

Abstract Title Page

Title: Improving Student Outcomes with mCLASS: Math, a Technology-Enhanced CBM and Diagnostic Interview Assessment

Authors and Affiliations:

Ye Wang (Amplify Education)

Matthew Gushta (Amplify Education)

Abstract Body

Background / Context:

The No Child Left Behind Act resulted in increased school-level implementation of assessment-based school interventions that aim to improve student performance. Diagnostic assessments are included among these interventions, designed to help teachers use evidence about student performance to modify and differentiate instruction and improve student outcomes.

The mCLASS: Math software (Ginsburg, Cannon, Eisenband, & Pappas, 2006) is comprised of screening/progress monitoring curriculum-based measures (CBMs) and Diagnostic Interviews to help teachers identify students' skill levels. mCLASS: Math enables teachers to target instruction to each student's needs and monitor each student's progress toward mastery. Educators are expected to use the constantly updated diagnostic information to improve ongoing instruction and increase student achievement. Schools have found that the use of mCLASS: Math helps support curriculum, instruction, and assessment (Ginsburg et al, 2006).

A large mid-western state implemented the technology-supported mCLASS: Math assessment to be administered to K-2 students multiple times each school year. While mCLASS: Math is implemented in that state, schools can choose to participate or not. If a school opts in to the K-2 Formative Assessment Grant (whereby the state supports implementation of mCLASS: Math), all teachers in K-2 must use mCLASS. However, there are instances in which individual students may be exempted from the assessment based on eligibility guidance provided by the state, the presence of an IEP or special education designation excluding a student from the assessment, mobility, illness, or other absence.

Therefore, the present study was designed to examine whether or not exempting students out of mCLASS: Math leads to differences in math achievement among students. This study can help to answer the question of how effective implementation of technology may improve students' math outcomes.

Purpose / Objective / Research Question / Focus of Study:

This study provides evidence about the effectiveness of mCLASS: Math in improving student outcomes on a statewide math achievement test, examining whether or not using the mCLASS: Math CBM impacts students' achievement test scores. Among the students who were administered the CBM, we investigate whether or not being administered the mCLASS: Math Diagnostic Interviews further influences subsequent test scores.

Propensity score matching methods (Fan & Nowell, 2011) were used in predicting the impact of Grade 2 student mCLASS: Math performance from 2010-2011 on Grade 3 statewide math achievement test performance, one year later (2011-2012). It was hypothesized that students who were administered the mCLASS: Math CBM would demonstrate higher achievement test scores than students who were exempted from mCLASS: Math CBM in Grade 2. Furthermore, among the students who were administered mCLASS: Math CBM, students who were also administered the Diagnostic Interviews were hypothesized to demonstrate improved outcomes over those who were not administered the Diagnostic Interviews.

Setting:

This study was conducted in a large mid-western state which provided data from 175 districts, 606 schools, and 1856 teachers for this study. All the schools and teachers had elected to implement mCLASS: Math; however, there are instances in which individuals students may be exempted from the assessment.

Population / Participants / Subjects:

Data from 41,363 Grade 3 students in 2011-2012 school year were analyzed. Among those students, 50.7% were male, 49.2% were female; 67.1% were white, 14.6% were African American, 11.3% were Hispanic, 1.4% were Asian or Pacific Islander, 5.6% were other ethnicities; 7.8% were English learners, 92.1% were native speakers; 13.5% were special education students, 86.2% were not special education students; 60.6 % were eligible for free or reduced priced lunch and 39% were not eligible for free or reduced priced lunch.

Among the 41,363 students, 29,631 were administered mCLASS: Math CBM in Grade 2 with high fidelity implementation during the 2010-2011 school year while 11,732 students were either exempted from mCLASS: Math or identified as low fidelity implementation in Grade 2. We control the fidelity of implementation based on the criterion informed by NCRTI(2010). We applied the following fidelity criteria: (1) students were administered CBM screening tools at each benchmark period; (2) low-performing students who were identified by the screening tools were progress monitored as needed according to the mCLASS: Math user guide.

Among the 29,631 students who were administered mCLASS: Math CBM with high fidelity implementation, 339 students from 27 schools were also administered at least one mCLASS: Math Diagnostic Interview.

Intervention / Program / Practice:

mCLASS: Math CBM is comprised of measures reflecting three broad categories of mathematical thinking: Number Sense, Written Computation, and Quick Retrieval (Ginsburg et al, 2006). After students complete all the applicable measures for their grade level and assessment period, an overall risk level status of Deficit, Emerging, or Established is assigned, signifying the amount of instructional support that may be necessary for a student to help guide them toward benchmark level. All Kindergarten and Grade 1 measures are individually administered for one minute each. The Grade 2 and 3 measures are administered to the entire class in one session. The duration of each Grade 2 and 3 measure is two minutes.

mCLASS: Math Diagnostic Interviews are comprised of individual student interviews that provide detailed analyses of the thinking that underlies student's mathematical performance. Thus, screening and progress monitoring CBM help teachers identify students' risk and mastery levels while the Diagnostic Interviews reveal students underlying thinking. Further, the technology used to administer mCLASS: Math harnesses the power of the Diagnostic Interviews in service of formative assessment by providing and capturing in-depth knowledge of student strategies, concepts, and thinking. Each DI is administered individually and requires approximately 5-7 minutes.

Research Design:

This study has two parts. The first part compares the performance of students who were administered mCLASS: Math CBM with high fidelity implementation to those who were either

exempted from mCLASS: Math or administered with low fidelity implementation; the second part is to compare the performance of students who were administered at least one mCLASS: Math Diagnostic Interviews to those who were not. As a statewide initiative supporting school-level implementation, random assignment to conditions was not available for this study. Therefore, we employed propensity score matching in the analysis of CBM and DI data. Propensity score methods rely on a model of treatment assignment to identify comparable individuals on the basis of similar probabilities of receiving treatment (Fan & Nowell, 2011).

In our mCLASS: Math database, there was a wide variety of student characteristics from which we could choose to build the propensity score. To ensure that we chose appropriate characteristics, we used an iterative model selection procedure with different subsets of characteristics as well as the full set of characteristics to perform the propensity score matching. The propensity score was estimated using a logistic regression model. For each model, we examined model performance and the balanced of covariates across treatment and comparison groups. The model that demonstrated the desired statistical characteristics was then used to estimate the treatment probabilities for the control and treatment groups.

In the first part of this study, because the control group had smaller sample size than the treatment group (i.e., 29,631 students in the treatment group versus 11,732 in the control group), in order to preserve treatment data, we relied on covariate adjustment using the propensity score approach (Austin, 2011). This approach maintains generalizability since no subjects were discarded. We first calculated the propensity score based on student demographic characteristics and subsequently regressed propensity score and an indicator variable denoting treatment status on the outcome variable: student achievement test score. The covariates included were ethnicity and eligibility for free or reduced-price lunch; other demographic characteristics did not perform well in the matching model and were excluded (the whole list of demographic information is in Table 1).

In the second part of this study, students were matched across conditions according to baseline characteristics (i.e., demographic information and CBM scores) which reduced selection bias due to observed potential confounders. Due to sample size disparity between conditions (339 in the treatment versus 29,292 in the control group), we used the nearest neighbor matching approach (Austin, 2011). In this approach, an individual from the control group is chosen as a match for an individual in the treatment group based on the closest propensity score. The covariates were students' CBM scores on the most relevant mCLASS: Math measures in Grade 2 at the beginning of the year; demographic characteristics did not perform well in terms of matching and balance and were excluded from the model. We also used the absolute value of the standardized difference in means between treatment and control groups introduced by Rosenbaum and Rubin (1983) to examine bias reduction before and after matching.

Data Collection and Analysis:

For the first part of this study, Table 1 shows descriptive statistics according to demographic characteristics of the treatment and comparison groups. As shown in Table 1, the two chosen covariates represent the largest difference in students' scores on demographic information between the two conditions. We then built a propensity score model whereby each student was assigned a propensity score, which is the predicted value of the logistic regression model estimated using the selected covariates. A linear regression analysis was then conducted using the whole data set, regressing statewide achievement test scores on the binary treatment indicator and the propensity scores.

For the second part of this study, Table 2 shows the means and standard deviations of the covariates from the two groups before matching, after matching, and the balance improvement on the covariates. The matching resulted in a subset of 339 control group students who matched closely with the 339 treatment group students. As shown in Table 2, the mean score difference between the two groups on the Computation measure in mCLASS: Math CBM reduced from -1.09 to 0.01 when treatment and comparison groups were matched; the mean score difference on Number Facts in mCLASS: Math CBM reduced from -2.02 to -0.01. An ANOVA analysis was conducted using the matched data to compare the achievement test scores.

Findings / Results:

For the first part of this study, the results suggest that, after controlling for the propensity score, mCLASS: Math CBM was positively associated with change in achievement test score ($b = 10.90$, $S. E. = 0.79$, $p < 0.01$). The Hedges's g effect size was 0.25; effect sizes of 0.25 or greater are considered to be "substantively important" by the What Works Clearinghouse (USDOE, 2010). The Outcome Results section of Table 1 provides the means and standard deviations of the two groups on the achievement test.

For the second part of this study, the results suggest that, after matching based on CBM scores, students who were administered at least one mCLASS: Math Diagnostic Interviews showed significantly higher achievement test scores ($F = 5.48$, $p < 0.05$). The Hedges's g effect size was 0.18. This effect, however, was not present before matching ($F = 0.25$, $n. s.$). The Outcome Results section of Table 2 provides the means and standard deviations of the two groups on the achievement test before and after matching.

Conclusions:

CBM measures quickly identify students who are at risk for poor mathematical performance while Diagnostic Interviews identify students' mathematical strengths and weaknesses which can be used to inform instruction. Overall, the administration of CBM measures provides broad, useful information about students' skill levels and identify those students in need of further assessments and development. The data obtained from the administration of the Diagnostic Interviews extends CBM results by providing specific information about students' problem solving strategies, concepts, and thinking. Using propensity score matching to reduce bias across treatment and comparison conditions, the results of this study revealed that when schools/teachers administered mCLASS: Math CBM and Diagnostic Interview(s) to Grade 2 students, achievement test scores were increased in Grade 3. However, while mCLASS: Math CBM is used widely in schools, Diagnostic Interviews are not - perhaps because of the amount of time it takes to administer each interview. The present study suggests that the coupling of CBM measures and Diagnostic Interviews provides teachers with a deeper understanding of students' math ability, allowing them to better target instruction and intervention, resulting in improved student achievement over the use of CBM measures alone.

Appendices

Appendix A. References

- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*, 399-424.
- Fan, X., & Nowell, D. L. (2011). Using propensity score matching in educational research. *Gifted Child Quarterly, 55*(1), 74-79.
- Ginsburg, H., Cannon, J., Eisenband, J., Pappas, S. (2006). Mathematical Thinking and Learning. In McCartney, K. (Ed.), Phillips, D. (Ed.). *Blackwell Handbook of Early Childhood Development* (pp. 208-229). Malden: Blackwell Publishing.
- National Center on Response to Intervention. (2010). *Essential components of RTI—A closer look at response to intervention*. Washington, DC: U.S. Department of Education, Office of Special Education Programs, National Center on Response to Intervention.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41-55.
- U.S. Department of Education, Institute of Educational Sciences. (2010). *What Works Clearinghouse Procedures and Standards Handbook* (Version 2.1).

Appendix B. Tables and Figures

Table 1. Part 1: CBM Analyses.

	Control Group		Treatment Group		<i>P</i>
	Mean	<i>SD</i>	Mean	<i>SD</i>	
Covariates					
*Ethnicity (White as 0, not White as 1)	0.94	0.24	0.98	0.14	<0.01
*Eligibility for free or reduced lunch (yes as 0, no as 1)	0.27	0.45	0.44	0.50	<0.01
Gender (male as 0, female as 1)	0.49	0.50	0.49	0.50	<i>n. s.</i>
English learner (yes as 0, no as 1)	0.92	0.28	0.92	0.27	<i>n. s.</i>
Special educational (yes as 0, no as 1)	0.85	0.36	0.87	0.34	<i>n. s.</i>
Outcome Results					
Math achievement test scores	448.53	75.42	467.24	73.87	<0.01

* Covariates included in the estimation of propensity scores.

Table 2. Part 2: Diagnostic Interview Analyses.

	Treatment		Unmatched Control		Matched Control		Difference Before Match	Difference After Match
	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>		
Covariates								
Computation CBM score	5.91	4.47	7.01	5.27	5.90	4.44	-1.09	0.01
Number Facts CBM score	15.27	7.73	17.29	8.68	15.28	7.76	-2.02	-0.01
Outcome Results								
Math achievement test scores	465.12	73.54	467.14	73.86	451.96	72.80	-2.02	13.16