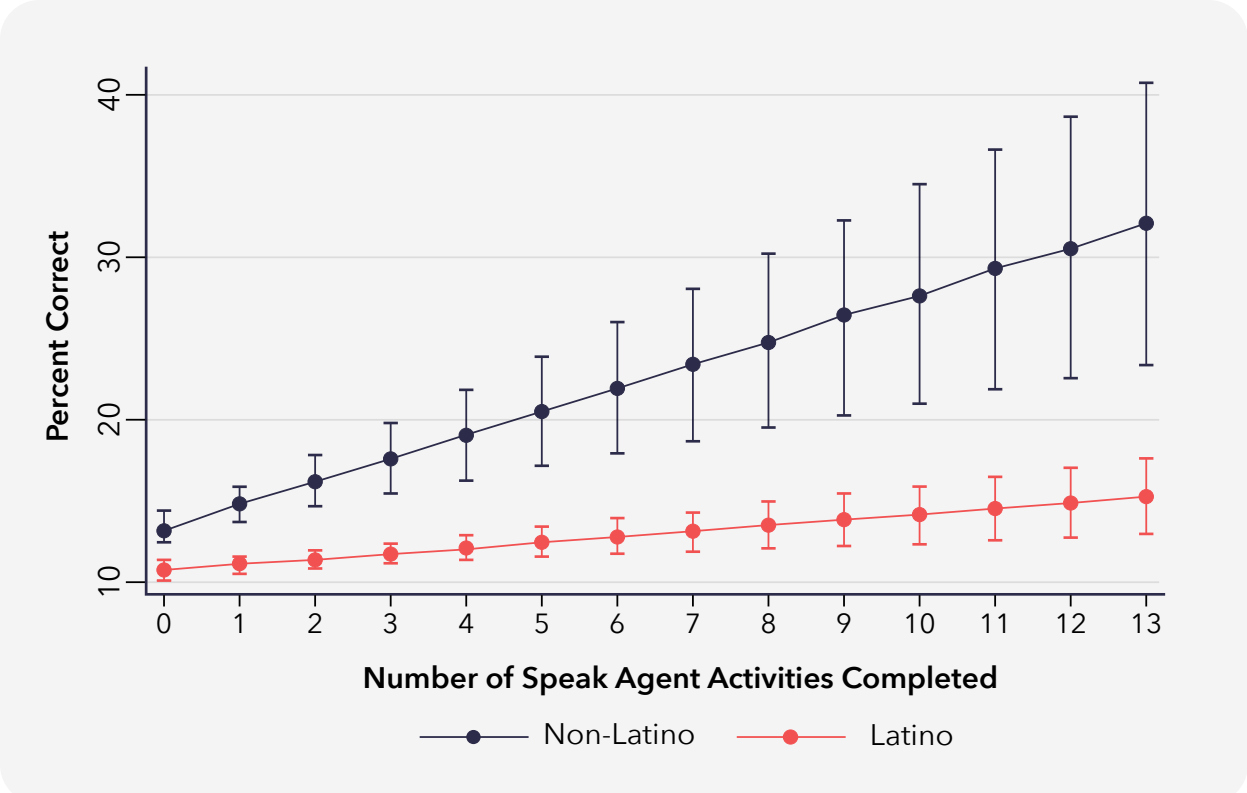
Proficiency Level	Effect Size	Sample Size (N)
Level 1.0 - 1.9	0.001	1,296
Level 2.0 - 2.9	0.002	2,235
Level 3.0 - 3.9	0.004**	3,990
Level 4.0 - 4.9	0.007*	865

Finding 3.

To assess the difference by race/ethnicity with Math Benchmark scores, the sample size drops dramatically. As noted earlier, race/ethnicity was only available for multi-lingual learners. Still, there was some overlap of English learners who took the ACCESS assessment and Math Benchmark assessment and those that only took the Math Benchmark assessment. Thus, the sample used to assess differences by race/ethnicity will include only multilingual learners.

The effects of Speak Agent are different for Latinos compared to non-Latinos (all multilingual learners), as illustrated in Figure 2. The blue line illustrates the effect for non-Latinos (N=865) and the red line illustrates the effect for Latinos (N=7,870). The blue line is steeper than the red line, demonstrating that for non-Latinos, there is a larger effect of completing Speak Agent activities on Math Benchmark scores than for Latinos. The effect size for multilingual non-Latinos is 0.015*** and the effect size for multilingual Latinos is 0.004***.

Figure 2. Predicted Benchmark Scores by Latino Identification



Other Analysis

- There is some evidence that Time on Task (hours) is positively associated with Benchmark Scores. On average, for every 5 hours of Speak Agent usage, scores are expected to be 1% higher among the population of users. This suggests that Time on Task itself is not as strong an indicator of progress as is activity completions, since time may be spent in more effective or in less effective ways.
- There is no evidence that Correct Answers or any of the other usage domains on Speak Agent are associated with Benchmark Scores.
- There is no effect of Speak Agent on Math Benchmark Scores among multilingual learners with an IEP status while there is an independent effect (0.005***) of Speak Agent on those multilingual learners without an IEP status. This relationship is true regardless of race/ethnicity.

TIME-SERIES ANALYSIS

In the time-series models in this section, the CHANGE of scores from one quarter to the next is assessed. Here, we can say that for any particular student, increasing or decreasing usage on Speak Agent would result in an increase or decrease in his/her own score.

We control for students' baseline scores and baseline Speak Agent usage (from quarter 1) to account for the fact that students with high scores, for example, have less room for growth than students with lower scores. The results show, therefore, the independent effect of Speak Agent on the change in individual student scores.

Finding 1.

The primary finding is that an increase in the number of activities completed on Speak Agent has a positive effect (increase) on change in Benchmark scores. The regression coefficient for Speak Agent Activities completed is 0.006 and statistically significant at the $p<0.001$ level. This means that for every additional activity a student completes on Speak Agent, as compared to the previous quarter, his/her score is expected to increase by 0.6 percentage points. As shown in Figure 7, if a student completed 10 additional activities on Speak Agent over one quarter, his/her Benchmark score would increase by 6 percentage points.

Table 5. Regression Results for CHANGE in Math Benchmark Scores

Speak Agent Explanatory Variables	Regression Coefficients
Change in # Activities Completed on Speak Agent	0.006***
Change in Time Spent on Speak Agent (hours)	0.001
Change in Correct Answers on Speak Agent	0.000***
STUDENT-LEVEL CONTROL VARIABLES	
Grade	0.003
SCHOOL-LEVEL CONTROL VARIABLES	
School Title	0.011**
Total Enrollment	0.000
Attendance Rate	0.000
% Proficient in ELA	0.001*
% White Population	-0.001**

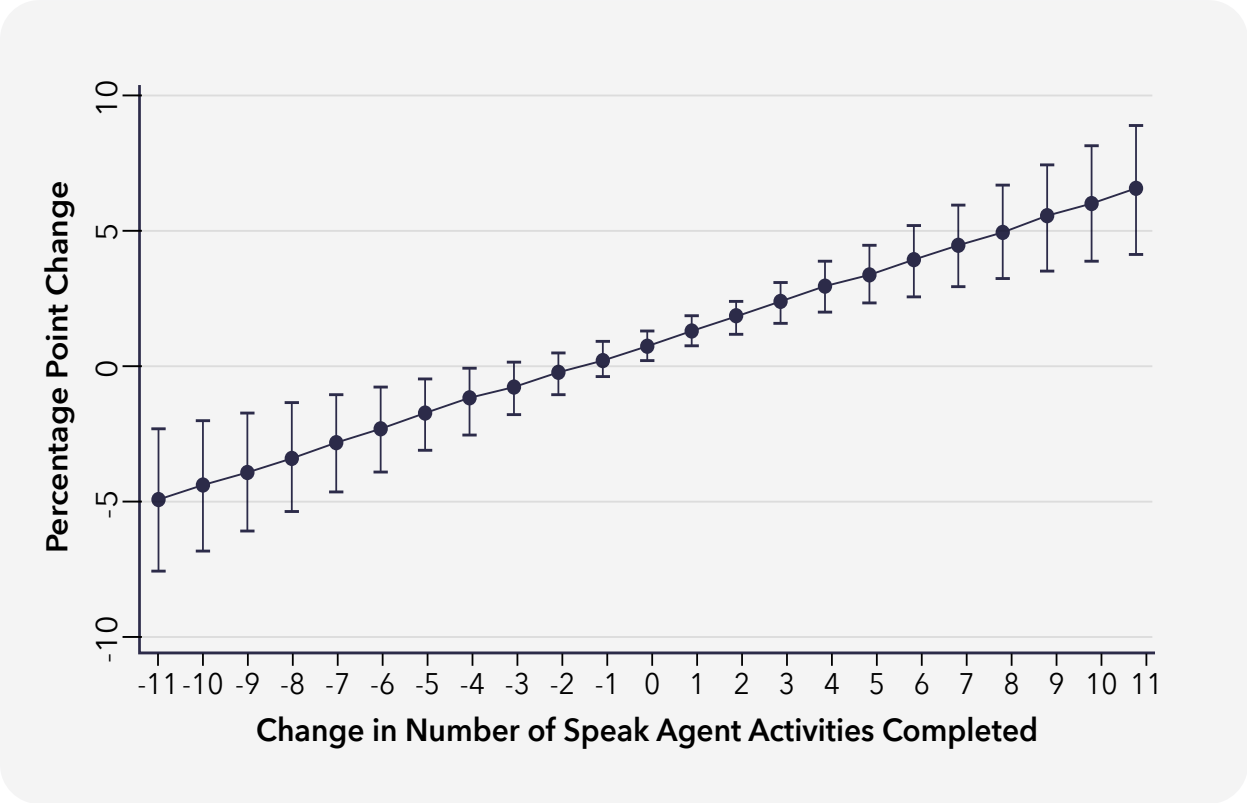
Key

- N=23,165
- *** p<0.001
- ** p<0.01
- * p<0.05
- R^2 = 0.2724

Notes

- *OLS Estimation with clustered standard errors at school level (not shown).*
- *A lag of Activities Completed and Percent Correct is included but not reported.*

Figure 3. Predicted CHANGE in Benchmark Scores

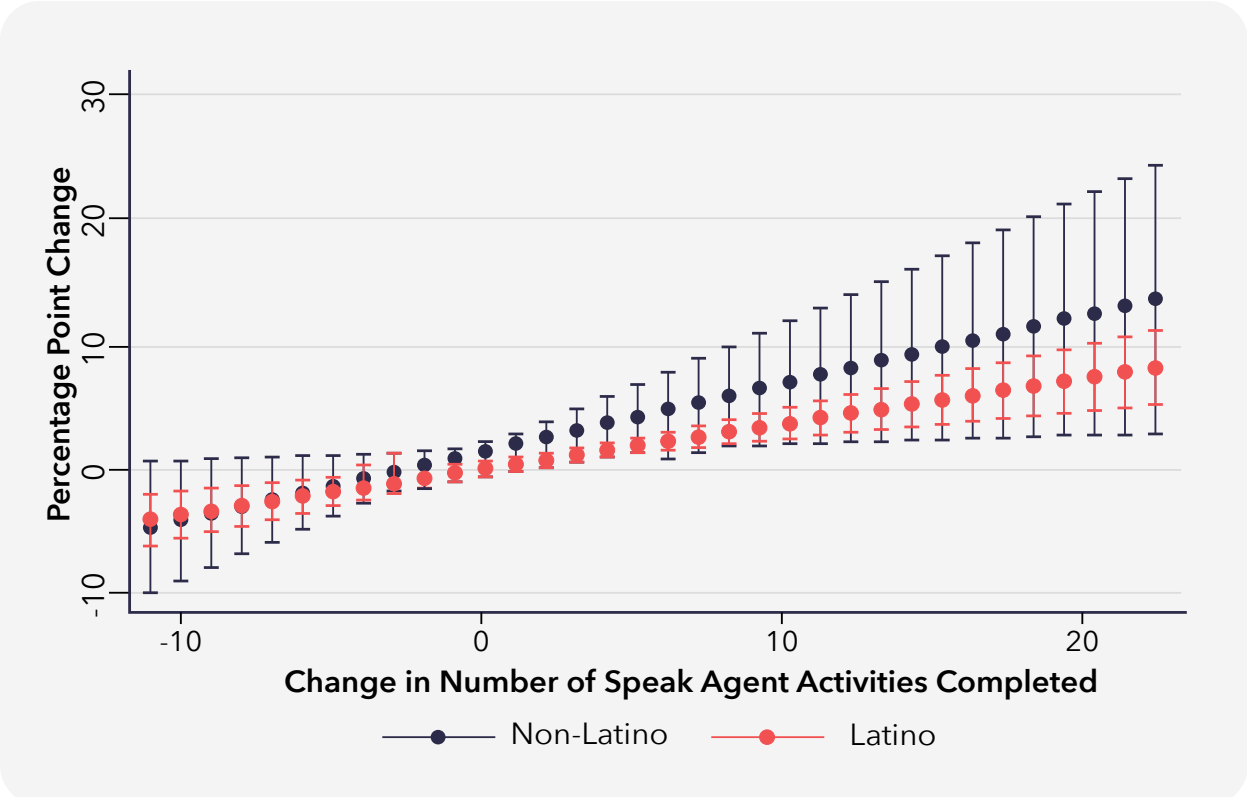


Finding 2.

Recall that gender and Latino identification are measures only available for those students that took the ACCESS assessment and are, therefore, all multilingual learners. The results here evaluate, therefore, the potentially different effect sizes of Speak Agent on Math Benchmark scores by race/ethnicity among only multilingual learners.

The results here show that there is no statistical difference in the effect size among Latinos and Non-Latinos. Unlike in the above analyses of race/ethnicity differences, the analysis here is able to control for baselines - or starting points - and compare whether there are differences among races in how much students gained from one quarter to the next. As illustrated in Figure 8, the gains are very similar and not statistically different. Both groups made statistically significant gains. For non-Latinos (N=447), the effect size is 0.009** and for Latinos (N=4,021), the effect size is 0.003***. While the effect size is greater for non-Latinos, that difference is not statistically different in large part due to more variation among the non-Latino population.

Figure 4. Predicted Change in Benchmark Scores by Latino Identification



WIDA ACCESS SCORES

The cross-sectional analysis below presents the results of the effect of Speak Agent usage on WIDA scores at the student level. That is, controlling for other possible explanatory variables of WIDA ACCESS scores, the results here show the independent effect of Speak Agent on WIDA ACCESS scores.

Finding 1.

The primary finding is that the Number of Speak Agent activities completed is positively associated with Overall Scale Scores on the ACCESS for ELLs assessment. The results in Table 6 show that the regression coefficient is 1.708 and statistically significant at $p<0.001$, which means that, on average, every additional activity completed on Speak Agent corresponds with 1.708 more points on the overall ACCESS score. Said differently, and illustrated in Figure 5, a student who completes one activity on Speak Agent is expected to get a score of about 329.54, a student who completes two activities is expected to get a score of 331.25, while a student who completes 10 activities on Speak Agent is predicted to get a score of 344.91.

Table 6. Regression Results for Overall ACCESS Scores

Speak Agent Explanatory Variables	Regression Coefficients
# Activities Completed on Speak Agent	1.708***
Time Spent on Speak Agent (hours)	0.531
Correct Answers on Speak Agent	0.007
STUDENT-LEVEL CONTROL VARIABLES	
Grade	4.178***
Female (vs. Male)	3.728***
Latino (vs. Non-Latino)	-7.107
SCHOOL-LEVEL CONTROL VARIABLES	
School Title	-2.018
Total Enrollment	-0.012***
Attendance Rate	4.025**
% Proficient in ELA	0.359**

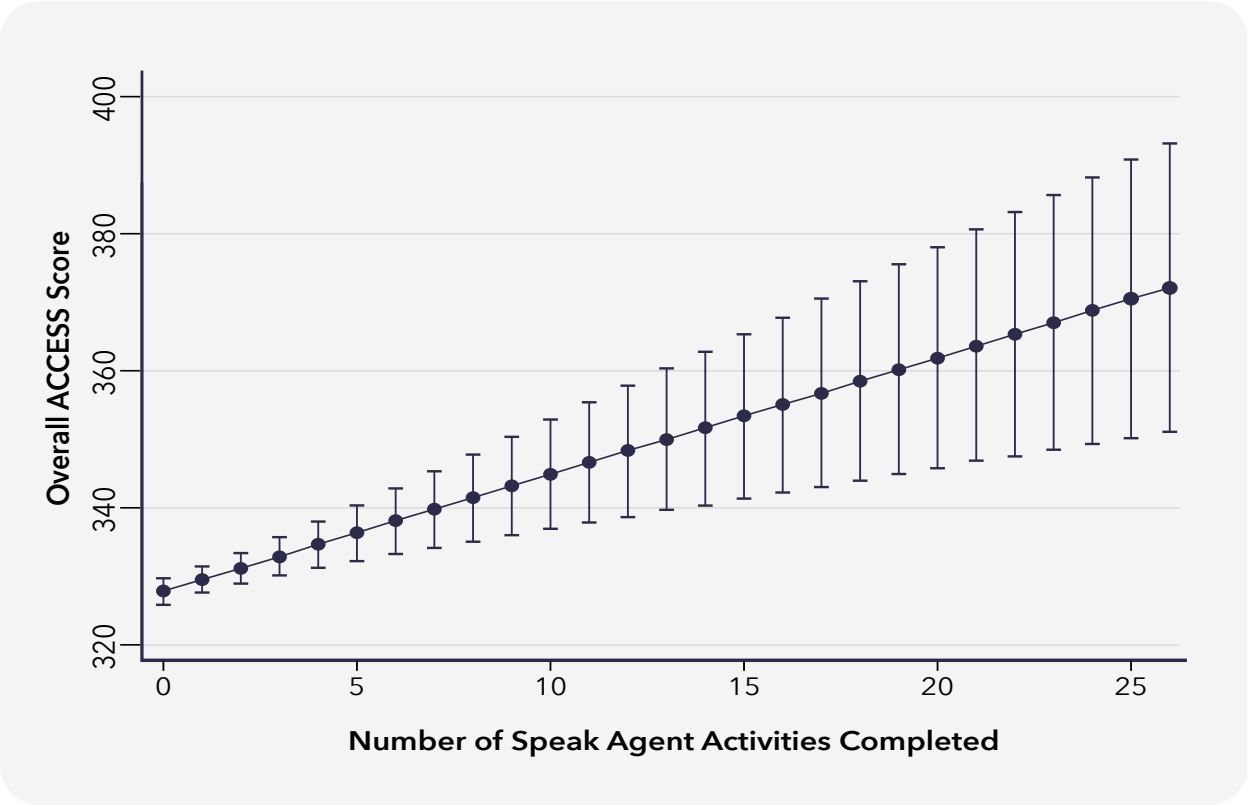
Key

- N = 5,134
- *** p<0.001
- ** p<0.01
- *p<0.05
- R^2 = 0.0411

Notes

- *OLS Estimation with clustered standard errors at school level (not shown). Multi-level models yield the same results.*
- *Speak Agent Explanatory Variables are the sum from quarter 1 and quarter 2 and Overall Score is from quarter 3.*

Figure 5. Predicted Overall ACCESS Scores



Finding 2.

The same effect of Speak Agent applies to every specific dimension on WIDA ACCESS scores. The cross-sectional regression results in Table 7 show that there was an independent, statistically significant positive effect of the Number of Speak Agent Activities Completed on every dimension of WIDA ACCESS scores. The effect was largest on the Speaking Score (2.15***) and Oral Score (2.072***) followed by Listening Score (1.984***).

Table 7. Regression Coefficients for Effect of Speak Agent Activities Completed on Specific Dimensions of WIDA ACCESS Scores

Separate regression model ran for each WIDA ACCESS dimension below

WIDA Dimension	Regression Coefficient
Listening Score	1.984***
Reading Score	1.317***
Speaking Score	2.150***
Writing Score	1.801***
Comprehension Score	1.517***
Literacy Score	1.561***
Oral Score	2.072***

Key

- *** p<0.001
- ** p<0.01
- *p<0.05

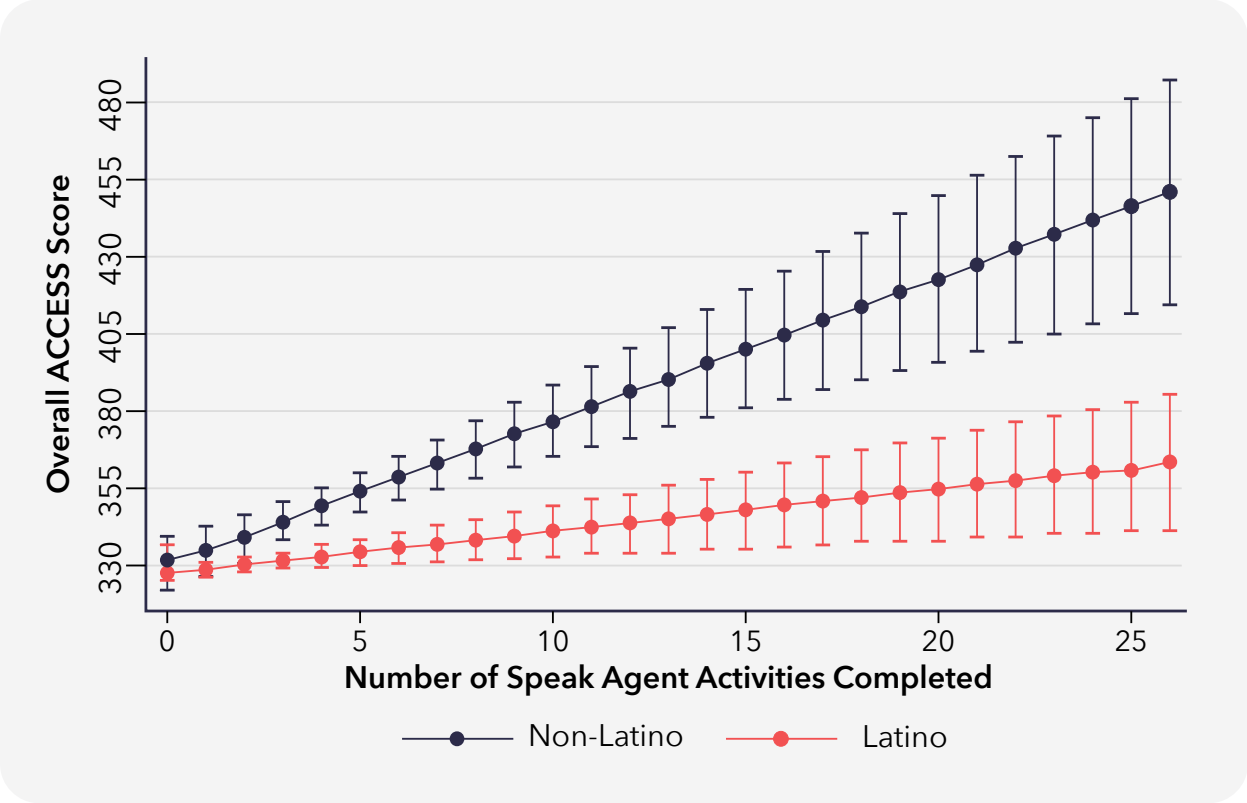
Notes

- OLS Estimation with clustered standard errors at school level (not shown). Multi-level models yield the same results.
- Speak Agent Explanatory Variables are the sum from quarter 1 and quarter 2 and Overall Score is from quarter 3.
- Control variables are the same as those in Table 6 (results are the same but not reported).

Finding 3.a

The effects of Speak Agent are different for Latinos compared to non-Latinos. The results here use the same model specification as in Table 6 but with an added interaction term to assess the difference between Latinos and non-Latinos. In Figure 6, the blue line illustrates the effect for non-Latinos (N=471) and the red line illustrates the effect for Latinos (N=4,663). The blue line is steeper than the red line, demonstrating that the effect of completing Speak Agent activities on WIDA ACCESS overall score is greater for non-Latinos. The effect size for non-Latinos is 4.585*** while the effect size for Latinos is 1.384**.

Figure 6. Overall ACCESS Score by Latino Identification



Finding 3.b

Figure 7 further illustrates the difference in effect size among four categories of race/ethnicity: Latino, Black, White, and Asian. The latter three groups include only those that do not also identify as Latino. As observed in Figure 7 and presented in Table 8, the effect size for the White population is significantly higher than all other races and statistically higher than the Black and Latino populations. The Asian and Black populations have similar effect sizes and remain statistically higher than the Latino population. Importantly, the effect size for all four groups is statistically significant and evaluates scores at one point in time (quarter 3) and does not, therefore, take into account changes/gains from a baseline starting point (i.e. previous years' scores).

Figure 7. Overall ACCESS Score by Race

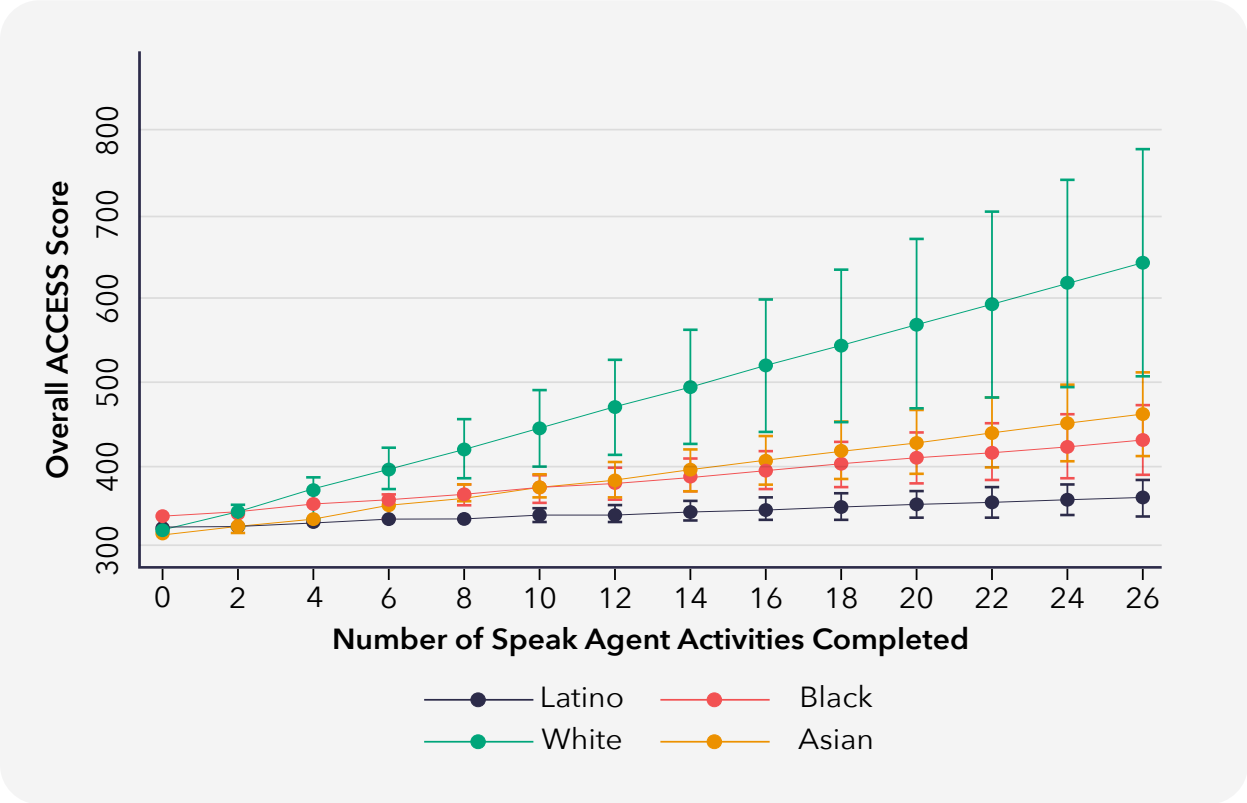


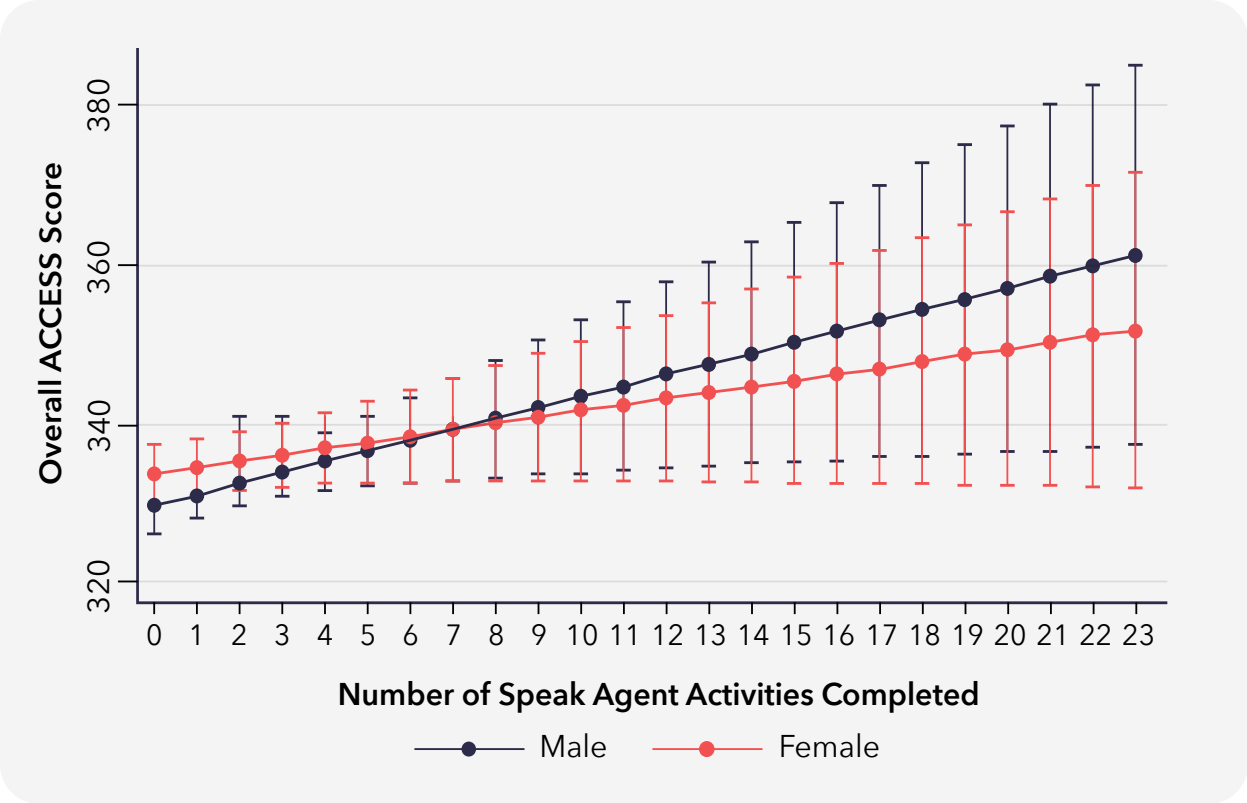
Table 8. Effect Size of Speak Agent on Overall ACCESS Score by Race/Ethnicity

Race	Effect Size	Sample Size (N)
Latino	1.384**	4,663
Black	3.649***	259
Asian	5.036***	172
White	11.437**	40

Finding 4.

The effect of gender has an independent effect on WIDA ACCESS scores: girls are more likely to have higher scores than boys. The effect of Speak Agent is, however, no different for boys than girls: the effect of completing Speak Agent activities is statistically significant and positive for both girls and boys (Figure 8).

Figure 8. Predicted Overall ACCESS Score by Gender



Finding 5.

The effect of Speak Agent on Overall ACCESS score is among only those students with a proficiency level in the 1.0 to 1.9 range, as presented in Table 9.

Table 9: Effect sizes of Speak Agent on Overall ACCESS Scores, by Proficiency Level

Proficiency Level	Effect Size	Sample Size (N)	Racial Difference
Level 1.0 - 1.9	0.640**	1,014	No racial differences
Level 2.0 - 2.9	0.129	1,299	No racial differences
Level 3.0 - 3.9	0.168	2,260	No racial differences
Level 4.0 - 4.9	0.029	540	No racial differences

Other Analysis

- There is no evidence of a non-linear relationship. While it is reasonable to believe that there would be diminishing returns at some level of Speak Agent usage, the data available here does not conform to that hypothesis. This is likely due to the fact that the data is extremely skewed towards 0 and we have too much variation at the higher levels of Speak Agent to confirm a statistically significant non-linear relationship.
- There is no evidence to suggest that the more views, speaks, reads, listens, or writes a student completes on Speak Agent, independent of activity completion, is associated with higher (or lower) WIDA scores on the respective domain.
- There is also no evidence that Time on Task or Correct Answers on Speak Agent is associated with ACCESS Scores overall or on each specific domain.
- There is no effect of Speak Agent on ACCESS scores for multilingual students with an IEP status (N=757) while there is an independent effect (1.834***) of Speak Agent on those students without an IEP status (N=4.377). This relationship is true regardless of race/ethnicity. It should be noted that none of the White population has an IEP status. The race/ethnicity differences observed in Figure 3 remain the same among those without an IEP status. Data for IEP status was not provided for non-multilingual students.

KEY TAKEAWAY FOR WIDA ACCESS SCORES

- The number of activities completed on Speak Agent is the feature of Speak Agent that is most likely to have an independent, positive effect on ACCESS scores overall and on each specific domain.
- The effect of Speak Agent is greater for non-Latinos than Latinos.
- The effect of Speak Agent is the same for boys and girls.