Applying a Partially Nested Regression Model to Evaluate the Impact of ASSISTments on Middle School Students' Math Achievement

Chun-Wei (Kevin) Huang,
Mingyu Feng, Linlin Li

September 2025

EIR Award #411B190024

WestEd.org

Suggested citation: Huang, C.-W., Feng, M., & Li, L. (2025). *Applying a partially nested regression model to evaluate the impact of ASSISTments on middle school students' math achievement*. WestEd.

Table of Contents

Introduction

ASSISTments is an evidence-backed educational technology program designed to enhance math instruction and student learning (Feng, Brezack, et al., 2025; Roschelle et al., 2016). In the 2022–23 and 2023–24 academic years, WestEd conducted a national study to evaluate the effectiveness of a large-scale ASSISTments intervention augmented by a virtual professional learning community (vPLC). Worcester Polytechnic Institute (WPI) collaborated with Lesley University, supported by a grant from the U.S. Department of Education's Education Innovation and Research (EIR) program, to develop the intervention that aimed to enhance and scale the use of ASSISTments in 6th through 8th grades. This intervention, vPLC-augmented ASSISTments, combines WPI's ASSISTments online math platform (Heffernan & Heffernan, 2014) with a vPLC designed to enhance the scalability of the program.

This report provides a technical supplement to the final study report (Feng, Li, et al., 2025) on the evaluation of the vPLC-augmented ASSISTments intervention. This technical report aims to demonstrate in more detail how the research team used an analytical approach based on a partially nested regression model (Lohr et al., 2014) to assess whether the intervention improved middle school students' math achievement overall and for rural students. In particular, this technical report focuses on the following two research questions from the study:

- What is the effect of vPLC-augmented ASSISTments on the math achievement of middle school students (grades 6–8) compared with the math achievement of the middle school students in the comparison group?
- Do the effects of vPLC-augmented ASSISTments vary for students with different prior achievement and for students with other policy-relevant characteristics (such as rural versus nonrural)?

Setting and Participants

The study was conducted in two cohorts (2022–23 and 2023–24 academic years) with a diverse sample of 36 schools across 24 states in the United States. The final treatment group involved 59 middle school teachers and 2,855 students in 6th through 8th grades (74% were in 7th grade). More than half of the students (55%) were eligible for free or reduced-price lunch. About half of the participants were in rural areas.

Intervention

Prior studies have shown that ASSISTments improves student learning when paired with in-person coaching (Feng, Brezack, et al., 2025; Roschelle et al., 2016). As part of an EIR grant, the program now includes a vPLC to support teacher training and peer collaboration—reducing travel and scheduling barriers, especially for rural districts.

Teachers participated in eight vPLC sessions focused on using ASSISTments reports to differentiate instruction, establish routines, and discuss student data. During implementation, teachers followed a four-step formative process (Heritage & Popham, 2013) as part of implementing ASSISTments: (a) create standards-aligned math assignments, (b) assist students through immediate feedback, (c) assess class performance, and (d) analyze answers together to address common errors and adjust instruction.

Research Design

The study used a quasi-experimental design (QED) with a matched sample. The study measured student math achievement with the online MAP Growth math assessment provided by NWEA. MAP assessments have demonstrated desired psychometric properties (NWEA, 2019). The test was administered twice yearly, first in the fall as a baseline and then in the spring as an outcome.

The study leverages the virtual comparison group (VCG) approach developed by NWEA as a proxy for an experimental control group (Ma & Cronin, 2009). Treatment students exposed to ASSISTments were matched and compared with the VCG from the NWEA's database based on their matching algorithm. The baseline scores, along with other information from the treatment group, were used to construct the VCG.

Construction of the Virtual Comparison Group

The VCG approach involves creating an equivalent virtual comparison for each student in the treatment group. Based on NWEA's matching algorithm, each treatment student was matched with up to 51 real students from NWEA's extensive national MAP test database based on school locale, school-level percentage of students eligible for free or reduced-price lunch, test-taking dates, student grade level, and pretest scores. After matching, the aggregated values (such as the average test scores) across that group of up to 51 students were then treated as the performance of a single virtual comparison student.

When the VCG was constructed, each virtual comparison student was assigned a matching level indicating the degree to which the matching criteria were from the most stringent set of values. This "comparison criteria level" ranged from 0 to 8, with smaller values indicating a closer match. Table 1 summarizes the number and percentage of virtual comparison students at each level in the VCG. Overall, approximately 89 percent of virtual comparison students were within the comparison criteria level of 0 to 2, meaning that the baseline scores of the matched students differed from those of the treatment students by no more than 2 points. These matched students also came from the schools with a similar percentage of students eligible for free or reduced-price lunch (within 5%), came from the same school locale, and had test administration dates (pretest or posttest) within 7 days of the corresponding treatment student.

Table 1. The Number of Virtual Comparison Students by Comparison Criteria Level

Comparison criteria level	Baseline score	School % FRL	Days between tests	School locale code	Number of virtual comparison students	Percentage	Cumulative percentage
0	"+/-0"	"+/-5"	"+/-7"	Exact Match	1,578	55.27	55.27
1	"+/-1"	"+/-5"	"+/-7"	Exact Match	728	25.50	80.77
2	"+/-2"	"+/-5"	"+/-7"	Exact Match	235	8.23	89.00
3	"+/-2"	"+/-5"	"+/-10"	Exact Match	67	2.35	91.35
4	"+/-2"	"+/-10"	"+/-10"	Exact Match	129	4.52	95.87
5	"+/-2"	"+/-15"	"+/-10"	Exact Match	39	1.37	97.23
6	"+/-3"	"+/-15"	"+/-10"	Exact Match	28	0.98	98.21
7	"+/-3"	"+/-15"	"+/-18"	Exact Match	42	1.47	99.68
8	"+/-4"	"+/-15"	"+/-18"	Exact Match	9	0.32	100.00
Total					**2,855**	**100**	

***Note*. FRL refers to free or reduced-price lunch eligibility. The Percentage column may not total exactly 100.00 due to rounding.**

Analytic Approach

To estimate the treatment impact for answering the research question about the effect of implementation on math achievement, a two-level hierarchical linear model (students nested within teachers) would have been used. However, with the current design, matching was done at the student level, which complicated the impact analysis. The intervention was still implemented at the teacher level, but there was no teacher level for the VCG. Therefore, a hierarchical linear model based on a partially nested randomized controlled trial (PN RCT) design (Lohr et al., 2014) was adopted. This type of model was initially designed for RCTs in which treatment students receive intervention services in groups (teachers or classrooms) but control students are not nested in any groups. The PN RCT design can also be applied to a QED, especially a matched comparison group design (Lohr et al., 2014) as was done in this study.

Since only treatment students were nested in teachers (referred to as intervention clusters, or ICs), the PN model allows additional variation at the cluster or teacher level for the treatment group. In other words, instead of treating the coefficient associated with the impact estimate (typically represented by the treatment indicator, 0 or 1) as fixed, it is treated as a random effect with an error term in the model. The model takes the following form:

Level 1 Model: $Y = \beta_0 + \beta_1 \text{(treat)} + r$

Level 2 Model:

$$\beta_0 = \gamma_{00}$$

$$\beta_1 = \gamma_{10} + \mu_1$$

Mixed Model: $Y = \gamma_{00} + \gamma_{10} \text{(treat)} + \mu_1 \text{(treat)} + r$

In this model, *treat* is coded 1 = treatment and 0 = control, γ_{00} is the expected value of the control condition, γ_{10} is the expected value of the treatment effect, μ_1 is the error term associated with treatment clusters, and *r* is the residual error. In this study, the pretest scores were also included as a covariate to adjust for the baseline difference in the model.

To fit the PN model in this study's sample, the IC values for the VCG needed to be created as if each control were a singleton cluster, rather than IC being treated as missing for all controls. For this purpose, researchers generated a markup teacher identification number for each control.

The following codes in Stata (StataCorp, 2025) were then used to estimate the treatment impact for the research question about the impact on math achievement, which aligned with the model specification discussed above:

```
mixed score_2 treat score_1 || tea_id: treat if optout~ = 1, noconstant residuals
(independent, by(treat)) reml dfmethod (sat)
```

where

- *mixed* is the Stata command for fitting linear mixed-effects models;
- *score_2* is the posttest score;
- *treat* is the treatment indicator (treat = 1 for the treatment group versus treat = 0 for the VCG);
- *tea_id* is the teacher identification number as IC values;
- *tea_id: treat* is a random effect for the *treat* variable at Level 2 (i.e., a random slope as shown in β_1 in the equation);
- *optout* excludes any students who opted out of the study (if optout = 1);
- *noconstant* suppresses the random intercept at Level 2, as shown in β_0 in the equation;
- *residuals(independent, by(treat))* models the residual variance structure, assuming the residuals are independent and differ for the treatment group and VCG (as recommended by Lohr et al., 2014);
- *reml* is an estimation method based on restricted maximum likelihood (REML); and
- *dfmethod(sat)* estimates the degrees of freedom (*df*) based on generalized Satterthwaite approximation for heterogeneous residual variances (as recommended by Lohr et al., 2014; to obtain the Satterthwaite *df*, the REML is required in Stata).

For the research question on whether the effects of vPLC-augmented ASSISTments vary for particular student subgroups, the same PN model was used, but the analyses were based on two distinct subsamples: the students with lower levels of performance and students in rural schools. Students with lower levels of performance were identified based on the treatment students' percentile rank on the pretest, with those scoring below the 50th percentile classified as low performing. Students' rural status was determined using the school locale code established by the National Center for Education Statistics (n.d.).

Results

For the main impact research question about the vPLC-augmented ASSISTments intervention's impact on student math achievement, the treatment impact was estimated to be 0.67 (the coefficient associated with "treat"), indicating that, on average, the treatment group scored 0.67 points higher than its VCG. However, this difference is not statistically significant at the .05 level, with the *p*-value equal to 0.172 (Table 2).

Table 2a. Model Estimation for the Main Research Question on Impact: Fixed Effect

Fixed effect	Coefficient	SE	*t* ratio	*p*-value
treat	0.67	0.488	1.38	0.172

Table 2b. Model Estimation for the Main Research Question on Impact: Random Effect

Random effect	Variance estimate	SE	95% CI (lower)	95% CI (upper)
Level 2, treat	11.86	2.626	7.689	18.307
Level 1 residual, VCG	2.88	0.076	2.738	3.038
Level 1 residual, treat	71.48	1.913	67.828	75.330

***Note*. SE refers to standard error; 95% CI (lower) refers to the lower bound of the 95 percent confidence interval; 95% CI (upper) refers to the upper bound of the 95 percent confidence interval.**

The second part of the output ("Random Effect") reported the variance estimates at each level. The estimated IC-level variance, var(treat) or µ1 in the equation, is 11.86. The residual variance at Level 1 (within-cluster) for the treatment group is 71.48 and for the VCG is 2.88.

For the research question on whether the effects of the implementation vary for particular student subgroups, similar output tables were produced (but are not presented here). The results indicate that treatment students who scored below the 50th percentile on the pretest showed a significant improvement compared to the VCG (214.38 versus 213.18, difference = 1.20, p = 0.020). Similar results were found when the analysis considered only students in rural areas. Table 3 summarizes the impact findings for both research questions.

Table 3. Impact of ASSISTments With vPLC, Overall and by Subgroup

	Tx (adjusted mean)	VCG (adjusted mean)	Diff	SE	*p*	ES	Tx (*n*)	Tx (SD)	VCG (*n*)	VCG (SD)
Posttest (all sample)	223.95	223.28	0.67	0.488	0.172	0.04	2,855	17.277	2,855	15.048
Posttest (low performing)	214.38	213.18	1.20	0.500	0.020	0.10	1,626	13.230	1,626	9.670
Posttest (rural only)	225.53	224.88	0.65	0.667	0.339	0.04	1423	16.438	1423	14.133
Posttest (low performing in rural)	215.51	214.26	1.26	0.607	0.049	0.11	733	12.964	733	9.217

***Note*. Tx refers to the treatment group; Diff represents the difference between Tx and VCG; SE refers to standard error; ES refers to the effect size and was calculated based on the pooled standard deviation (SD).**

Conclusion

This paper discusses in detail how the VCG was constructed based on the NWEA's matching algorithm. Overall, approximately 89 percent of virtual comparison students were within a comparison criteria level of 0 to 2. The baseline equivalence test on the pretest (presented in the main study report) revealed that the average mean of VCGs was very close to that of the treatment group, with an effect size of -0.03. This indicates that the matching algorithm performed as expected. Additionally, the paper demonstrates the use of a PN regression model to analyze data in which the clusters existed only in the treatment group and not in the control or comparison group. The paper discusses how the model was set up to allow variation only between teachers in the treatment group (a random slope, but suppressing a random intercept at Level 2) and at Level 1 to allow an independent residual variance estimate separated by the treatment condition, following the discussions and recommendations by Lohr and colleagues (2014). Based on the impact estimates, it appears that ASSISTments with vPLC benefited the students with lower levels of performance in particular.

References

Feng, M., Brezack, N., Huang, C., & Collins, K. (2025). Long-term effects of an online math tool on U.S. adolescents' achievement. *British Journal of Educational Technology*. http://doi.org/10.1111/bjet.13579

Feng, M., Li, L., Brezack, N., Huang, C.-W., Luttgen, K., Grady, J., Yang, K., Lavine, A., Collins, K., & Schneider, M. (2025). *Scaling teachers' professional development for ASSISTments: Evaluation final report*. WestEd.

Heffernan, N. T., & Heffernan, C. L. (2014). The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, *24*(4), 470–497.

Heritage, M., & Popham, W. J. (2013). *Formative assessment in practice: A process of inquiry and action*. Harvard Education Press.

Lohr, S., Schochet, P. Z., & Sanders, E. (2014). *Partially nested randomized controlled trials in education research: A guide to design and analysis* (NCER 2014-2000). National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/use-work/resource-library/report/research-report/partially-nested-randomized-controlled-trials-education-research-guide-design-and-analysis

Ma, L., & Cronin, J. (2009). *Evaluating the effect of random selection on virtual comparison group creation*. Northwest Evaluation Association. https://eric.ed.gov/?id=ED521965

National Center for Education Statistics. (n.d.). *Elementary/Secondary Information System (ElSi)* (Public Elementary/Secondary School Universe Survey, 2022-23 v.1a; Public Elementary/Secondary School Universe Survey Geographic Data (EDGE), 2022-23 v.1a.) [Data set]. U.S. Department of Education. https://nces.ed.gov/ccd/elsi/

NWEA. (2019). *MAP® Growth™ technical report: March 2019*. https://www.nwea.org/uploads/2021/11/MAP-Growth-Technical-Report-2019_NWEA.pdf

Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA Open*, *2*(4). https://doi.org/10.1177%2F2332858416673968

StataCorp. (2025). *Stata Statistical Software: Release 19.5* [Computer software]. StataCorp LLC.