Effectiveness of Agile Mind's *Agile Assessment*

June 2020

Jefferson Education Accelerator contracted with Empirical Education to study the effectiveness of the implementation of Agile Mind's *Agile Assessment* in an urban school district in the Northeast United States during the 2016-17 school year. *Agile Assessment* is a system for building formative assessments that includes research-based items aligned to state standards.

This study focused on students enrolled in Algebra and examined the impact of *Agile Assessment* usage on student performance, as measured by the PARCC Algebra assessment. We investigated the following questions.

(1) Do students who use *Agile Assessment* perform better on the PARCC Algebra assessment than comparable students who did not use *Agile Assessment?*

(2) Is the impact of *Agile Assessment* different for students with different characteristics?

(3) Are differences in *Agile Assessment* usage associated with differences in student performance?

Results

(1) Do students who use *Agile Assessment* perform better on the PARCC Algebra assessment than comparable students who did not use *Agile Assessment*?

Students who used Agile Assessment performed better on the PARCC Algebra assessment than comparison students who did not use Agile Assessment.

We find that Algebra scores for students who actively used *Agile Assessment* are better than for comparison students. The result corresponds to a 12-percentile point gain on the PARCC Algebra assessment, adjusting for differences in student demographics and pretest between *Agile Assessment* students and comparison students.

As shown in Figure 1 (and Table 1), the effect size of the overall impact of *Agile Assessment* on the PARCC Algebra Assessment is 0.30, and we have strong confidence in this result ($p < .05$).
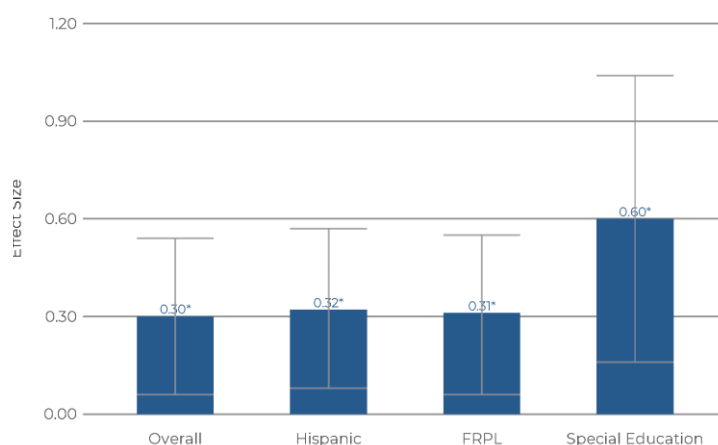

FIGURE 1. OVERALL IMPACT AND SUBGROUPS

Note. 95% confidence intervals convey that we have strong confidence of the results falling within that area.

(2) Is the impact of *Agile Assessment* different for students with different characteristics?

There is a positive impact of Agile Assessment for several student subgroups, including Hispanic students, students who qualify for free- or reduced-price lunch (FRPL), and Special Education students.

As shown in Table 1, there were significant positive effects for several student subgroups. Many of these were close to the average effect size of *Agile Assessment* for all users (0.30 with $p < .05$). The majority of students in the analytic sample are in 9th grade, Hispanic, non-English Language Learners (ELL), and qualify for FRPL; the low significance of the estimates for other student subgroups reflects the small sample sizes for those populations.

(3) Are differences in *Agile Assessment* usage associated with differences in student performance?

There is a positive association between average Agile Assessment scores and PARCC Algebra outcomes.

We estimated the effects of selected usage metrics on student PARCC Algebra assessment scores for all *Agile Assessment* users included in the impact study. We found that outcomes on the PARCC assessment are positively associated with the average score on the *Agile Assessment* benchmark tests. However, the predictive power of *Agile Assessment* is limited; adding the average score on *Agile Assessment* benchmark tests to the linear model increased its explained variance from 48% to 54%. It should be emphasized that since students and/or teachers have freedom in choosing the level of usage and the type of activities in *Agile Assessment*, none of these behavioral relationships can be considered causal. It is still a possibility that usage is affected by unmeasured student abilities or interest in using formative assessments.

LEVELS OF CONFIDENCE IN OUR RESULTS

Results are reported based on statistical calculations that give a measure of confidence expressed as a probability or *p* value. A low *p* value indicates a low probability that we would detect a difference like the one found in the study if no difference actually existed. A *p* value less than .05 gives us strong confidence in the result (a level conventionally called statistically significant). A *p* value between .05 and .15 gives us moderate confidence, while a *p* value between .15 and .20 gives us limited confidence. A *p* value greater than .20 gives no confidence.

STUDY DESIGN

The study compared achievement for students in two groups: one group of students in grades 9 through 12 who implemented *Agile Assessment*, and the other group who did not, adjusting for the differences in student characteristics. *Agile Assessment* was used in Algebra classes to deliver three benchmark assessments in November, January, and March. The experimental group included students who took at least one benchmark assessment, and the comparison group included students who did not use *Agile Assessment* at all.

PARTICIPANTS

The study used data from the 2016-17 school year in an urban school district in the Northeast United States. The district provided student data including unique student ID; school, teacher, and course data; student demographics; and Fall STAR Math pretest and Spring PARCC outcome data. We combined these data with data from the 2016-2017 *Agile Assessment* system, including the time spent, problems attempted, and problems correct on three or fewer benchmark assessments.

Conclusion

We found a positive impact of *Agile Assessment* on the PARCC Algebra assessment, and these impacts were significant across multiple student subgroups. Additionally, the average *Agile Assessment* score was found to have a positive association with student outcomes on the PARCC assessment.

CAUTIONS FOR INTERPRETING THESE RESULTS

Results shown in the figures and tables are not actual differences in test outcomes but estimates that adjust for the differences between users of *Agile Assessment* and a comparison group, and they should be interpreted as the hypothetical improvement in outcome for the average comparison student if they were in an *Agile Assessment* classroom. The actual outcomes for actual *Agile Assessment* students may vary depending on their characteristics.

Results reported as no difference do not imply that no real differences exist, but that a large study is needed to estimate them accurately.

This case study was conducted on behalf of the school district with the technical assistance of Empirical Education. In conducting or supporting the agency's conduct of the study, Empirical does not intend to generate evidence valid beyond the agency in which the case study was conducted.

Technical Details

DATA PREPARATION

The district provided student data for the 2016-2017 school year for all students enrolled in Algebra.

Agile Mind provided the district with student log data for the three benchmark Algebra assessments for all students in the district, which the district merged with their own data.

ANALYTIC SAMPLE

The analytic sample for the comparison study, shown in Table 2, consisted of students in grades 9 through 12 with both fall 2016 STAR Math pretest and spring 2017 PARCC Algebra outcome test scores. *Agile Assessment* students were matched using a stratified matching procedure, including the following covariates: gender, ethnicity, ELL status, eligibility for free- or reduced-price lunch, Special Education status, and STAR pretest score. In the group of *Agile Assessment* users, there were 764 students, and the comparison group had 146. The sample for the usage analysis was consistent with the impact study. All covariates had differences of less than .25 standard deviations between the comparison and *Agile Assessment* groups, as shown in Table 2.

ANALYSIS

We used a hierarchical linear mixed effects regression model, which accounts for the clustering of students within classes and within schools, adjusting for student demographics and pretest scores to compare performance for *Agile Assessment* and comparison students. Table 2 below displays the results.

The usage analysis was also performed using hierarchical linear mixed effects regression with the PARCC Algebra assessment as the outcome variable, and student characteristics, pretest, and usage metrics as covariates.

TABLE 1. DETAILED RESULTS

Values	Effect size	*p* value	Percentile Gain
All Agile Assessment users	0.30	.01	12
Grade 9	0.34	.01	13
Grades 10-12	N/A	1	N/A
Male	0.31	.02	12
Female	0.28	.07	11
White	N/A	.28	N/A
Black	N/A	.32	N/A
Hispanic	0.32	.01	13
Eligible for free- or reduced-price lunch (FRPL)	0.31	.01	12
Not eligible for free- or reduced-price lunch	N/A	.56	N/A
Special Education	0.60	.01	23
Not Special Education	.23	.09	9
English Language Learner (ELL)	N/A	.29	N/A
Not English Language Learner (ELL)	.34	.01	13

Note. Effect sizes and growth percentiles are not given for estimates that do not reach levels of statistical significance.

TABLE 2. BASELINE EQUIVALENCE

	Comparison	*Agile Assessment*	Pooled SD	Diff % SD
Pretest	727	726	118	0.01
% Male	58	56	50	0.04
% FRPL	93	95	32	0.07
% ELL	24	28	44	0.09
% Special Education	10	9	35	0.03
% White	3	4	26	0.07
% Black	16	12	37	0.13
% Hispanic	74	75	47	0.02
% Asian	7	9	30	0.07
Students	146	764		

Reference this report: Empirical Education Inc. (2020). *Effectiveness of Agile Mind's Agile Assessment.* (Empirical Education Rep. No. Empirical_JEA-6059-DDS-M-2019-O.2). San Mateo, CA: Empirical Education Inc. Retrieved June 17, 2020 from https://www.empiricaleducation.com/blog/agile-assessment-impact-formative-testing-student-achievement-algebra/