Pre-analysis Plan for Extended ASSISTments Replication Study

July 15, 2020

The study is an extension and expansion of the ASSISTments replication study funded by the Institute of Education Sciences (IES) (R305A170641) that is currently being conducted in North Carolina. The purpose of this extended study is to power up the replication study with a larger sample, and to measure the long-term effects of ASSISTments on student performance one year after the intervention, when the students are in Grade 8. The study design and analysis plan of the original IES-funded replication study was posted on Registry of Efficacy and Effectiveness Studies (REES) in November 2018.[1]

Study Design

The study uses a school-level, delayed-treatment, clustered randomized experimental design. Schools are randomly assigned to use the ASSISTments platform to support math homework for their Grade 7 students or to a control condition. In the control condition, teachers will continue their current homework practices, including any use of online tools, but will not have access to the ASSISTments platform. The intervention will be implemented by all Grade 7 teachers in treatment schools over 2 consecutive years.

We will be measuring the immediate impact for students in teachers' second year of experience with the system in Grade 7 (after teachers have a warm-up year with a different cohort of students). This group of Grade 7 students will comprise the analytic sample for the main confirmatory research questions. The students will maintain their conditions and be followed longitudinally for another year to Grade 8 when their Grade 8 performance (long term impact) is measured at the end of Grade 8. During the follow-up year in Grade 8, no interventions will be provided to Grade 8 teachers or students.

Sample and setting. The study will be conducted in approximately 63 public schools in the state of North Carolina and surrounding states, each with 2 teachers teaching Grade 7 mathematics on average. (See Power Analysis section for justification). We assume that on average, at least 50 students would be enrolled in each teacher' regular education Grade 7 classrooms. Based on these assumptions, the sample will include approximately 6,300 students.

The North Carolina schools (N=63) were recruited from fall 2017 to spring 2019, and randomly assigned to the treatment or control conditions. Grade 7 teachers in the treatment schools began using ASSISTments with their students during the 2018–19 school year (practical year) and the 2019–20 school year (experimental year). In the experimental years, students' Grade 7 outcome will be measured for school year 2019-20 and students' Grade 8 outcome will be measured for the school year afterwards, i.e., 2020-21.

Research Questions

The study addresses two confirmatory research questions:

1. What is the impact of ASSISTments on student math outcomes at the end of Grade 8?

[1] The original pre-analysis plan is available at https://sreereg.icpsr.umich.edu/framework/pdf/index.php?id=2064 (accessed October 22, 2019).

2. What is the impact of ASSISTments on student math outcomes for the subgroup of students with lower levels of prior math achievement – defined as those scoring below the median on baseline math achievement tests at the end of Grade 8?

The exploratory research questions include the following:

1. Do the effects of ASSISTments vary for students of different socio-economic status, race/ethnicity, or with other policy-relevant student characteristics?
2. Do participating schools implement ASSISTments as intended by the developer? How much usage occurs? To what extent is each ASSISTments feature used?
3. What are the factors that hinder or facilitate implementation?

Data Sources

To estimate the impact of the ASSISTments intervention on student achievement for the confirmatory research questions, we will collect individual student and school-level demographic and achievement data for three school years, as detailed in the following table.

2018-2019	2019-20	2020-21
Grade 6 (baseline)	Grade 7 (Immediate, missing outcome due to COVID-19)	Grade 8 (Primary)

Outcomes. The primary outcomes will be measured by state standardized Grade 8 mathematics assessments. Students' academic performance in 6th grade will serve as student-level covariates in the 2-level hierarchical linear analytic model (HLM) (as discussed below). Among all the performance measures reported by the state assessments, we plan to use the scaled scores as outcomes.

Covariates. Student level covariates include student achievement pre-test and student demographic data, specifically: race/ethnicity, gender, eligibility for free or reduced lunch (an indicator of income status). All student-level variables will be aggregated at the school-level in order to serve as additional school-level covariates. For example, school means on previous mathematics testing may reflect enrollment in an advanced, standard, or more remedial mathematics course in 7th grade. School-level covariates will be 7th grade enrollment size and school Title 1 eligibility.

Implementation data. ASSISTments backend system log data will be collected to monitor the use of the system by students and teachers during the study and to gauge the overall fidelity of implementation. Surveys and interviews will be administered to teachers in both conditions to understand the implementation status and challenges as well as help establish the contrast between conditions. Classroom observations will be conducted to see how teachers are using ASSISTments for math work review with students.

Data Analysis Plan

Confirmatory analysis. An analysis of the efficacy of the use of the ASSISTments for improving student mathematics achievement will be performed using hierarchical linear regression models to account for the effect of clustering on the variance structure of the data. To address **Confirmatory Questions**, the efficacy of the system will be analyzed at the end of Grade 8 **(Confirmatory Question 1)** of the focal analytic sample that is the group of students who were in Grade 7 during the experiment year.

We will model the Grade 8 state test scores (the primary, long term outcome) for students in treatment and control schools, controlling for the students' Grade 6 state test score and other student-, and school-level covariates as described above under "Covariates." The hierarchical linear model for the analysis of student achievement is illustrated below.

$$Math_{ij} = a_0 + b_1\text{PreMath}_{ij} + b_2\text{Tx}_j + \sum b_j\text{S}_j + \sum b_i\text{I}_{ij} + \mu_j + \varepsilon_{ij} \qquad \textbf{[1]}$$

where subscripts *i*, and *j* denote student and school, respectively; *Math* represents student achievement in math; *PreMath* represents the baseline measure of math performance; *Tx* is a dichotomous variable indicating student enrollment in a school that has been assigned to the treatment condition; and *I* and *S* are vectors of student-level covariates, and school-level covariates (as described above), respectively, measured prior to exposure to the intervention. Lastly, μ_j and ε_{ij} represent the random effect of school and student, respectively. In this model, the causal impact of ASSISTments is represented by b_2.

To examine whether the intervention has differential impact on low-achieving treatment students relative to controls (**Confirmatory Question 2**), as defined as students scoring below the median on the pre-test, we will augment the above analytical model with a cross-level interaction term of the school-level treatment variable and whether the student's Grade 6 state test was below the group median. Interpretation of any comparison between low- and high-achieving students will be careful because of possible regression-to-the-mean effects.

Exploratory analysis. To address **Exploratory Question 1**, we will also conduct a similar moderator analysis examining whether the intervention had differential impact on minority (e.g., Hispanic, Black) versus nonminority (White) students. Similar models will be used to estimate the effects of the intervention on the students with other policy-relevant background variables such as FRPL, IEP and ELL status. In these hierarchical models, the school characteristic will be interacted with the treatment status variable to estimate the differential effect of ASSISTments for schools with certain qualities such as those serving high-poverty communities (e.g., greater than 50% FRPL status). We plan to address **Exploratory Questions 2** based on the ASSISTments backend data and calculate the implementation metrics at the individual level (i.e., teacher and student), focusing on the amount of assignments made, the frequency and dosage of usage, and to what extent different ASSISTments features are used. **Exploratory Question 3** will be addressed using qualitative research data (e.g. surveys, interviews, and observations) that the research team will collect.

General analysis issues. The primary study analyses will be conducted on *an intent-to-treat (ITT)* basis. When available, state test outcome data will be collected and analyzed for *all* students who are in the sample schools during the focal 7th grade year, even if their math teacher does not implement ASSISTments or opts out of other researcher-administered data collection activities (i.e., classroom observations and/or surveys). Analysis will be conducted to establish the baseline equivalence of groups (i.e., students, and schools) in various demographic and prior year mathematics performance variables.

The study will monitor attrition at the school- and student-level in both conditions. We will track the out-migration and in-migration of the focal students in each school during their 7th and 8th grade years based on school attendance data which is reported in the Fall and at the end of each school year. Students who switch between treatment/control schools after Fall enrollment report date of the 7th grade year (typically two to four weeks into the school year) will be included in their original group assignment for intent-to-treat estimate. Students who join or leave the treatment/control schools after the Fall enrollment report date of the 7th grade year will be excluded from the analytic sample, and students who leave will be counted as attrition.

Any missing student outcomes data will be treated as attrition. No imputation or trimming of outliers will be used on student outcome data. The final analytic sample will be comprised of students with End-of-Grade (EOG) math achievement scores at Grade 8 (for that follow-up period). Any missing student-level data on pretest scores, or other covariates may be imputed by multiple imputation for chained equations if more than five percent of students are missing the data and if available student demographic data explain adequate variance in the data. The decision on whether to impute the covariates data will be made when such data are available and before we examine any outcome data. This analysis plan will be updated at that time to reflect our decision, prior to examining any outcome data. If covariate data are imputed, a dummy variable indicating whether pretest data was imputed for each student will be added to the model in our primary analysis, and sensitivity analyses will be conducted.

The current plan is to conduct the confirmatory analysis using a two-level hierarchical linear model, given the assumption that on average there are two teachers in each school (see Power Analysis section below). We may decide to switch to a three-level model, depending on the number of teachers in each school in the final sample. The pre-analysis plan will be updated with the final model to reflect the decision, after we examine the final sample of schools and teachers, and before examining the outcome data.

Effect size (Hedge's *g*) will be calculated. The practical significance of the intervention effects will be assessed against four benchmarks as described by Lipsey et al. (2012), namely normal student academic growth, policy-relevant gaps in student performance, the size of the effects found in prior educational interventions, and the costs and benefits of the intervention.

Power Analysis

3-level vs. 2-level model. As recommended by Schochet (2005), we did a power analysis for two-level cluster randomized trials using NC state test as outcome variable, given that this replication study randomly assigned schools to conditions and all grade 7 math classes in a school are included in the study.

Design parameter estimations. We referred to the [online intraclass correlation (ICC) database] by Hedges & Hedberg (2013) for empirical estimates of the design parameters, including ICC at school level and the portion of variation explained by covariates used at student and school levels. For 7th grade math studies in North Carolina, the estimated ICC is 0.21, and variation explained at school- and student- level by pretest and demographics are 0.88, and 0.68

respectively. Separately, the project team has calculated ICC at school level based on 2012-13 NC state EoG test data and the value was 0.32, higher than what's reported by Hedges & Hedberg (2013). Considering the differences in ICC estimates, we conducted the power analysis with a few different assumptions of ICCs. We've also used more conservative estimation for the portion of variance explained at different levels (80% and 60%) than Hedges & Hedberg.

Fixed vs. random effect models. The current study uses a *simple cluster random assignment* design; students are nested within school and school is the unit of random assignment. Therefore, we rely on what *PowerUp!* (Dong & Maynard, 2013) provides for two-level cluster random assignment designs with treatment at level 2 to calculate the power of the study. For such simple cluster random assignment designs, the software only provides random effect estimation.

The results of the power analysis are presented in the tables below.

Two-level, student nested within schools; School level randomization

Number of schools in the study	63
Average number of students per school	80
Portion of variance explained at school level by demographics and prior performance measures	80%
Portion of variance explained at student level by demographics and prior performance measures	60%
Power level	80%
Significance level	5%

ICC estimation at school level	0.15	0.21	0.25	0.30	0.32
Minimum Detectable Effect Size (MDES)	0.133	0.154	0.167	0.181	0.186

Correspondence with Ethical Standards for Research

The research protocol will be reviewed by the WestEd Institutional Review Board (IRB). Privacy of all data will be maintained and identifiable student information (e.g., names) will be removed from the study data. A data security plan will be reviewed by WestEd data security officer to help ensure all data is handled, transferred, and stored using secure approach.

References

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions.

Dong, N. & Maynard, R. (2013). PowerUp!: A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quasi-Experimental Design Studies. *Journal of Research on Educational Effectiveness*, *6*, 24–67.

10.1080/19345747.2012.673143.

Gargani, J., & Cook, T. (2005). *How many schools? Limits of the conventional wisdom about sample size requirements for cluster randomized trials* (Working Paper). University of California, Berkeley.

Hedges, L. V. & Hedberg, E. C. (2013). Intraclass correlations and covariate outcome correlations for planning 2 and 3 level cluster randomized experiments in education. *Evaluation Review*. *37*(6), 445-489.

Lipsey, M.W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M.W., Roberts, M., et al. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. *National Center for Special Education Research*. National Center for Special Education Research. Washington, DC 20202. Retrieved July 15, 2016, from https://ies.ed.gov/ncser/pubs/20133000/pdf/20133000.pdf

Rochelle, J., Feng, M., Murphy, R. & Mason, C. (2016). Online Mathematics Homework Increases Student Achievement. AERA OPEN. October-December 2016, Vol. 2, No. 4, pp. 1–12. DOI: 10.1177/2332858416673968 Retrieved from http://ero.sagepub.com/content/2/4/2332858416673968

Schochet, P. Z. (2005). *Statistical power for random assignment evaluations of education programs*. Mathematica Policy Research Inc.

Schochet, P. Z. (2008). Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioral Statistics, 33*(1), 62–87.