prodigy

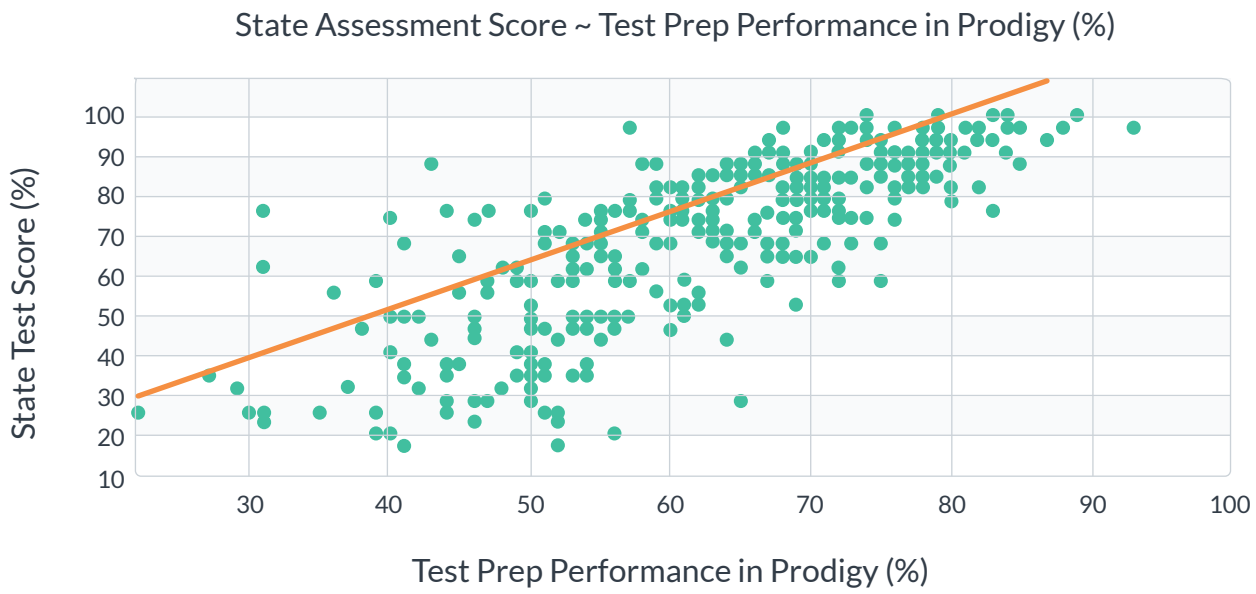
Evaluation of Prodigy

A Correlational Study on the Effectiveness of Prodigy on Improving Math Performance

Jiawen Chen, PhD
Kevin DeSimone, PhD
Saajid Mohammed

Executive Summary

The purpose of the current study was to evaluate Prodigy's impact on students' mathematics learning. Using data from a sample of Grade 4 students in the southern U.S. from the 2018-2019 school year, quantitative analyses were conducted with statistical controls in place. The findings showed that students who had greater usage and better performance in Prodigy scored significantly higher on the state standardized math test. In addition, we found that metrics derived from Prodigy-collected user answers also provided an accurate assessment of student math abilities. Even though causal claims cannot be made due to the correlational nature of the study, the overall findings suggest that Prodigy usage is positively associated with enhanced learning outcomes.



JOHNS HOPKINS
SCHOOL of EDUCATION

We reviewed the data analyses and findings reported by Chen, DeSimone, and Mohammed in their "Correlational Study on the Effectiveness of Prodigy on Improving Math Performance" from February 2020. We believe the study meets the ESSA Tier 3 "Promising" standards of evidence. Importantly, results indicated that Prodigy usage is significantly and positively associated with improved performance on the state standardized assessment. In addition, performance on Prodigy's test preparation questions significantly predicted performance on the state standardized assessment after controlling for prior mathematics achievement and socioeconomic background.

Jennifer R. Morrison, Ph.D.
Associate Research Scientist
Center for Research and Reform in Education
School of Education
Johns Hopkins University
JRMorrison@jhu.edu

Michael A. Cook, Ph.D.
Senior Education Researcher
Center for Research and Reform in Education
School of Education
Johns Hopkins University
MACook@jhu.edu

Background

In the fourth quarter of 2019, Prodigy conducted a correlational study with approximately 450 Grade 4 students from a school district in a southern state in the U.S. The primary research question was: What impact does Prodigy usage have on students' mathematics learning over one school year? To assess end-of-year learning, we used the state standardized mathematics test scores from Spring 2019. The predictor variables were derived from students' Prodigy usage and performance metrics. It is hypothesized that, after controlling for prior math abilities and socioeconomic background, greater usage of and more engagement with Prodigy will have a significant and positive effect on students' end-of-year standardized math test scores. This study is part of the ongoing effort by Prodigy to assess its educational efficacy. Results from this study meet Tier 3 evidence as set out by the Every Student Succeeds Act (ESSA), which requires evidence-based interventions to demonstrate "a statistically significant effect on improving student outcomes or other relevant outcomes based on promising evidence from at least one well-designed and well-implemented correlational study with statistical controls for selection bias" (U.S. Department of Education, 2016, *p.* 7).

Prodigy is a supplemental digital game-based learning program that aspires to make learning math fun for students (Grade 1 to 8) through its immersive environment, attractive graphics and innovative format. As the main objective of the game storyline, students must save the world by collecting six magic stones scattered around the game. Along the journey, they engage in and win battles by correctly answering math questions. Based on students' answers, the game's algorithm identifies students' abilities and knowledge gaps and introduces either prerequisite skills or more advanced concepts that are ability-matched to enhance learning outcomes. Prodigy is currently used by over 50 million students worldwide.

Most broadly, we classify students' behavior in Prodigy in terms of engagement (e.g., number of questions answered) and performance (e.g., ratio of correct to total answers). These engagement and performance metrics are then used to make predictions about students' achievement, as assessed by end-of-year standardized math test results. Recent research on the effectiveness of traditional learning tools such as homework assignments suggests that effort may be as important a predictor of student achievement as usage (Fan, Xu, Cai, He, & Fan, 2017). To that point, in the current study efforts were accounted for vicariously through performance in Prodigy. The assumption was that getting more questions right was the result of a student having mathematical ability and the willingness to put in the effort to work through the questions.

Prodigy has a number of distinct algorithms (e.g., placement test, standard, test preparation) that could be used to serve math questions to users depending on the learning context. When a student enters Prodigy for the

first time, they self-select a grade level. Whenever a student is attached to a teacher, the teacher's self-selected grade is substituted for the student's self-selected grade. The student also completes a placement test, which allows the program to place the student at a grade level that matches their placement test performance. During regular gameplay, the difficulty of the questions generated by the standard algorithm is dependent upon a combination of the student's grade, the placement test assessed grade, and the algorithm assessed grade determined by the student's recent performance. For each standard algorithm question, students are given a second opportunity if the first response was incorrect. In this study, only response accuracy on the first attempt was used as a predictor. In addition, the test preparation questions – created by Prodigy's expert content creators and aligned to each state's standardized test materials – are generated for students when teachers specifically assign this type of test to their students. Response accuracy on test preparation questions was used as another predictor in the study.

Methods

Sample

Prodigy built a learning partnership with a suburban school district in the southern U.S. The program was implemented district-wide as a supplemental learning tool in addition to the regular mathematics lessons at school. Teachers and school administrators received training from Prodigy prior to the program implementation and had continuous support from Prodigy throughout the academic year. As state assessment scores from 2018 and 2019 were made available for a group of 448 Grade 4 students from six schools in the school district, these students formed the study sample. Within the sample, 57% were Hispanic, 63% were eligible for the Free or Reduced Price Meals (FRPM) program, and 29% were of limited English proficiency (LEP).

Measures

Outcome Variable

The school district provided the percentile score on the state standardized mathematics test from Spring 2019, which served as the outcome variable in the study. The scores in the sample ranged from 18% to 100%.

Predictor Variables

Based on the available student information, school administrators matched the students with their Prodigy data for the 2018-2019 school year. A set of predictor variables were derived from the data. Specifically, information was available for all 448 students on their mean response accuracy (ranging from 50% to 100%) and total response time (in hours; ranging from 0.01 to 92.29 hours) on all standard algorithm questions answered, the number of standard algorithm questions answered (ranging from 2 to 5236 questions), self-selected/ teacher-determined grade level (ranging from Grade 1 to Grade 8) and grade level determined by the standard algorithm (i.e., algorithm grade; ranging from Grade 1 to Grade 7). Among all the students, 428 (95.5%) had a teacher profile attached to them, which meant that we could derive information about the teachers' Prodigy usage. A variable was computed by tallying all actions a teacher performed on the Prodigy platform (e.g., opening web pages, viewing student reports, assigning questions; ranging from 10 to 14,205 actions performed by 43 teachers). Another 342 students (76.3%) in the sample had completed test preparation questions. Their mean response accuracy on the test preparation questions ranged from 22% to 100%.

Control Variables

To account for students' math abilities prior to the school year, the study controlled for the students' percentile score on the 2018 state standardized math test (ranging from 16% to 100%). The study also controlled for demographic variations, as indicated by being eligible for FRPM program (dummy coded 1 for being eligible and 0 for being ineligible) and being LEP (dummy coded 1 for having limited English proficiency and 0 otherwise).

Analysis

The main analysis used to answer the research question was multiple regression. This was an appropriate analytic technique as learning outcomes are often considered as a function specified by a linear and additive combination of learning inputs (e.g., skills learned, individual characteristics, school and teacher characteristics, and family characteristics; Todd & Wolpin, 2003). In this study the learning outcome was performance on state standardized math test in 2019. The learning inputs were Prodigy usage and performance and individual and teacher characteristics. The goal of the study was to test the educational efficacy of Prodigy. Therefore, performance and usage in Prodigy as well as students' self-selected and algorithm grade were treated as the predictor variables. Prior math abilities as indicated by the 2018 state standardized test score, eligibility for free or reduced price meal, and English language ability were included as control variables. The outcome variable was first regressed on the full set of predictors and control variables. Subsequently, with backward elimination, nonsignificant variables were trimmed from the regression model (Royston & Sauerbrei, 2008). For students with teacher data and test preparation data, separate regression analyses were conducted.

Background

Descriptive statistics were computed for a preliminary understanding of Prodigy usage patterns and their associations with performance on state standardized tests (see Table 1). Students in this sample had a mean correct response rate of 73% on basic algorithm questions and 63% on test preparation questions, with the overall sample responses resembling a normal distribution (see Figure 1a and 1b). On average, students spent a total of 5.42 hours answering questions on Prodigy and answered 662.43 questions over the course of the school year. However, usage varied significantly among the students as implementation standards differed by teachers and schools. For example, there was a great amount of variation in time spent (*SD* = 8.48) and questions answered (*SD* = 672.57) by students. There was also much variation in the number of actions teachers performed on Prodigy (*M* = 2095.49, *SD* = 3007.30). To reduce the variability of these variables and enhance their interpretability, the variables for time spent answering questions and number of questions answered were recoded such that the bottom one-third of the values were categorized as "low", the top one-third as "high", and the middle one-third as "moderate". The number of teacher actions was recoded such that the bottom 25% values were categorized as "low", the top 25% as "high", and the middle 50% as "moderate". In addition, the algorithm grade (*M* = 3.46, *SD* = .94) was lower than students' self-selected/teacher-determined grade (*M* = 4.25, *SD* = .84) and students' actual grade (i.e., Grade 4).

Table 1

Means, standard deviations, and bivariate correlations on study variables

	1	2	3	4	5	6	7	8	9
2019 state standardized test score	-								
2018 state standardized test score	.80	-							
Response accuracy[a]	.44	.38	-						
Hours answering questions[a]	.18	.17	-.06	-					
Number of questions answered[a]	.12	.16	-.16	.69	-				
Algorithm grade[a]	.51	.46	-.29	.08	.02	-			
Self-select grade	-.01	-.01	-.03	-.02	-.03	.18	-		
Teacher activity	.08	.11	.13	.15	.14	.08	.01	-	
Response accuracy[b]	.77	.72	.47	.14	.07	.56	-.04	.10	-
Mean (*SD*)	.70 (.20)	.69 (.21)	.73 (.07)	5.42 (8.48)	662.43 (672.57)	3.46 (.94)	4.25 (.84)	2095.49 (3007.30)	.63 (.13)

Note. [a]Variables derived from user data on standard algorithm questions.

[b]Variable derived from user data on test preparation questions.

Figure 1 a) Distribution of Response Accuracy on Basic Algorithm Questions

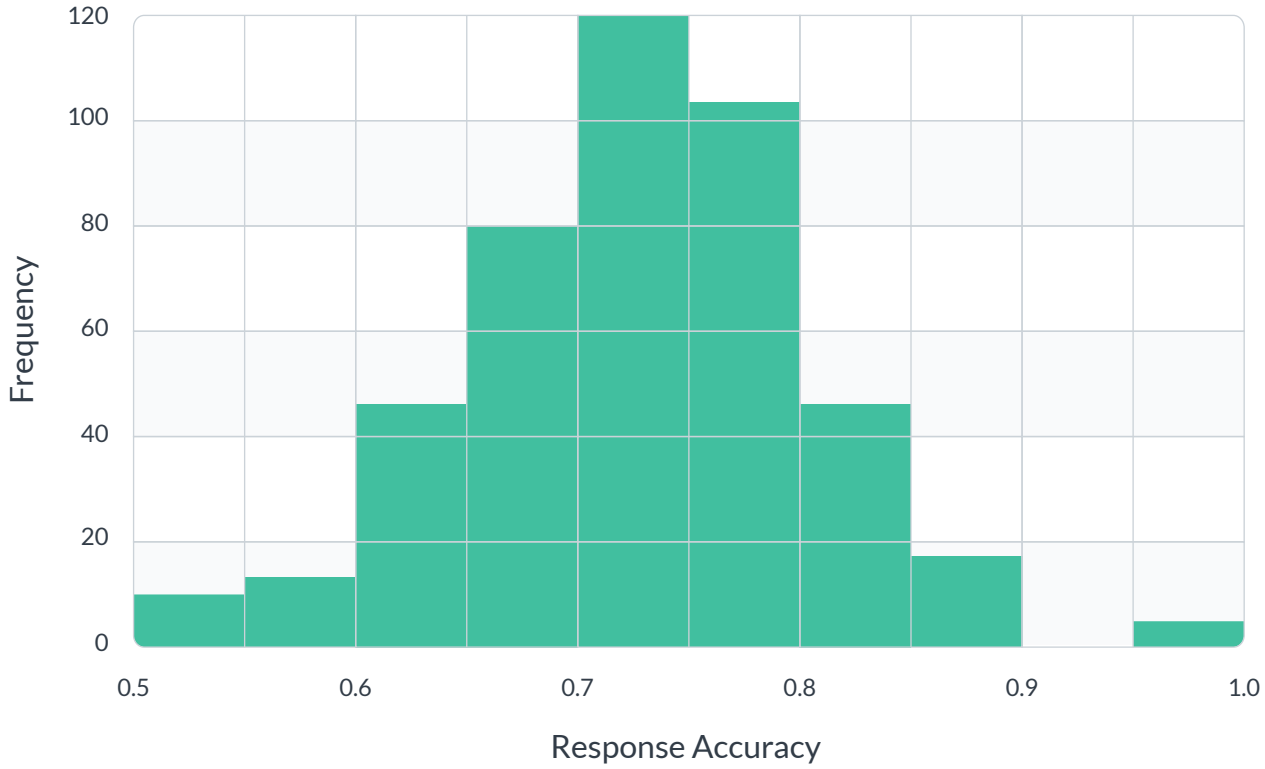


Figure 1 b) Distribution of Response Accuracy on Standardized Test Preparation Questions

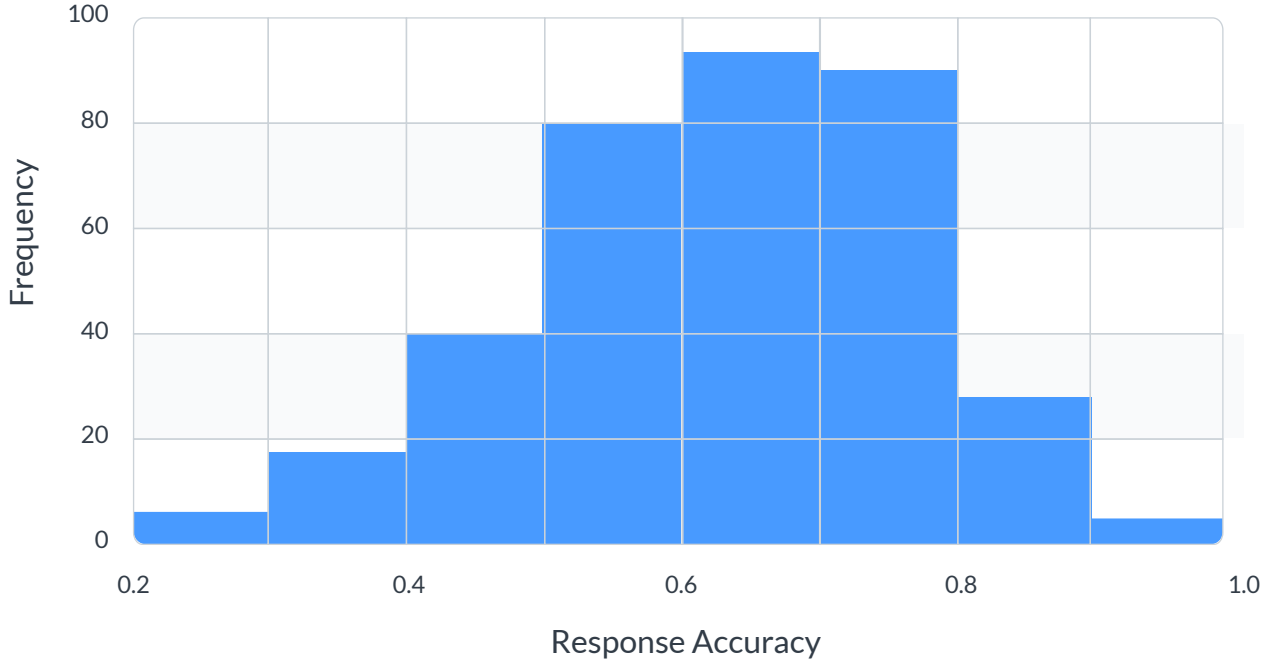


Figure 1. a) Distribution of response accuracy on standard algorithm questions. b) Distribution of response accuracy on test preparation questions.

The correlations suggested that all variables were associated in the expected directions (see Table 1). For example, the number of standard algorithm questions answered ($r = .12$, $p = .01$) and time spent answering those questions ($r = .18$, $p < .001$) both showed small but significant correlations with the 2019 standardized test score, which means students who played Prodigy more also performed better on the standardized test in comparison to students who played less (see Figure 2). Response accuracy on standard algorithm questions ($r = .44$, $p < .001$) and on test preparation questions ($r = .77$, $p < .001$) showed moderate to strong correlations with the 2019 standardized test score. It is worth noting that the strength of the correlation between response accuracy on test preparation questions was comparable to that of the correlation between 2018 and 2019 standardized test scores ($r = .80$, $p < .001$), suggesting the potency of the test preparation questions as an accurate assessment tool of students' math abilities. Algorithm grade also correlated significantly with the 2019 standardized test score ($r = .51$, $p < .001$) whereas self-selected/teacher-determined grade did not ($r = -.01$, $p = .84$).

Figure 2

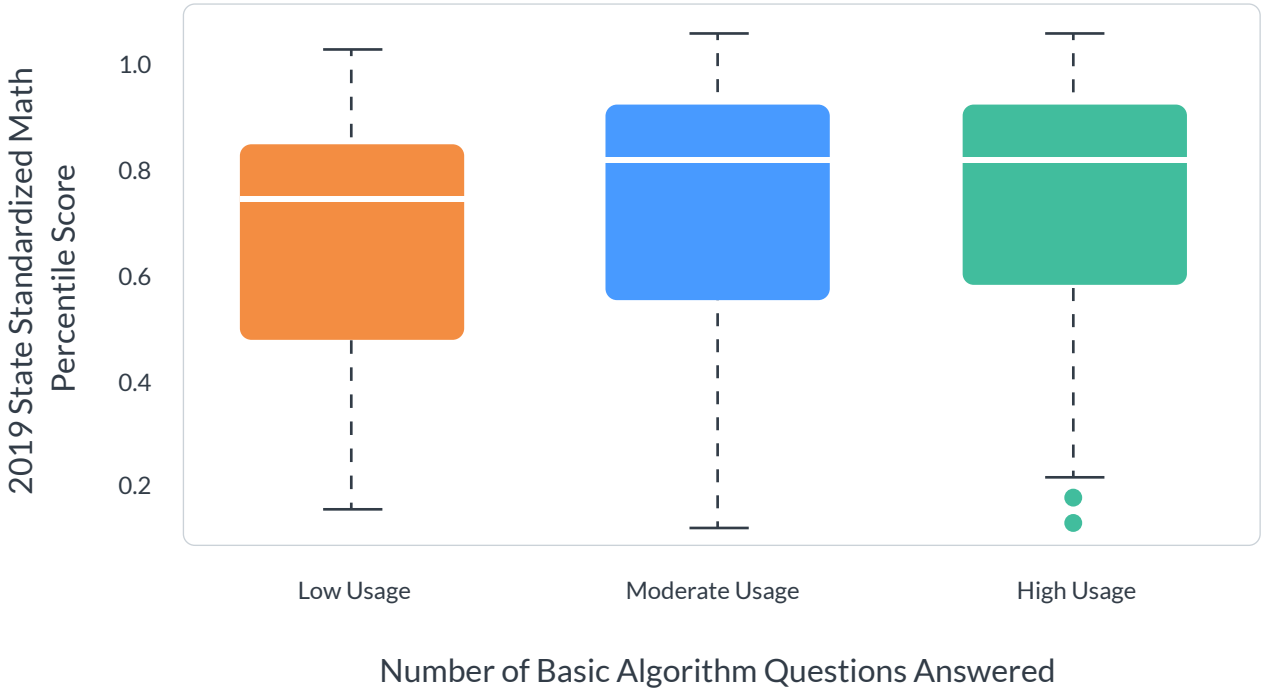


Figure 2. Prodigy usage and state standardized math test performance.

Prodigy Impact on Math Achievement

Table 2 shows the main regression models that examined the impact that Prodigy had on students' math learning. In Model 1, all predictor and control variables were included to predict state standardized test percentile scores in 2019. In Model 2, only significant variables were retained while nonsignificant variables were trimmed.

After controlling for 2018 standardized test scores, Prodigy usage and performance remained significant in predicting 2019 standardized test scores. Specifically, in terms of Prodigy performance, 10% higher in response accuracy on standard algorithm questions was associated with 3.7% higher standardized test scores ($b = .37$, $p < .001$). With respect to Prodigy usage, students who were in the bottom one-third in the number of questions answered scored 3% lower on the standardized test ($b = -.03$, $p = .02$) than students in the top two-thirds of the sample. In addition, being one grade level higher as determined by Prodigy's algorithm was associated with scoring 3% higher on the standardized test ($b = .03$, $p < .001$). Finally, students with limited English proficiency scored 3% lower on the standardized test ($b = -.03$, $p = .02$). Variables in Model 2 explained 69% of the variance in the 2019 state standardized test percentile scores.

Table 2 – Regression predicting 2019 state standardized test percentile scores

	Model 1		Model 2	
	b	*SE*	*b*	*SE*
Low number of questions answered[a]	-.02	.02	-.03*	-.01
Moderate number of questions answered[a]	.02	.02		
Response accuracy[a]	.35***	.08	.37***	.08
Low time answering questions[a]	-.002	.02		
Moderate time answering questions[a]	-.01	.02		
Algorithm grade	.03***	.01	.03***	.01
Self-select grade	-.01	.01		
2018 state standardized test	.65***	.03	.65***	.03
FRPM program	-.01	.01		
LEP	-.02	.01	-.03*	.01
Adjusted R^2	69%		69%	

Note. a) Variables derived from user data on standard algorithm questions.
***p < .001. **p < .01. *p < .05.

Table 3 shows the models that included variables for teacher activity and student test preparation response accuracy, respectively. Because not all students in the sample had data on teacher activity or test preparation answers, sample sizes for Models 3 (n = 428) and 4 (*n* = 342) were smaller than the previous models. Model 3 indicated that the level of teacher activity on Prodigy did not predict students' standardized test scores. On the other hand, Model 4 showed that students who had 10% higher response accuracy on Prodigy's test preparation questions also scored 5.1% higher on the state standardized test ($b = .51$, $p < .001$). After the addition of response accuracy on test preparation questions, limited English

proficiency was no longer a significant predictor of standardized test scores. A low number of standard algorithm questions answered became marginally significant (b = -.02, p = .09). With the inclusion of response accuracy on test preparation questions, the variables in Model 4 accounted for 73% of the variance in 2019 state standardized test scores.

Table 3 – Regression predicting 2019 state standardized test percentile scores

	Model 3		Model 4	
	(n = 428)	*SE*	*(n = 342)*	*SE*
Low number of questions answered[a]	-.03	.01	-.02[†]	-.01
Response accuracy[a]	.38***	.08	.26**	.09
2018 state standardized test	.65***	.03	.47***	.04
LEP	-.03*	.01	-.01	.01
Algorithm grade	.03***	.01	.02**	.01
Low teacher activity	.004	.02		
Moderate teacher actvity	.01	.02		
Response accuracy[b]			.51***	.07
Adjusted R^2	68%		73%	

Note. a) Variables derived from user data on standard algorithm questions.
b) Variable derived from user data on test preparation questions.
***p < .001. **p < .01. *p < .05. ,[†]p < .10.

Discussion

This study examined the effectiveness of Prodigy as a supplemental educational technology tool to boost math learning in a sample of Grade 4 students in the southern U.S. The findings supported the main hypothesis. In particular, more time spent answering questions and more questions answered in Prodigy were significantly correlated with better performance on state standardized math tests. Effort put into Prodigy as evidenced in performance on standard algorithm questions and test preparation questions positively predicted achievement on standardized tests after taking into consideration students' prior math achievement and their socioeconomic background. It is worth noting that performance on Prodigy's test preparation questions had comparable correlational strength with the standardized test score as the previous year's standardized test score. This strength of correlation is also comparable to the correlations found between NWEA's MAP, a widely recognized assessment, and the state standardized test examined in the present study (Li & Tran, 2017). While we strive to further validate the strong association between Prodigy's test preparation questions and other state level assessments, this finding has potential implications for teachers and schools that can use the test preparation questions as a tool to gauge student readiness before a state standardized test. In addition, grade level as determined by Prodigy's algorithm, but not self-selected/ teacher-determined by students, positively predicted test performance. It suggests that continued monitoring of student performance in Prodigy may shed light on students' progress, achievement, or struggle for teachers, school administrators, and parents, and give clues about students who may need interventions. Overall, these findings warrant the continued use of Prodigy for enhancing students' math learning.

Limitations

This study revealed significant and positive association between Prodigy usage and students' mathematics learning. However, there are several drawbacks that merit follow-up assessments to better evaluate Prodigy's efficacy. Due to the correlational nature of the study, there was no control group that did not use Prodigy. Having a treatment and a control group with students randomly selected into either group would provide stronger evidence to support Prodigy's efficacy. In addition, there was significant inconsistency in the implementation of Prodigy within the sample which might have dampened the significant results found in the study. Thirdly, we did not have information on Prodigy usage prior to the 2018-19 school year and were not able to control for it in our analyses. This may potentially affect the estimation of Prodigy's effect. Lastly, this study only examined a group of Grade 4 students from a southern U.S. state, future evaluations should include a more diverse sample of students consisting of different grade levels from different regions in the U.S.

Conclusion

In evaluating Prodigy's effectiveness as a tool to improve students' mathematics learning, this study revealed that both usage and performance in Prodigy are significantly and positively related to students' state standardized math test achievement. Prodigy could be a valuable tool for students' math learning and for teachers to gauge their students' abilities and test readiness. Future evaluation efforts that implement rigorous experimental research designs will add stronger evidence to the assessment of Prodigy's educational efficacy. Given the increased adoption of digital game-based learning technology in educational settings, Prodigy is a promising and attractive option for teachers, schools, and parents alike who are searching for an engaging and effective learning program.

References

Fan, H., Xu, J., Cai, Z., He, J., & Fan, X. (2017). Homework and students' achievement in math and science: A 30-year meta-analysis, 1986-2015. Educational Research Review, 20, 35-54.

Li, S., & Tran, S. (2017). 2017 concordance study: Linking the Texas STAAR scales to the NWEA MAP growth scales. NWEA Research.

Royston, P., & Sauerbrei, W. (2008). Multivariable model-building: A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. Chichester: John Wiley & Sons, Ltd.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. The Economic Journal, 113, F3-F33.

U.S. Department of Education. (2016). Non-regulatory guidance: Using evidence to strengthen education investments. Washington, DC: U.S. Department of Education. Retrieved from https://www2.ed.gov/policy/elsec/leg/essa/guidanceuseseinvestment.pdf