Abstract Title Page

Title:

Assessing Impacts of *Math in Focus*, a ‘Singapore Math’ Program for American Schools

Authors and Affiliations:

Andrew P. Jaciw (Empirical Education)

Whitney Hegseth (University of Michigan)

Megan Toby (Empirical Education)

Background

The Common Core State Standards (CCSS) have been developed in response to the criticism that students in the U.S. are graduating from high school without being college and career ready and that they are falling behind their counterparts in other countries in key subject areas (Common Core State Standards Initiative, 2014). The mathematics curricula in the U.S. have been described as "a mile wide and an inch deep" compared to curricula in countries that outperform the U.S. on international tests, which focus on a smaller number of topics in greater depth (Schmidt, Wang, & McKnight, 2005). CCSS attempts to address this deficit in the mathematics curriculum by stressing conceptual understanding of key ideas and a focused, coherent, and rigorous approach to the subject matter organized around eight principles of practice (Common Core State Standards Initiative, 2014). Currently forty-four of the fifty states, plus the District of Columbia, belong to the Common Core State Standards Initiative.

As participating states integrate the CCSS, the expectation is that SEAs and LEAs will select specific supporting curricula, teaching tools, and resources to plan units and create lessons with. In this work, we report the results of an efficacy study that investigated the impact of one such curriculum—Math in Focus: Singapore Math (*MIF*)—developed by Houghton Mifflin Harcourt (HMH) that, according to the program developer, provides comprehensive support for CCSS. We also examine whether impacts of *MIF* vary by ethnicity, reflecting the priority of proponents of CCSS to raise performance while closing the achievement gap (National Governors Association, the Council of Chief State School Officers, & Achieve, Inc., 2008). Black students in the U.S. have performed especially poorly on international assessments (Baldi et al., 2007).

Because CCSS are being newly implemented, there is no track record of studies of impact of CCSS-aligned curricula on student achievement outcomes. However, CCSS grew out of prior standards—eight principles of mathematical practice that were adapted from five process standards of the National Council of Teachers of Mathematics and five strands of proficiency in the *Adding it Up Report* from the National Research Council. This allows us to briefly examine the literature on the impact of programs aligned with these reform-based predecessor standards.

Slavin & Lake (2007) reviewed research on a variety of elementary mathematics curricula, ranging from the reform-based, NSF-supported *Everyday Mathematics*, to *Saxon Math*, which is described as the "antithesis of constructivist approaches." They found most studies of reform-based curricula to be of "marginal methodological quality" and impacts on standardized assessments were "thin." The authors note that reform-based mathematics programs may have positive effects on other outcomes not measured by standardized tests. Due to the lack of evidence in support of the effects of different math curricula, the authors determined that more research is needed on these programs.

Given the conclusions of Slavin and Lake (2007), we adjusted our focus to interventions reviewed by WWC that have been found to meet evidence standards with or without reservation. Agodini et al. (2010) compared four curricula head to head, using a randomized control trial (RCT). The study authors (as well as Slavin & Lake) considered *Investigations in Number, Data and Space* (*Investigations)* to be a student-centered program, *Math Expressions* to be a blend of student-centered and teacher-directed instruction, and *Saxon Math* and *Scott Foresman -Addison Wesley Elementary Mathematics* (*SFAW*) to be more traditional, teacher-led programs. A comparison of *Investigations* against both *Saxon* and *SFAW* showed no impact in Grades 1 or 2 on the ECLS-K math assessment. There was a positive impact of *Math Expressions* compared to *SFAW* in first and second grades on the ECLS-K. Another RCT (Gatti & Giordano, 2010) compared *Investigations* to a traditional skills-based program and found no impact in 1st grade and a .25 standard deviation positive impact in 4th grade on a standardized multiple choice test (GMADE). An experimental study by Waite (2000) of *Everyday Mathematics*, another NSF-funded intervention, found positive and statistically significant effects on overall math and subtests (concepts, operations and problem solving); however, the WWC deemed the results from the study to not be statistically significant. An RCT of *enVisionsMATH* by Pearson (Resendez & Azin, 2008), which features problem-based instruction and small-group interaction, found a positive impact on tests of concepts and communications, math computation and problem solving and reasoning. (*enVisionsMATH* or *Investigations* were used in 81% of control classes in the RCT reported in this study.)

Our review, which is shortened to fit the space of this proposal, reveals that there is no clear-cut and generalizable conclusion concerning the efficacy of programs based on standards that are predecessors to CCSS. Many studies used inferior methods and results from studies that meet methodological standards are equivocal. It may take many rigorously designed studies to work out the complexities concerning the efficacy of CCSS-based curricula. The results from our study of *MIF* provide initial evidence to what will be an emerging picture of the general impact of CCSS-based curricula on mathematics achievement.

Purpose and Research Questions

The purpose of the current work is to report the results of an efficacy study that investigated the impact and differential impact of one CCSS-aligned curriculum – *MIF* – on mathematics achievement. The research questions are as follows:

- Is there a positive impact of *MIF* on student skills in mathematics problem solving?
- Is there a positive impact of *MIF* on student math procedural skills?
- Is *MIF* differentially effective in its impact on student achievement depending on (1) the ethnicity of the student? (2) the incoming achievement level of the student?

In addition to addressing these questions, we document levels of fidelity of implementation.

The work reported in this paper provides an assessment of the impact of a CCSS-aligned intervention using an experimental and <u>within-culture</u> comparison. This is important because while international comparisons of student achievement lead to discussions of discrepancies in curricula as the cause of difference in performance, other kinds of contextual and systemic differences in schooling that exist between cultures may drive the performance differential. This study provides an apples-to-apples assessment of impact by researching the questions within a specific U.S. context.

Setting

The research took place during the 2011-2012 school year across twelve elementary schools in one urban school district in Nevada. The district has a total enrollment of approximately 300,000 students. Seventeen percent of students were English Language Learners, 32% were white, 12% black, 42% Hispanic, and 7% Asian.

Participants

Ninety-three teachers of grades 3, 4 and 5 were recruited to participate in the study, with 41 teachers randomized to the *MIF* group and 52 teachers randomized to the control group. Rosters were provided for 2235 students in participating teachers' classrooms.

Intervention

As its name implies, *Math in Focus* is specifically modeled after pedagogical approaches used in Singapore. CCSS are considered well-aligned to Singapore's Mathematics Syllabus. *MIF* meets three core criteria around which the CCSS are organized (Common Core State Standards Initiative, 2010) (1) <u>Coherence:</u> CCSS emphasize mathematics as a coherent body of knowledge with topics introduced in earlier grades being connected and extended to coverage in later topics and with reinforcement of major topics in a grade. With *MIF*, concepts are connected across grade levels with higher-level coverage as one proceeds through the grades. (2) Focus: CCSS focus on fewer topics emphasizing depth over breadth. Consistent with this, *MIF* is organized around fewer topics and the goal is to teach them more thoroughly. (3) Rigor: CCSSs emphasize a rigorous approach, balancing conceptual understanding, procedural skills and fluency, and application. An important implication is that problem solving skills do not come at the expense of procedural skills because they are part of a single coherent approach to learning and using mathematics. A central feature of *MIF* is the "concrete to pictorial to abstract" (CPA) approach which is designed to support conceptual understanding.

The *MIF* curriculum has the following components. (1)Teachers lead students through an Instructional Pathway consisting of guided practice, and then student practice and apply their learning. (2)Teachers differentiate instruction and iterate between teaching and letting students solve problems on their own. (3)*MIF* materials include textbooks, student workbooks, implementation guides, transition guides (to make connections to prior grade-level materials), a 30-student manipulative kit, and digital resources (a test generator, virtual manipulatives, online Transition Resource Map, math background videos, student interactivities, common core Focus Lessons and Activities).

The program duration was one year. The counterfactual included the following curricula (values in parentheses show the number of teachers reporting use of each program): *Envisions* (27), *Investigations* (17), *Scott Foresman* (5), *Pearson SuccessNet* (2), *Everyday Math* (2)
No set curriculum (1).

Research Design

The design was a group randomized trial lasting one year. We worked with HMH to recruit 12 schools with grades 3, 4, and 5. We randomized intact grade-level teams that volunteered for participation to the *MIF* and control groups. Randomizing whole teams allows collaboration within grades, which is an important component of *MIF*. Technically, each school constituted a randomized block, with the two randomized teams (grades 4 and 5 in one team, and grade 3 in the other) forming a matched pair. For the schools that did not have a participating grade 3, we randomized one of grades 4 or 5 to treatment and the other to control. Altogether we randomized 22 grade-level teams. Twelve were assigned to *MIF*, the rest to control. The achievement outcomes were assessed in the spring of the year following random assignment. Using the available samples and plausible values for the design parameters we powered the study to detect impacts as small as .28 standard deviations in the outcome, assuming Type-1 error of 5% and Power 80%. Math performance was assessed using the Stanford Achievement Test (SAT10) problem solving and math procedures scales**.** We chose the SAT 10 because it is closely aligned with the Common Core Standards for these two scales.[1] A second assessment was the Nevada Criterion-Referenced Test (CRT) which is a state standards-based assessment that functions as an indicator of student performance.

Data Collection and Analysis:

We used a two-level hierarchical linear regression model (Raudenbush & Bryk, 2002) to estimate the impacts of *MIF* on student achievement. Students were modeled at level 1 and grade teams at level 2, which reflects the design, with teams randomized to conditions and outcomes

[1] Information on Pearson's alignment study of SAT 10 with Common Core Standards may be found at: http://www.pearsonassessments.com/hai/images/PDF/Stanford_10_Alignment_to_Common_Core_Standards.pdf

assessed at the student level. Assignment to condition was modeled using a dummy variable. Outcomes were analyzed together for grades 3, 4, and 5. The model included grade-level random effects and dummy variables for schools to reflect the randomized block design. A series of covariates, including the pretest, were used to increase precision. Differential impacts were assessed through a term for the interaction between the moderator and the treatment dummy.

Findings / Results

Attrition. Table 1 shows changes in the samples between the point of randomization and analysis. The rates of overall attrition for SAT 10 Problem Solving were 18% and 27% at the randomization and student levels, respectively. Similar levels of attrition were experienced for the SAT 10 Procedures scale. Attrition was lower for the CRT. Equivalence tests conducted on the analysis samples showed no significant differences for each of the three scales.

Implementation Fidelity. Three criteria were used to assess fidelity: Thirty-eight percent of *MIF* teachers (n = 15) reported teaching *MIF* at least 80% of the time they devoted to math instruction in their classrooms. Eighty-two percent of teachers (n = 32) reported implementing with fidelity in terms of incorporating elements of the Instructional Pathway. Sixty-five percent of teachers (n = 22) met the third criterion of using the CPA approach.

Impact and Differential Impact. Results for the three impact analyses are displayed in Table 2 and for differential impact analyses in Table 3. The main results are as follows.

- A high level of confidence in a positive impact of *MIF* on SAT10 Problem Solving (p=.05). The standardized effect size is 0.12, and the difference in percentile standing is 5%.
- Some confidence in a positive impact of *MIF* on SAT10 Procedures (p=.10). The standardized effect size is .14 with a difference in percentile standing of 6%.
- No impact on the CRT (p=.54).
- No difference in impact by level of pretest or minority status.

Conclusions

The study gives preliminary evidence concerning the impact of one CCSS-aligned math intervention on student performance on two mathematics strands. The result gives us confidence that *MIF* is beneficial for problem solving and may be advantageous for procedural skills also. Importantly, the impact is achieved *in spite of the counterfactual conditions consisting largely of other reform-based programs*. The results do not support the hypothesis that CCSS-aligned curricula narrow the achievement gap. Importantly, the RCT reported in this work involves a within-U.S. comparison, allowing us to assess impact while holding constant other factors that may be responsible for performance differentials observed on international assessments of achievement.

Assessing impacts of CCSS-aligned curricula with RCTs gives us a fresh start to understanding impacts of student-centered and reform-based curricula. Results from past studies have been inconclusive in part because of weaker research designs, and because studies of the question are necessarily complex—each study involves a combination of specific program characteristics, counterfactual treatments, assessments and subscales, populations and subgroups, and contexts of schooling and instruction. We cannot avoid this complexity. Therefore, it is important as we build a track record of results of studies of impacts of CCSS-aligned curricula to account for these differences to be able to draw accurate generalized inferences concerning program impacts..

Appendices

Appendix A. References

Agodini, R., Harris, B., Thomas, M., Murphy, R., & Gallagher, L. (2010). *Achievement Effects of Four Early Elementary School Math Curricula: Findings for First and Second Graders* (NCEE 2011-4001). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Baldi, S., Jin, Y., Skemer, M., Green, P.J., & Herget, D. (2007, December). *Highlights from PISA 2006: Performance of U.S. 15-Year-Old Students in Science and Mathematics Literacy in an International Context*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Common Core State Standards Initiative. (2014). *Read the Standards*. Retrieved September 30, 2014 from http://www.corestandards.org/read-the-standards/

Common Core State Standards Initiative. (2010, June). *Common Core State Standards for Mathematics*. Retrieved September 30, 2014 from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf

Cronbach, L. J., (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 116-127.

Gatti, G., & Giordano, K. (2010). *Pearson Investigations in Number, Data, and Space efficacy study: Final report.* Pittsburgh, PA: Gatti Evaluation, Inc.

National Governors Association, the Council of Chief State School Officers, & Achieve, Inc. (2008). *Benchmarking for Success: Ensuring U.S. Students Receive a World-Class Education.* Retrieved September 30, 2014 from http://www.corestandards.org/assets/0812BENCHMARKING.pdf

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.

Resendez, M., & Azin, M. (2008). *A study on the effects of Pearson's 2009 enVisionMATH program. 2007–2008: First year report.* Jackson, WY: PRES Associates, Inc.

Schmidt,W.H., Wang, C.H., & McKnight, C.C. (2005). Curriculum coherence: An examination of U.S. mathematics and science content standards from an international perspective. *Journal of Curriculum Studies, 37*(5), 525-559.

Slavin, R.E., & Lake, C. (2007, February). *Effective Programs in Elementary Mathematics: A Best-Evidence Synthesis*. Retrieved September 30, 2014 from http://www.bestevidence.org/word/elem_math_feb_9_2007.pdf

Waite, R.D. (2000). A study of the effects of *Everyday Mathematics* on student achievement of third-, fourth-, and fifth-grade students in a large north Texas urban school district. *Dissertation Abstracts International, 61*(10), 3933A. (UMI No. 9992659)

Appendix B. Tables and Figures

TABLE 1. NUMBERS OF UNITS IN THE EXPERIMENTAL GROUPS AND ATTRITION OVER TIME

	Control				*MIF*			
Event	No. of schools	No. of teams	No. of teachers	No. of students	No. of schools	No. of teams	No. of teachers	No. of students
Randomization	10	10	41	n/a	12	12	52	n/a
(Loss prior to rosters)	(1)	(1)	(4)	n/a	(1)	(1)	(6)	n/a
Fall rosters received	9	9	37	1025	11	11	46	1210
SAT 10 Problem Solving Analytical sample								
(Loss due to lack of posttest)	0	0	(2)	(241)	(2)	(2)	(7)	(353)
Final count of units with SAT 10 Problem Solving[a]	9	9	35	784	9	9	39	857
SAT 10 Procedures Analytical sample								
(Loss due to lack of posttest)	0	0	(2)	(233)	(2)	(2)	(7)	(375)
Final count of units with SAT 10 Procedures[b]	9	9	35	792	9	9	39	835
CRT Analytical sample								
(Loss due to lack of posttest)	0	0	0	(84)	0	0	0	(84)
Final count of units with CRT posttest	9	9	37	941	11	11	46	1126

[a] Of the 241 control students without posttests, 57 were lost because of lack of responses from the two attrited teachers in that condition, and 184 were lost from teachers for whom we have responses for at least some other students; of the 353 *MIF* students without posttests, 168 were lost due to no outcomes from the two randomized teams, and 185 were lost from teachers for whom we have responses for at least some other students.

[b] Of the 233 control students without posttests, 57 were lost because of lack of responses from the two attrited teachers in that condition, and 176 were lost from teachers for whom we have responses for at least some other students; of the 375 *MIF* students without posttests, 168 were lost due to no outcomes from the two randomized teams, and 207 were lost from teachers for whom we have responses for at least some other students.

Note. In the above table, most schools are double counted because grade-level teams from both conditions are in most of the participating schools.

TABLE 2. EFFECT SIZES FOR IMPACTS ON MATH

	Condition	Means	Standard deviations	No. of students	No. of teams	No. of schools	Effect size	*p* value	Percentile standing
SAT10 Problem Solving	Control	639.96	43.05	784	9	9	0.12	.05	5%
	MIF	644.47	40.62	857	9	9			
SAT10 Procedures	*Control*	634.28	47.33	792	9	9	0.14	.10	6%
	MIF	640.38	45.05	835	9	9			
CRT	Control	0.00	1.00	941	9	9	0.05	.54	2%
	MIF	0.05	1.07	1126	11	11			

The adjusted effect size was computed by dividing the regression-adjusted effect estimate by the standard deviation of the posttest scores for the control group. Between-grade differences in the posttest were factored out of the standard deviation in the denominator of the effect size. The *p* value corresponds to the significance test for the effect of *MIF* in the regression model. The program mean was obtained by adding the regression-adjusted estimate of the average one-year effect of *MIF* to the unadjusted control mean.

Modeling separate school effects leads to estimates of control-group performance which are specific to schools. For purposes of display, to set the performance estimate for the control group, we compute the overall average performance for the sample of control cases used to calculate the adjusted effect size. The estimated *MIF* effect, which is constrained to be constant for each grade block (i.e., it is modeled as fixed), is added to this estimate to show the relative advantage or disadvantage to being in the *MIF* group.

TABLE 3. DIFFERENCES IN EFFECTS OF *MIF* ON STUDENT ACHIEVEMENT FOR SUBGROUPS OF STUDENTS

	SAT10 problem solving			SAT10 Procedures			CRT		
	Estimated Effect	*p* value	Effect size	Estimated Effect	*p* value	Effect size	Estimated Effect	*p* value	Effect size
Added Effect for non-Minorities	1.57 (2.96)	.60	.04	4.79 (3.76)	.20	.10	.14 (.07)	.06	.14
Added Effect for a one SD increase in pretest	.04 (1.29)	.97	<.01	1.52 (1.61)	.35	.03	.01 (.03)	.82	.01

Note. Number in parentheses is the standard error.