Progress Learning

McREL Research Summary

The Impact of the Progress Learning Online Assessment, Practice and Instruction programs on Fourth Grade Mathematics and Reading Achievement

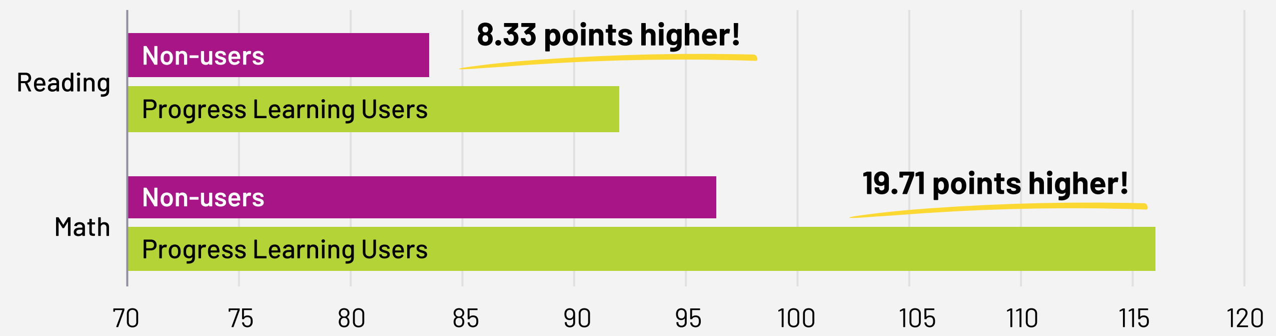
About the Research

McREL International conducted evidence-based research and used a quasi-experimental (matched comparison) design that conforms with the requirements of the Every Student Succeeds Act (ESSA) evidence standards for providing *Tier 2 Moderate Evidence*. The outcomes of interest were grade-level gain (from third to fourth grade) on the mathematics and reading scales of the State of Texas Assessment of Academic Readiness (STAAR) assessment. Achievement gain of Progress Learning user schools was compared to gain of non-user schools from the 2016-17 to 2017-18 school years, controlling for prior achievement and other school-level demographic variables to ensure baseline equivalence of the two groups. Schools in the study utilized Progress Learning and Liftoff programs and answered a minimum number of questions based on recommended usage levels.

Impact Results

McREL's research found that schools using Progress Learning had scores grow at a substantial level when comparing their 2017 state test scores in 3rd grade in comparison with their 4th grade results in 2018 for both math and reading. The results are considered substantively important according to What Works Clearinghouse (2017) requirements. Progress Learning user schools grew 19.71 points higher than non-user schools in math and grew 8.33 points higher than non-user schools in reading.

Growth Comparison for 1st Year Progress Learning Schools compared to non-users



This quick summary provided by Progress Learning, LLC
View the attached McREL Research document for full review.

For questions, pricing, or more information, contact us at: info@progresslearning.com

Progress Learning

The Impact of the Progress Learning on Fourth Grade Mathematics and Reading Achievement

Conducted by:
McREL International
4601 DTC Boulevard, Suite 500
Denver, CO 80237-2596

Contact:
Tedra F. Clark, Ph.D.
P: 303.632.5629
tclark@mcrel.org

McREL INTERNATIONAL

Table of Contents

Study Overview

The purpose of this study was to conduct a rigorous, external evaluation of the impact of the Progress Learning Online Assessment, Practice, and Instruction programs[1] on mathematics and reading achievement. Specifically, the study used a quasi-experimental (matched comparison) design that conforms with the requirements of the Every Student Success Act (ESSA) evidence standards for providing Tier 2 Moderate Evidence. The outcomes of interest were grade-level gain (from third to fourth grade) on the mathematics and reading scales of the State of Texas Assessment of Academic Readiness (STAAR) assessment. Achievement gain of Progress Learning user schools was compared to gain of non-user schools from the 2016-17 to 2017-18 school years, controlling for prior achievement and other school-level demographic variables to ensure baseline equivalence of the two groups.

Progress Learning user schools and non-user comparison schools were selected based on the following criteria:

1. Progress Learning user schools newly adopted the program for the 2017-18 school year, with no exposure to the program prior to that year. Thus, data from 2016-17 were used for matching user schools to non-user schools and to establish baseline equivalence.
2. Progress Learning use schools met cut-off criteria for program usage (average number of questions answered per student) indicative of fidelity of implementation (more information about the cut-off criteria for usage is provided in the Research Design section).
3. Non-user schools were selected using a propensity score matching algorithm that accounted for baseline (2016-17 school year) achievement variables and demographic characteristics to ensure that user and non-user schools were equivalent prior to user-school exposure to Progress Learning.

Results of the study revealed a positive impact of Progress Learning on grade level gain from third to fourth grade on both mathematics and reading subscales of the STAAR compared to non-user matched comparison schools, with statistical significance levels (p-values) of <.001 for mathematics and .068 for reading. The effect sizes of 0.58 and 0.30 for mathematics and reading, respectively, are considered substantively important positive effects, according to What Works Clearinghouse[2] Evidence Standards (WWC, 2017).

Progress Learning Description

The current study addresses the Progress Learning online assessment practice platform with an adaptive intervention component[3] to help struggling learners get up to grade level quickly. The programs provide online assessment practice and individualized study plans for students to make progress toward learning important concepts built to the rigor of state standards. Thousands of questions can be accessed through the online platform or via printable worksheets covering mathematics, reading, writing, and science content areas, in both English and Spanish, and allowing English language learners to switch between the two languages as needed. The programs are designed to be highly engaging to students through games and reward structures that make learning fun and foster student persistence. The programs provide support for key elements of formative assessment practice, such as providing real time feedback to students and allowing students to understand their own progress toward mastery. At the same time, teachers gain formative information from progress reports to identify areas of strength and opportunity for their individual classrooms and students and to identify areas of focus for further instruction.

1 The Progress Learning platform was previously entitled Education Galaxy
2 "The What Works Clearinghouse (WWC) addresses the need for credible, succinct information by identifying existing research on education interventions, assessing the quality of this research, and summarizing and disseminating the evidence from studies that meet WWC standards." (WWC, 2017, pg. 1).
3 Previously entitled Liftoff

Previous research reported by Progress Learning shows that schools that adopted the programs during the 2016-2017 school year in both Texas and Georgia showed greater percentages of students demonstrating proficiency on their respective state assessments than schools that did not adopt the Progress Learning platform (Education Galaxy, 2017ab). These studies provide the basis for the next step in building the rigor of research to support the Progress Learning platform as impactful for improving student achievement.

Research Design: Quasi-Experimental Matched Comparison Study

The current study aimed to answer the following research question:

- Did elementary schools that newly adopted Progress Learning platform for the 2017-2018 school year experience greater achievement gains on the STAAR assessment (mathematics and reading scales) from grade 3 (Spring 2017) to grade 4 (Spring 2018) than schools that did not adopt Progress Learning?

This research question was answered via secondary analysis of the STAAR data publicly available for Progress Learning user schools and non-user schools, disaggregated by grade level. Specifically, McREL conducted an analysis of grade-level performance data from the 2016-17 and 2017-18 school years using a matched comparison quasi-experimental design. This matched comparison design conforms with ESSA standards for Tier 2 moderate evidence because it controls for any bias in impact estimates that may be due to baseline differences between users and non-users related to prior achievement and other demographic factors highly correlated with achievement.

Potential bias resulting from school self-selection as Progress Learning users was controlled with a rigorous matching strategy called Propensity Score Matching (PSM) – a computer-based algorithm that minimizes the overall distance between groups of cases (Rosenbaum & Rubin, 1985). Using this strategy, elementary schools that newly adopted Progress Learning for the 2017-18 school year were matched to other elementary schools throughout the state of Texas that did not adopt the program and were not previous Progress Learning users. The literature on quasi-experimental studies suggests that matching based on pre-treatment measures of the eventual outcome of the study – in this case, academic achievement in reading and mathematics – is a key variable for optimizing matching and controlling bias. Thus, McREL used the prior year's (2016-17) school-level achievement data from grade 3 to match the schools. In addition to prior achievement, McREL used the following demographic variables for matching on the basis of their relationship to student achievement:

- School size
- School locale
- School gender composition (percentage of male versus female students)
- School racial/ethnic composition (percentage of minority students)
- Percentage of students from families of low-socio-economic status as indicated by free- and reduced priced lunch status)
- Percentage of students who are English Language Learners (ELL)
- Percentage of students served by special education (SpED)
- Percentage of students who met or mastered grade level standards at baseline
- Mean grade level student achievement scores from baseline (grade 3 students in the 2016-17 school year)

Sample Selection and Power

Selection of Progress Learning user schools was an important piece of the study design because it was necessary to balance optimal program usage with inclusion of enough schools to have the power necessary to detect statistically significant effects. After consultation with the program developer, it was determined that an average of 400 questions answered per student (in both mathematics and reading) is optimal for a school to achieve implementation fidelity. The number of questions answered per student was calculated by dividing the school's total questions answered by fourth grade students by the number of fourth grade students in the school. This calculation allowed us to weight the number of questions answered by the school size to ensure that schools of all sizes would be included in the study. (Using the total of questions answered per school – without weighting by the number of students per school – would have biased the study toward pre-dominantly including larger schools.)

A preliminary power analysis using Optimal Design software[4] (Spybrook, Bloom, Congdon, Hill, Martinez, & Raudenbush, 2011) indicated that a total of approximately 300 schools (including user and non-user schools) would be needed for a minimum detectable effect size[5] of 0.25 – the effect size cut-off set by the WWC (2017) as indicative of a substantively important positive effect.

Based on an analysis of the average number of questions answered per Progress Learning school, it was confirmed that the 400 questions per student cut-off could be achieved for mathematics and still maintain power (i.e., at least 300 schools included in the analysis) if a 1:5 matching procedure were used (see more about the matching strategy in the Propensity Score Matching Section below and in Appendix A). However, because reading items take longer than mathematics items for students to complete, it was determined that the cut-off for the number of questions answered would need to be lowered in order to maintain adequate power, even though it does not conform with optimal implementation. Thus, the average number of questions per student cut-off for reading was set at 300 items, which lowered both power to detect statistically significant effects and the level of implementation for the reading sample as compared to the mathematics sample.

Propensity Score Matching

Within each dataset (mathematics and reading), matching was done using logistic regression to obtain a propensity score representing the probability that a unit with certain characteristics was assigned to the Progress Learning user group. After propensity scores were estimated, a one-to-five nearest neighbor matching algorithm with a caliper of 0.05[6] and with replacement was used to identify five non-user comparison schools per user school based on the aforementioned list of demographic and achievement variables. (More details on the propensity score matching procedures and its results are presented in Appendix A.) Baseline characteristics of the Progress Learning user and non-user schools for both the final (post-matching) mathematics and reading samples are shown in Table 1. Analysis of variance did not reveal any statistically significant baseline differences between user and non-user schools on any of the variables (all *p*s > .40).

---

4 The following assumptions were made based on the educational literature (Cook, 2005; Hedges & Hedberg, 2007ab): (1) the value of significance level is 0.05; (2) variances explained by school-level variables (e.g., average school-level pretest for 3rd grade, percentage of free and reduced lunch students, percentage of minority students, etc.) is 0.50, and the desired power is 0.80.
5 The minimum detectable effect size (MDES) represents the smallest true effect, in standard deviations of the outcome, that is detectable for a given level of power and statistical significance.
6 Implementing a caliper in the matching process helps to avoid the risk of bad matches (Guo & Fraser, 2010; Parsons, 2001).

Table 1. Comparison of Progress Learning (User) and Comparison (Non-user) School Baseline (Post-matching) Characteristics for final Mathematics and Reading Samples

	Mathematics Sample		Reading Sample	
Characteristic	Progress Learning (User) Schools (N = 57)	Comparison (Non-user) Schools (N = 265)	Progress Learning (User) Schools (N = 31)	Comparison (Non-user) Schools (N = 145)
	Mean (SD)		Mean (SD)	
Mathematics Achievement – Scaled Score (Mean, SD)	1455.77 [b] (41.68)	1456.80 (55.54)	1413.58 (54.51)	1419.39 (54.07)
	Percent		Percent	
Mathematics Achievement – Percent Meeting Proficiency	21.70%	21.46%	15.21%	15.32%
School Size [c]				
Small	17.30%	17.73%	16.13%	14.48%
Medium	75.44%	77.74%	77.42%	76.55%
Large	5.26%	4.53%	6.45%	8.97%
School Locale [d]				
MU	28.07%	28.68%	29.03%	31.03%
MS	17.45%	16.98%	12.90%	16.55%
IT	1.75%	1.51%	6.45%	3.45%
NFG	0.00%	0.00%	0.00%	0.00%
NS	8.77%	5.28%	6.45%	3.45%
OCC	14.04%	15.47%	22.58%	22.76%
OCS	19.30%	20.75%	16.13%	17.93%
RURAL	10.53%	10.53%	10.53%	10.53%
Percentage of male students	51.47%	51.47%	51.55%	52.33%
Percentage of racial/ethnic minority students	73.98%	75.46%	80.04%	82.50%
Percentage of students in the free- or reduced-price meal program (FRPM)	65.95%	65.44%	62.76%	61.20%
Percentage of students who were English Language Learners (ELL)	5.78%	6.80%	5.58%	6.77%
Percentage of students served by special education (SpEd)	9.15%	9.00%	9.25%	8.99%

---

a There were no statistically significant differences between Progress Learning user and non-user schools on any of the baseline (post-matching) school-level variables according to Analysis of Variance (all *ps* > .40).
b Numbers in parentheses are standard deviations.
c For school locale, IT =Independent town; MS = Major Suburban ; MU = Major Urban ; NFG = Non-metropolitan: fast growing ; NS = Non-metropolitan: stable; OCC = Other central city ; OCS = Other central suburban; and RURAL = rural.
d For school size, schools with equal or less than 400 students were categorized as small size schools; schools with student enrollment between 401 and 800 were categorized as medium size schools; schools with equal or greater than 801 schools were categorized as large size schools.

After the matching procedure, all school-level covariates that were used in the matching (demographic and prior achievement variables) were controlled statistically in the analytic models to assess the impact of Progress Learning. Also, because comparison schools were allowed to be matched more than once (matching with replacement), weights were created to apply when analyzing the data.

Data Analysis

Publicly available STAAR data for mathematics and reading were used to calculate the achievement gain outcomes for the study. Specifically, gain scores were calculated by subtracting 2017 third grade average school scale scores from 2018 fourth grade average school scale scores. Two separate single-level multiple linear regression models (one for mathematics and one for reading) were used to assess the impact of Progress Learning on reading and mathematics gains in elementary schools that newly adopted Progress Learning for the 2017-18 school year as compared to matched comparison schools that did not adopt the resource. All variables used in the matching process, including STAAR achievement from the prior school year (grade 3 from the 2016-17 school year) and demographic characteristics of schools were entered in the model control purposes.

Specifically, including school level demographic characteristics (e.g., school size; percent of ELLs; percent minority students; percent of students in families of low socio-economic status, etc.) added an additional level of control for potential differences between user and non-user groups on these factors. Weights from the PSM were also included to account for matching with replacement. The full conditional analytic model used to analyze the impact of Progress Learning on school level achievement gains is specified in Appendix B. Prior to conducting the impact analysis, McREL examined the STAAR mathematics and reading scale scores from the baseline (2017) year to check for baseline equivalence. Specifically, descriptive statistics (unadjusted means and standard deviations) were calculated along with linear regressions and effect sizes (Hedge's *g*) to ensure that the Progress Learning user schools were equivalent to non-user schools on mathematics and reading achievement prior to exposure to Progress Learning. Results from the baseline equivalence and impact analyses are described next.

Results

Baseline Equivalence

The first step in the analysis was to establish baseline equivalence by computing descriptive, regression, and effect size statistics on the grade 3 (2017) mathematics and reading STAAR scale scores. Results are presented in Table 2.

Table 2. Results of baseline equivalence analyses

Content Area	Progress Learning (User) Schools			Comparison (Non-user) Schools			Mean difference	Test Statistic (*t*-value)	Significance Level (*p*-value*)	Effect size (Hedge's *g*)
	Unadjusted Scale Score Mean	SD	N	Unadjusted Scale Score Mean	SD	N				
Math	1455.77	41.68	57	1456.80	55.54	265	-1.03	.131	0.896[a]	-0.02[b]
Reading	1413.58	54.51	31	1419.39	54.07	145	-5.81	-.542	0.589[a]	-0.11[c]

a p-value does not approach statistical significance
b effect sizes < 0.05 satisfy baseline equivalence according to the WWC (2017)
c effect sizes > 0.05 to ≤0.25 satisfy baseline equivalence with statistical adjustment according to the WWC (2017)

For the baseline equivalence analysis on the mathematics subscale, the mean difference was negative 1.03, meaning the Progress Learning user schools scored slightly lower than the non-user schools at baseline (grade 3; 2017). This difference according to the linear regression analysis was not statistically significant ($p = 0.896$) and the effect size of negative 0.02 (calculated using Hedge's g7) satisfies baseline equivalence according to WWC (2017) standards.

For the baseline equivalence analysis on the reading subscale, the mean difference was negative 5.81, meaning the Progress Learning user schools scored slightly lower than non-user schools at baseline (grade 3; 2017). The difference according to the linear regression analysis not statistically significant ($p = 0.589$). The effect size of negative 0.11 falls between negative 0.05 and negative 0.25, which, according to WWC (2017), requires statistical adjustment to satisfy baseline equivalence. It should be noted, however, that the difference between Progress Learning user schools and non-user schools is in the opposite direction of what was hypothesized for the results of the outcome analysis (e.g., non-user schools scored slightly higher than user schools at baseline). Regardless, impact analyses for both mathematics and reading included baseline scale score achievement along with several other demographic variables as covariates in the analytic model, which satisfies the WWC baseline equivalence requirement (see Data Analysis section).

Impact Analysis

Results of the impact analyses on gain outcomes for both mathematics and reading can be seen Table 3.

Table 3. Results of impact analyses

Content Area	Progress Learning (User) Schools			Comparison (Non-user) Schools			Mean difference	Test Statistic (*t*-value)	Significance Level (*p*-value*)	Effect size (Hedge's *g*)
	Unadjusted Scale Score Mean	SD	N	Unadjusted Scale Score Mean[a]	SD	N				
Math	116.20	30.50	57	96.49	35.00	265	19.71	4.18	4.18[b]	0.000[d]
Reading	92.01	25.42	31	83.68	28.16	145	8.33	1.84	1.84[c]	0.068[d]

a Means are adjusted for covariates in the regression model
b *p*-values ≤ 0.05 are considered statistically significant
c *p*-values > .05 and ≤ 0.10 are considered marginally significant
d Effect sizes ≥ .25 are considered substantively important according to the WWC (2017)

For the mathematics outcome, Progress Learning user schools showed an adjusted mean gain of 116.20 scale score points from grade 3 (2017) to grade 4 (2018) as compared to an adjusted mean gain of 96.49 for non-user schools. This adjusted mean difference of 19.71 scale score points was statistically significant ($p < .001$), with an effect size of .58. For the reading outcome, Progress Learning user schools showed an adjusted mean gain of 92.01 scale score points from grade 3 (2017) to grade 4 (2018) as compared to an adjusted mean gain of 83.68 for non-user schools. This adjusted mean difference of 8.33 scale score points was marginally significant ($p < .068$), with an effect size of .30. Although the *p*-value for the reading outcome did not reach the conventionally-used cut-off of ≤. 05, the effect size is considered substantively important even in the absence of statistical significance according to the WWC (2017). As described previously in the Research Design Section, unlike the analysis for mathematics, the analysis for reading was compromised on both implementation fidelity (the user schools did not meet the developer's recommendation of 400 questions answered per student), as well as power to detect statistical significance (the total sample size was less than 300 schools). Even with these limitations, the effect for reading is considered strong, as it indicates a 0.30 standard deviation difference in gain between Progress Learning user schools and non-user schools.

Conclusions and Recommendations

This study was conducted to estimate the impact of the Progress Learning Online Assessment, Practice, and Instruction platform on the outcomes of mathematics and reading gains from grade 3 to grade 4. Results of the study suggest that if schools purchase Progress Learning and implement it under similar conditions as schools included in this study, positive impacts may be found on both mathematics and reading gain outcomes over the course of only one school year of implementation. The impact analyses on achievement gains were conducted in the context of a quasi-experimental (matched comparison) design to establish baseline equivalence between Progress Learning user schools and non-user schools. As a result, the analyses provided unbiased estimates of the impact of Progress Learning on student achievement gains.

An important factor to consider when interpreting these findings is that, while Progress Learning is typically implemented under real-world conditions (schools/districts purchase the intervention and implement it without requirements), only those schools that implemented the program in a way that approached or reached the developer's recommendation were included in the study. Therefore, the results may not generalize to districts/schools that purchase the resource and do not use it at the level recommended by the developer. The importance of implementation fidelity is highlighted by the lower impact on the reading outcome as compared to the mathematics outcome found in this study. Specifically, across the schools considered for inclusion in this study (schools in Texas that newly adopted the resource for the 2017-18 school year), a small number of schools met the recommended cut-off of 400 questions answered (on average) per student for reading[8]. Therefore, in order to include enough schools for the reading outcome analysis, the cut-off for inclusion was reduced to 300 questions answered per student. This, along with the smaller sample size for the reading outcome analysis, provides a possible explanation for the smaller effect size for reading as compared to mathematics, as schools included in the mathematics analysis met the recommended 400 questions per student.

Considerations of implementation fidelity aside, this study provides robust evidence that Progress Learning is a valuable resource for schools and districts wishing to provide their students with engaging standards-aligned assessment practice and instruction opportunities for improving their achievement results. In fact, effect sizes in this study range from almost two-thirds (for reading) to over one-half (for math) of a standard deviation of achievement gain for fourth grade classrooms that used Progress Learning as compared to those that did not. Given these robust findings, Progress Learning is well positioned to expand both the scope and scale of research to support their product. As the user-base for Progress Learning grows larger, recommended next steps for research include: expanding to additional states (beyond Texas) and grade levels (beyond grade 4); exploring gain trends across multiple years of usage (beyond the first year); and conducting a randomized controlled trial where schools are randomly assigned to use the Progress Learning platform to conform with the highest level of ESSA evidence standards (strong evidence as compared to moderate evidence in the current study).

8 One possible reason for lower usage on reading questions is that a reading question generally takes longer to complete than a mathematics question.

References

Cook, T. D. (2005). Emergent principles for the design, implementation, and analysis of cluster-based experiments in social science. *The Annals of American Academy of Political and Social Science*, 599(1), 176-198.

Education Galaxy (2017a). *Improving Student Achievement with Evidence from the State of Texas Assessment of Academic Readiness*. Retrieved from https://educationgalaxy.com/wp-content/uploads/2018/10/STAAR-2017-Comparison-Texas.pdf

Education Galaxy (2017b). *Improving Student Achievement with Evidence from the Georgia Milestones EOG Assessments*. Retrieved from https://educationgalaxy.com/wp-content/uploads/2018/10/Milestones-EOG-2017-Comparison.pdf

Guo, S., & Fraser, M. W. (2010). *Propensity score analysis: statistical methods and applications*. Thousand Oaks, California: Sage

Hedges, L. V. & Hedberg, E. C. (2007a). *Intraclass correlations for planning group-randomized experiments in education*. *Educational Evaluation and Policy Analysis*, 29, 60-87.

Hedges, L. V., & Hedberg, E. C. (2007b). *Intra-class correlations for planning group randomized experiments in rural education*. *Journal of Research in Rural Education*, 22(10), 1-15.

Parsons, L. S. (2001). *Reducing bias in a propensity score matched-pair sample using greedy matching techniques*. Paper presented at the SAS SUGI 26.

Rosenbaum P. R. & Rubin D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39, 33-38.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodolog*y, 2, 169-188.

Spybrook, J., Bloom, H., Congdon, R., Hill, C., Martinez, A., & Raudenbush, S. (2011). *Optimal Design Plus Empirical Evidence: Documentation for the "Optimal Design" Software*. William T. Grant Foundation. Retrieved from http://hlmsoft.net/od/od-manual-20111016-v300.pdf.

What Works Clearinghouse (WWC) (October 2017). Procedures and Standards Handbook Version 4.0. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

About McRel International

McREL International, a private 501(c)(3) nonprofit corporation, was established in 1966 and is headquartered in Denver, Colorado. McREL's primary goal is to make a difference in the quality of education for all learners through excellence in applied research, program evaluation, professional learning, technical assistance, product development, and service to those who are committed to improving lives and the organizational conditions that facilitate success. McREL achieves its objectives by focusing on what matters most to change the odds of success for students and staff, and by collaborating with educators to create better ways to help learners of all ages flourish.

In operation for over 50 years, McREL houses expertise in conducting research and evaluations; developing resources, tools, and standards-based programs; providing technical assistance, professional learning, and leadership development; consulting in system improvement; evaluating policies; and engaging in strategic planning. McREL's understanding of the current issues and challenges facing PreK–12 education is based on over five decades of research, development, and service to clients at the international, national, regional, state, and local levels. McREL's research provides education stakeholders with valuable information and practical tools for research-based, effective approaches to the challenges of education today.

Appendix A:
Propensity Score Matching Procedure and Results

McREL researchers used publicly available Texas STAAR data to identify matched comparisons for Progress Learning schools using propensity score matching (PSM). The original dataset provided included 301 schools that used Progress Learning during the 2018-19 school year, and 4358 potential comparison schools that have never used Progress Learning during or before the 2018-19 school year. Because PSM does not allow missing data, schools with missing data on key covariates were removed from the matching. After data cleaning, 274 Progress Learning schools and 3824 potential comparison schools remained. Across the Progress Learning schools, the level of program usage varied widely. For this study, usage is measured by the average number of questions answered by the targeted students (i.e., 4th graders). After consulting with the program developer, McREL used the following cut offs to identify the Progress Learning user schools for each subject area:

- Reading: schools with an average number of questions answered equal or greater than 300.
- Mathematics: schools with an average number of questions answered equal or greater than 400.

The matching was conducted separately for each subject area. Table A-1 summarizes the final sample size for each dataset before matching.

Table A-1. Study Sample Size by Subject Area

Dataset	Number of Treatment Schools	Number of Potential Comparison Schools
Reading	34	3447
Mathematics	58	3471

For each dataset, matching was done using logistic regression to obtain a propensity score representing the probability that a unit with certain characteristics was assigned to the Progress Learning user group. After propensity scores were estimated, a one-to-five nearest neighbor matching algorithm with a caliper of 0.059 and with replacement was used to identify five comparison schools per user school based on a list of demographic and achievement characteristics (i.e., covariates) that were found to be associated with the outcomes of interest. Table A-2 shows the list of covariates that were included in the matching, including school-level characteristics and grade-level specific student demographic characteristics and achievement score at baseline.

9 Implementing a caliper in the matching process helps to avoid the risk of poor matches (Guo & Fraser, 2010; Parsons, 2001).

Table A-2. Covariates included in propensity score matching and subsequent impact analyses

Covariate	Definition	Variables included in the matching
School size	Schools with equal or less than 400 students were categorized as small size schools; schools with student enrollment between 401 and 800 were categorized as medium size schools; schools with equal or greater than 801 schools were categorized as large size schools. Two dummy variables were created and used in the matching with small schools serving as the reference group.	Medium; Large
School locale	Texas Education Agency classified schools into nine categories: (1) Major urban (MU), (2) major suburban (MS), (3) other central city (OCC), (4) other central city suburban (OCS), (5) independent town (IT), (6) non-metropolitan: fast growing (NFG), (7) non-metropolitan: stable (NS), (8) rural (RURAL), and (9) charter school districts (CSD). Eight dummy variables were created and used in the matching with the charter school districts serving as the reference group.	MU, MS, OCC, OCS, IT, NFG, NS, RURAL
Percentage of male students	Of those who took the STAAR reading or mathematics test in the 2017-18 school year, the percentage of fourth grade students who were male.	Male
Percentage of racial/ethnic minority students	Of those who took the STAAR reading or mathematics test in the 2017-18 school year, the percentage of fourth grade students who are from racial/ethnic minority groups.	Minority
Percentage of students in the free or reduced-lunch meal program (FRPM)	Of those who took the STAAR reading or mathematics test in the 2017-18 school year, the percentage of fourth grade students in the FRPM program.	FRPM
Percentage of students who were English Language Learners (ELL)	Of those who took the STAAR reading or mathematics test in the 2017-18 school year, percentage of fourth grade students who were ELL.	ELL
Percentage of students served by special education (SpEd)	Of those who took the STAAR reading or mathematics test in the 2017-18 school year, the percentage of fourth grade students who were served by SpEd.	SpEd
Percentage of students who meet or master grade level standards	Percentage of students who met or mastered grade level standards on the STAAR reading or mathematics test in the 2016-17 school year (school-level aggregated).	PctProficient0
School-level student achievement score	School mean of students' STAAR reading or mathematics test scores in the 2016-17 school year.	STAAR_R0 (Reading) STAAR_M0 (mathematics)

After the matching process was complete, balance diagnostics were conducted to check the quality of the matches. It was expected that the selected comparison (non-user) group would be similar to the user group on all covariates being used for the PSM process (Rubin, 2001). As shown in Figure B1, an examination of the distribution of propensity scores was first conducted to assess common support via a graphic diagnostic; then, three numerical balance measures were used to check covariate balances (Rubin, 2001):

- The ratio of the variances of the propensity scores in the two groups must be close to 1.0. Rubin (2001) suggests that the variance ratios should be between 0.5 and 2.0.
- The difference in the means of the propensity scores in the two groups being compared must be small. Rubin (2001) suggests that the standardized differences of means should be less than 0.25.
- For the percent of balance improvement[10], the larger the percent, the better the PSM results.

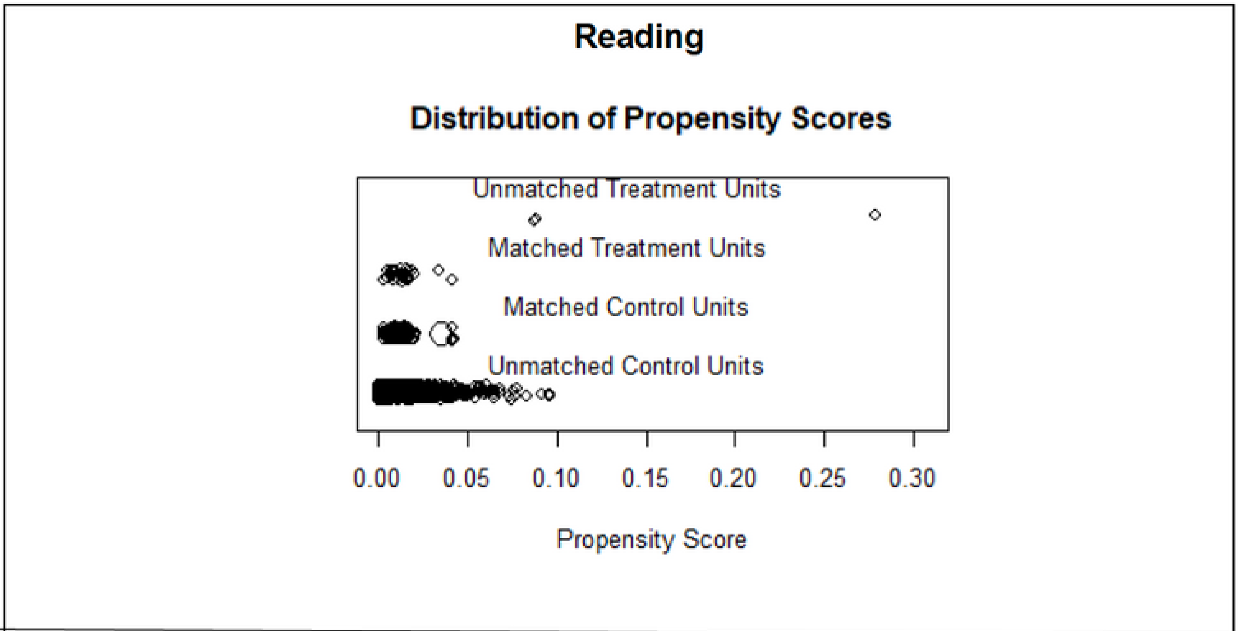
For the mathematics dataset, the result of PSM identified 265 matched comparisons for 57 treatment schools[11]. For the reading dataset, the result of PSM identified 145 matched comparisons for 31 treatment schools[12]. These were the final samples included in the baseline equivalence and impact analyses. Because comparison schools were allowed to be matched more than once, weights were applied when analyzing the data.

A visual examination of Figure A-1 suggests that the selected comparison schools and treatment schools have similar distributions of propensity scores in both mathematics and reading datasets. As shown in Tables A-3 and A-4, the ratio of the variances of the propensity scores equals 1.00, which is within the range suggested by Rubin (2001). The analyses of standard mean differences suggest that the matching procedures have significantly minimized the group mean differences between the treatment and comparison schools across all four datasets. Most importantly, after the PSM process, all covariates had a standardized mean difference smaller than 0.25, as suggested by Rubin (2001). Results of the percent of balance improvement suggests that PSM, overall, improves the balance between the treatment and comparison schools, especially for the covariates that had larger standardized mean difference before matching. For the covariates that had small a standardized mean difference before matching, the percent of balance improvement seem to be smaller or in some cases had negative values. This is expected because PSM decreases the standardized mean difference between user and comparison schools; hence, for the covariates that already have small standardized mean differences before the matching, the change in balance is likely to be small.

---

10 The percent improvement in balance is defined as 100*((|a| - |b|) / |a|), where a is the balance before and b is the balance after matching.
11 One school in the mathematics dataset was removed from the matching because there were no good matches for the school.
12 Three schools in the reading dataset were removed from the matching because there were no good matches for them.

Figure A-1. Jitter plots of the distribution of propensity scores by matching groups



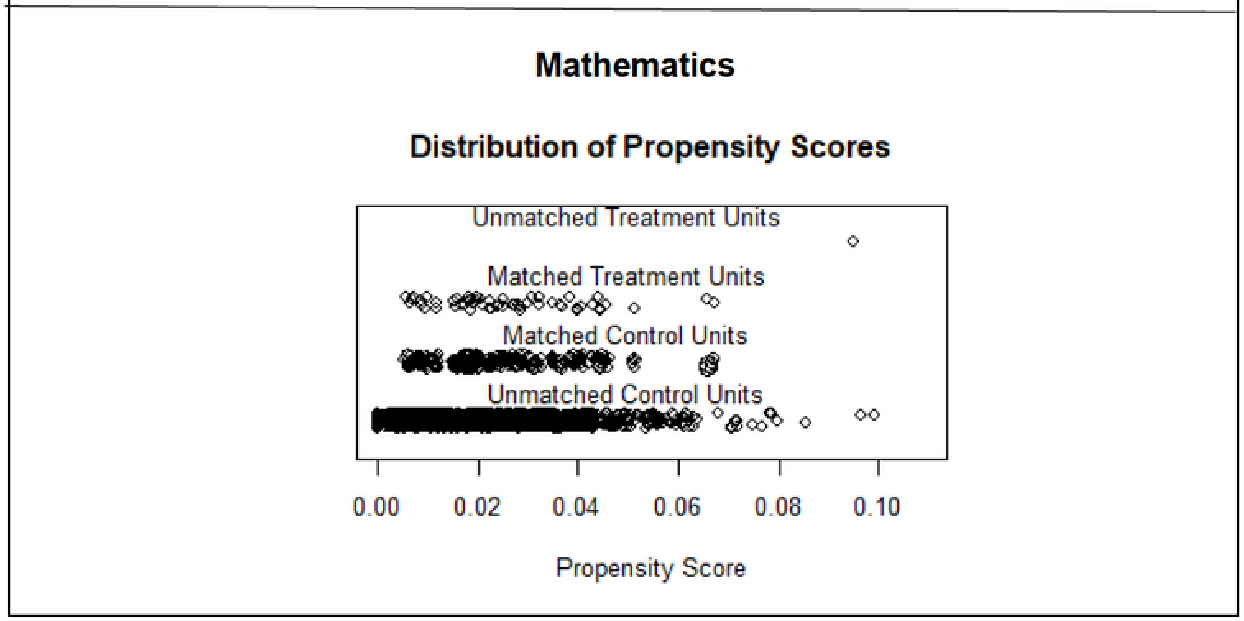


Table A-3. Balance Diagnosis Before and After the PSM Process: 4th Grade Reading

Variables	Treatment				Comparison				Balance Diagnosis			
	Before		After		Before		After			Standard Mean Differences		
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	Variance Ratio	Before	After	% Balance Improvement
Propensity Score	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	1.00	0.33	0.00	99.77
Medium	0.74	0.45	0.77	0.43	0.70	0.46	0.78	0.42		0.08	0.01	81.85
Large	0.06	0.24	0.06	0.25	0.14	0.34	0.08	0.28		0.32	0.08	74.85
IT	0.06	0.24	0.06	0.25	0.04	0.20	0.05	0.21		0.08	0.08	-6.30
MS	0.12	0.33	0.13	0.34	0.30	0.46	0.16	0.37		0.56	0.10	82.28
MU	0.26	0.45	0.29	0.46	0.21	0.41	0.30	0.46		0.56	0.10	82.28
NFG	0.00	0.00	0.00	0.00	0.01	0.08	0.00	0.00		0.00	0.00	100.00
NS	0.15	0.36	0.06	0.25	0.04	0.19	0.06	0.23		0.31	0.02	94.12
OCC	0.21	0.41	0.23	0.43	0.16	0.37	0.21	0.41		0.11	0.03	71.79
OCS	0.15	0.36	0.16	0.37	0.13	0.33	0.17	0.38		0.06	0.04	38.15
RURAL	0.06	0.24	0.06	0.25	0.06	0.23	0.05	0.21		0.00	0.08	-2313.14
Male	51.46	6.27	51.55	6.46	51.03	6.02	52.25	5.57		0.07	0.11	63.60
Minority	80.97	22.97	80.04	23.78	73.32	25.52	82.39	20.50		0.33	0.10	69.33
FRPM	64.04	22.87	80.04	23.51	53.06	28.09	60.95	24.83		0.48	0.08	83.56
ELL	9.24	15.43	5.58	7.72	6.96	9.46	6.85	10.52		0.15	0.08	43.88
SpEd	8.99	3.45	9.25	3.33	8.99	4.47	8.86	4.52		0.00	0.11	-9446.4421
Pct Proficiency	37.06	13.82	38.03	14.03	43.53	16.70	39.44	13.49		0.47	0.10	78.25
STAAR_R0	1410.71	53.71	1413.58	54.51	1435.30	63.05	1420.33	53.05		0.46	0.13	72.53

Note. See Table A-2 for variable definitions.

Table A-4. Balance Diagnosis Before and After the PSM Process: 4th Grade Mathematics

Variables	Treatment				Comparison				Balance Diagnosis			
	Before		After		Before		After			Standard Mean Differences		
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	Variance Ratio	Before	After	% Balance Improvement
Propensity Score	0.03	0.01	0.03	0.01	0.02	0.01	0.03	0.01	1.00	0.65	0.00	98.84
Medium	0.74	0.44	0.75	0.43	0.70	0.46	0.78	0.42		0.09	0.06	38.54
Large	0.05	0.22	0.05	0.23	0.13	0.34	0.05	0.21		0.37	0.03	91.56
IT	0.02	0.13	0.02	0.13	0.04	0.20	0.02	0.12		0.18	0.03	84.81
MS	0.17	0.38	0.18	0.38	0.30	0.46	0.17	0.38		0.33	0.01	97.20
MU	0.28	0.45	0.28	0.45	0.21	0.41	0.29	0.45		0.14	0.02	83.87
NFG	0.00	0.00	0.00	0.00	0.01	0.08	0.00	0.00		0.00	0.00	100.00
NS	0.10	0.31	0.09	0.29	0.04	0.19	0.05	0.22		0.22	0.05	75.30
OCC	0.14	0.35	0.23	0.35	0.16	0.37	0.15	0.36		0.07	0.04	39.29
OCS	0.19	0.40	0.16	0.40	0.13	0.33	0.21	0.41		0.16	0.03	83.64
RURAL	0.10	0.31	0.06	0.31	0.06	0.23	0.11	0.32		0.15	0.02	79.58
Male	51.54	6.05	51.55	6.08	50.93	5.86	51.45	5.58		0.10	0.03	82.78
Minority	74.43	28.61	80.04	28.65	73.79	25.52	75.46	25.92		0.02	0.03	-56.39
FRPM	66.18	19.62	80.04	19.72	53.48	28.35	65.44	23.82		0.65	0.02	96.35
ELL	7.01	12.44	5.58	8.29	6.72	9.22	6.80	8.67		0.02	0.09	-276.30
SpEd	0.09	0.04	9.25	3.93	0.09	0.04	9.00	3.99		0.05	0.08	-60.65
Pct Proficiency	42.33	12.46	38.03	12.44	46.53	16.90	42.92	15.12		0.34	0.02	93.16
STAAR_R0	1454.62	42.33	1413.58	41.68	1469.09	61.55	1456.80	55.54		0.34	0.02	92.56

Note. See Table A-2 for variable definitions.

Appendix B: Analytic Model for Impact Analyses

The full conditional analytic model used to analyze the impact of Progress Learning on school level achievement gains for both mathematics and reading is specified as:

$$Yi = \pi_0 + \pi_1 (TREAT)_i + \left(\sum_{S>0} \pi_{si} X_{si}\right) + \varepsilon_i$$

where:

Y_i = the gain outcome for school i,

π_0 = the regression-adjusted mean value of gain outcome for school i,

π_{1i} = is the adjusted mean difference in the school gain outcome between Progress Learning user schools and non-user schools,

TREAT = an indicator variable for the intervention coded as 1 for Progress Learning users schools and 0 for non-user schools,

π_{si} = the value of the coefficient on the sth school-level covariate (baseline achievement, percent of student qualifying for free-reduced priced lunch, percent minority students etc.),

X_{si} = the value of the sth school-level covariate for school i, and

ε_i = the residual error for school i which is assumed to be independently and identically distributed.