Abstract Title Page

Title: Findings from a Multi-Year Scale-Up Effectiveness Trial of Everyday Mathematics

Authors and Affiliations: Michael Vaden-Kiernan (AIR), Geoffrey Borman (University of Wisconsin-Madison), Sarah Caverly (AIR), Nance Bell (AIR), Veronica Ruiz de Castilla (AIR), Kate Sullivan (AIR) and Debra Rodriguez (AIR).

Abstract Body

Background / Context:
This study addresses the effectiveness of *Everyday Mathematics*, a widely used core mathematics curriculum that reflects over two decades of National Science Foundation (NSF)-sponsored research and development studies (Klein, 2007; National Research Council, 2004) and aligns well with recommended policies and practices by the National Council of Teachers of Mathematics (NCTM) Curriculum Focal Points (2006) and National Mathematics Advisory Panel (NMAP) (2008). This and other similar curricula are increasingly needed to strengthen student math skills and ensure that all children are at or above grade level in math proficiency. Statistics present concerns to reaching this goal - only 42% of fourth grade students and 35% of eighth grade students are at or above proficient-level on National Assessment of Educational Progress (NAEP) math scores (National Center for Educational Statistics, 2013). Recent initiatives emphasize the role of early math instruction and curricula, recognizing that students need to develop a well-defined set of skills in critical math content areas (e.g., whole numbers, fractions, elements of geometry, and measurement) in early grades in order to prevent difficulties in later grades. The NMAP (2008) recommended that all students receive effective preparation from an early age to ensure their later success in algebra and emphasized the need for early math programs that mitigate and prevent difficulties. To date, however, *Everyday Mathematics* is only one of a few elementary school math curriculum reviewed by the What Works Clearinghouse (WWC) (2007) that has demonstrated "potentially positive effects," with the evidence for effectiveness on math achievement rated as medium to large[1].

The *Everyday Mathematics* (EM) curriculum, developed by the University of Chicago School Mathematics Project (UCSMP) during the mid-1980s, and published by Wright Group/McGraw-Hill, relies on research-based practices, which according to the National Research Council (2004), are supported by more researchers and empirical studies than any other elementary mathematics curriculum. EM is a PK-6 curriculum emphasizing six content strands (numbers and numeration; operations and computation; data and chance; measurement and reference frames; geometry; and patterns, functions, and algebra) with learning targets or curriculum focal points identified for each of the six strands at each grade level. Developed largely from NSF-sponsored studies, the curriculum shows promise in preventing math difficulties in early grades. It emphasizes a constructivist philosophy, with a strong emphasis on real-life problem solving, manipulatives, concept development, and targeted use of technology and parent participation. Besides being the highest rated elementary school math curriculum in terms of effectiveness by the WWC (2007), it has also been widely adopted and used across the country. EM is used by nearly 4 million students in more than 11,000 schools in more than 3,000 districts in all 50 states (SRA/McGraw-Hill, 2009). Results from quasi-experimental evaluations have revealed statistically significant math achievement advantages of +0.16 relative to other math curricula (Carroll, 1998; Carroll & Isaacs, 2003; Riordan & Noyce, 2001; SRA/McGraw-Hill, 2003; Waite, 2000; WWC, 2009; Woodward & Baxter, 1997). In addition, a cluster randomized

[1] "Potentially positive effects" is evidence of a positive effect in a domain with no overriding contrary evidence. Programs rated as having a "medium to large" amount of evidence require at least two studies that meet the WWC evidence screen with 2 schools and a total sample size of at least 350 students or 14 classrooms across the studies. The screened studies reviewed for EM included a total of approximately 12,600 students in grades 3-5 from a range of socioeconomic backgrounds and attending schools in urban, suburban, and rural communities in multiple states.

controlled trial (RCT) currently being conducted by the publisher is documenting potential impacts of EM on math achievement in grades K-5 students in three schools in Washington D.C. The study is assessing classroom-level impacts of treatment assignment on the TerraNova math achievement test scores (SRA/McGraw-Hill, 2009b). Despite the curriculum's widespread use and promising research findings, EM has not been evaluated rigorously on a large scale as part of an objective, third-party evaluation.

Purpose / Objective / Research Question / Focus of Study:
Given the importance of early mathematics instruction and curricula for preventing mathematics difficulties in later grades, it is necessary to identify effective mathematics curricula and instruction to ensure that children become proficient in early mathematics content and procedures. *Everyday Mathematics* was reviewed by the What Works Clearinghouse and is reported to have "potentially positive effects" on students' mathematics achievement. However, most of the studies that have evaluated EM have used quasi-experimental designs or are small-scale randomized control trials. This study reports the preliminary year one findings for Kindergarten and 3rd grade cohorts of the first scale-up evaluation of this widely used curriculum. The results of this study will contribute to understanding whether EM is effective in promoting mathematic proficiency in the elementary grades when implemented "at scale" with typical "real world" levels of support. The study was designed to address the following research questions:

- **Overall Impacts**. Does school-level assignment to the *Everyday Mathematics* curriculum intervention produce stronger effects on math achievement than assignment to the "business-as-usual" control condition?
- **Impacts by Subgroups**. Is there significant variation in the outcomes of *Everyday Mathematics* or do the effects reliably replicate across student subgroups, the sampled classrooms/teachers, schools, and districts?
- **Fidelity of Implementation**. To what extent was the intervention delivered as the curriculum developers indicated it should be implemented? Was there significant variation in implementation fidelity of *Everyday Mathematics* among the classrooms/teachers, schools, and districts? In what ways were *Everyday Mathematics* students' experiences similar or different to those of students in the control condition?
- **Proximal Outcomes as Mediators of Impacts**. Is there a significant relationship between proximal student and teacher outcomes, such as fidelity of implementation or student motivation/engagement, and student math achievement outcomes and does this relationship vary by classrooms/teachers, schools, and districts?

Setting:
The study was conducted in a sample of 48 elementary schools (kindergarten through 5th grade) in 7 districts across the country across a two year time frame.

Population / Participants / Subjects:
The study participants include approximately 4,500 elementary school students and 1,200 teachers per year.

Intervention / Program / Practice:
EM is a core mathematics curriculum for grades prekindergarten to six. The curriculum emphasizes six strands of mathematics knowledge with learning targets identified for each strand by grade level. It includes student materials, teacher manuals, assessment and practice guides, and home links to support parent involvement. The curriculum includes a 2- to 3-day summer workshop to train teachers to implement the curriculum, as well as follow-up support by EM consultants. The instructional format follows a consistent three-part lesson plan in all grades focusing on teaching the lesson, ongoing learning and practices for students, and differentiated instruction options. Teachers, as part of the intervention, can also use informal and formal assessments to monitor student progress and inform instruction.

Research Design:
The evaluation of the OCR program involved two key elements: the multi-site cluster randomized trial (CRT) and the implementation study. The CRT includes 48 elementary schools across 7 districts that were randomized to receive training and delivery of the EM curriculum (treatment group) or to deliver the standard reading curriculum for the school (control group) blocking at the district level. Districts were recruited over a 3 year timeframe. Schools were randomly assigned to treatment (N=24) and comparison (N=24). The study focused on a grade cohort (grade K-3 and a grade 1-4) each year of the study.

Data Collection and Analysis:
Data from teachers and students in two cohorts (grades K&3 and grades 1&4) were gathered over two school years. This paper is presenting findings from both years of data collection; however, the sample is cross sectional by design. The pre- and post-test outcomes were assessed in the fall and spring using the *Group Mathematics Assessment and Diagnostic Evaluation* (GMADE) and Student Motivation Form (SMF). Fidelity of implementation was captured using classroom observations, interviews, and surveys with teachers and other key staff (e.g., curriculum trainers).

The main intent-to-treat impact analyses uses a three-level model with school-level fall pre-test scores on the GMADE as a covariate and spring post-test scores as the dependent variable, nested within schools, which in turn were nested within districts. Additional subgroup (moderator) analyses were used to investigate the effects of the EM program as a function of student baseline characteristics (e.g., age/grade, gender, baseline math proficiency, student engagement), teacher/classroom characteristics (e.g., class size, fidelity of implementation), and school characteristics (e.g., geographic region or locale). Student, teacher, and school characteristics will also be examined as potential mediators of the effects.

The analysis of fidelity of implementation investigates the latent construct of fidelity of implementation underlying four separate components: dosage, adherence, quality of delivery, and student responsiveness.

Findings / Results:
Table 1 provides school level characteristics of students in the study schools at baseline for each year of the study. The analytical sample is comprised of 4,520 students in grades K and 3 and 4,467 students in grades 1 and 4 across 48 elementary schools with valid scores on the Spring

GMADE assessment. Results indicate that at baseline, there were no significant differences between the demographic characteristics or meant pre-treatment scores on the GMADE for either cohort.

Results from the overall ITT impact analyses are presented in Table 2. The intraclass correlation (ICC) of student math achievement was .154 for schools in year 1 and 0.176 in year 2. The three level model (student, school, and district) includes grand-centered, school-mean GMADE pretest scores and an indicator for treatment condition. Both predictor variables of pretest and treatment condition are included in level 2 of the model.

Level-1:

$$Y_{ijk} = \pi_{0jk} + \sum_{n=1}^{x} \pi_{xjk} + e_{ijk}$$

Level-2

$$\pi_{0jk} = \beta_{00k} + \beta_{01k} Pretest_{jk} + \beta_{02k} EM_{jk} + \sum_{n=2}^{7} \beta_{0xk} + r_{ijk} \quad (2.1)$$

$$\pi_{xjk} = \beta_{0xk} + \sum_{n=1}^{x} \beta_{xjk} \quad (2.2)$$

Level 3:

$$\beta_{00k} = Y_{000} + u_{000} \quad (5.1)$$

$$\beta_{01k} = Y_{100} + u_{100} \quad (5.2)$$

$$\beta_{02k} = Y_{200} \quad (5.3)$$

The multilevel model results indicate that the school mean pretest is predictive of the posttest achievement. For instance, a one standard deviation increase in school pretest score is associated with a .88 increase in outcome scores in Year 1.

The ITT analysis (see Table 2) indicates that the EM program does not have a statistically significant impact on students' math achievement compared to the business as usual curriculum in place in control schools after the first or second year of implementation. None of the subgroup analyses (see Table 3) suggested statistically significant differential impacts of EM except for the subgroup by grade in year 2 with positive impacts for fourth grade students.

Implementation analyses demonstrate surprisingly low levels of uptake in treatment schools. In Year One, only 55% of teachers in treatment schools showed evidence of implementing both the unique and essential components of EM. In Year Two teachers had even lower levels of implementation; only 44% of teachers in treatment schools were in the Unique Class. These analyses also uncovered limited contamination in two schools in Year One (3 teachers) and one school in Year Two (4 teachers). All contamination occurred within the same district.

Conclusions:

This study provides preliminary evidence that the impacts of EM are not significant on overall students' mathematics performance when implemented at scale in a large sample of schools after one or two years relative to other core math curricula. However, there were positive differential impacts for 4th grade students. Future exploratory mediator analyses will explore other indirect relationships with reading outcomes. These findings are particularly important given the large number of students that are exposed to the program across the country and relatively small number of third-party evaluations of this math program as well as others for elementary school students.

Appendices

Appendix A. References

Carroll, W.M. (1998). Geometric knowledge of middle school students in a reform-based mathematics curriculum. *School Science and Mathematics*, *98*(4), 188-197.

Carroll, W.M., & Isaacs, A. (2003). Achievement of students using the University of Chicago School Mathematics Project's Everyday Mathematics. In S.L. Senk & D.R. Thompson (Eds.), *Standards-based school mathematics curriculum: Where are they? What do students learn?* (pp. 79-108). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Klein, D. (2007). A quarter century of U.S. "math wars" and political partisanship. *Journal of the British Society for the History of Mathematics, 22*(1) 22-33.

National Center for Educational Statistics. (2013). *A First Look: 2013 Mathematics and Reading (National Assessment of Educational Progress at Grades 4 and 8).* Washington, DC: U.S. Department of Education, Institute of Education Sciences.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics: A quest for coherence*. Reston, VA: National Council of Teachers of Mathematics.

National Mathematics Advisory Panel. (2008). Foundations for success: The final report of the National Mathematics Advisory Panel. U.S. Department of Education: Washington, DC.

National Research Council. (2004). *On Evaluating Curricular Effectiveness: Judging the Quality of K-12 mathematics Evaluations*. Committee for a Review of the Evaluation Data on the Effectiveness of NSF-Supported and Commercially Generated Mathematics Curriculum Materials. Mathematical Sciences Education Board, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Riordan, J.E., & Noyce, P.E. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education*, *32*(4), 368-398.

SRA/McGraw-Hill (2003). Everyday Mathematics student achievement studies: Volume 4. Chicago: Author.

SRA/McGraw-Hill (2009). Kate Powell, Marketing Manager, Everyday Mathematics, Wright Group McGraw-Hill, personal communication, May 21, 2009.

SRA/McGraw-Hill (2009b). District of Columbia Public Schools Everyday Mathematics, Third Edition Evaluation Research. Unpublished manuscript.

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)

What Works Clearinghouse (2007, July). *What Works Clearinghouse topic report: Elementary School Math.* Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved from: http://ies.ed.gov/ncee/wwc/reports/elementary_math/topic/

What Works Clearinghouse. Revised (2009, March). *What Works Clearinghouse intervention report: Everyday Mathematics.* Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved from: http://ies.ed.gov/ncee/wwc/reports/elementary_math/eday_math/

Woodward, J., & Baxter, J. (1997). The effects of an innovative approach to mathematics on academically low-achieving students in inclusive settings. *Exceptional Children*, *63*(3), 373-388.

Appendix B. Tables and Figures

Table 1 Equivalence test of school characteristics by treatment assignment, Baseline

	Year 1			Year 2		
	All Schools	Treatment	Control	All Schools	Treatment	Control
School Characteristics						
Urbanicity (percentage)						
City	43.8	41.7	45.8	43.8	41.7	45.8
Suburb	6.3	12.5	0.0	6.3	12.5	0.0
Town	14.5	12.5	16.7	14.5	12.5	16.7
Rural	35.4	33.3	37.5	35.4	33.3	37.5
Region (percentage)						
South	83.3	83.3	83.3	83.3	83.3	83.3
West	16.7	16.7	16.7	16.7	16.7	16.7
Title I Eligible (percentage)	83.3	83.3	83.3	87.5	87.5	87.5
Schoolwide Title I Eligible (percentage)	83.3	83.3	83.3	81.25	83.3	79.16
Total FTE classroom teachers	28.6	28.6	28.7	28.2	27.5	28.8
Student-teacher ratio	17.0	16.8	17.3	17.4	17.2	17.6
Total school enrollment	483	472	494	486	465	508
Student Demographics						
Students eligible for free/reduced-price lunch (percentage)	72.8	74.7	70.9	70.3	67.1	73.5
Student enrollment (average)						
Enrollment K/Grade 1	69	70	67	67.9	67.2	68.6
Enrollment Grade 3/Grade 4	67	66	68	67.0	65.8	68.3
Student gender (percentage)						
Female	48.1	47.8	48.3	48.4	48.2	48.6
Male	51.3	51.8	50.7	50.7	50.5	51.0
Age 1st day of school	6.5	6.5	6.5	7.6	7.6	7.6
Race/Ethnicity (percentage)						
Black	51.1	52.4	49.8	49.2	52.5	50.3
White	40.1	39.9	40.4	42.8	38.4	39.5
Hispanic	5.5	5.2	5.9	5.3	5.6	5.7
American Indian/Alaska Native	1.1	0.3	1.8	1.0	1.4	1.7
Asian	1.1	1.4	0.9	1.2	1.2	0.9
Hawaiian Native/Pacific Islander	0.1	0.1	0.1	0.2	0.2	0.2
Two or more races	0.2	0.2	0.2	0.3	*0.3*	0.2
School Sample Size	48.0	24.0	24.0	48.0	24.0	24.0

Sources: CCD Data 2011-2012 and 2012-2013, Fall evaluation data.

Table 2. Multilevel Model Estimates for Impact of Everyday Mathematics on Student Math Achievement

	Year 1			Year 2		
	Fixed Effects					
	Estimate		SE	Estimate		SE
Level-1 (student)						
Intercept	96.04	***	0.83	94.92		0.81
Race/Ethnicity (White)						
Black	-3.74	***	0.82	-3.27	***	0.83
Hispanic	-4.34	***	1.03	-3.16	**	1.00
Other	2.79	*	1.45	5.63	***	1.35
Free-Reduced Price Lunch	-3.67	***	0.60	-3.77	***	0.60
Level-2 (school)						
Treatment Effect	0.88		0.51	-1.01		0.53
Pretest Mean	0.82	***	0.06	1.15	***	0.08
District (Pike)						
Jackson	2.80	**	1.11	1.50		1.11
Los Alamos	-1.07		1.40	-1.81		1.41
Muskogee	3.49	*	1.51	1.83		1.59
Nye	1.92		1.18	2.60	*	1.18
Pointe Coupee	4.28	**	1.31	5.55	***	1.36
Rapides	0.20		1.00	-0.43		0.99
	Random Effects					
Variance Components	**Variance**			**Variance**		
School	1.05	*		1.18	*	
Student	195.39	***		197.57	***	
	Random Effects (from unconditional model[a])					
Variance Components	**Variance**		**ICC**	**Variance**		**ICC**
School	36.12	***	0.154	42.89	***	0.176
Student	199.11	***		201.34	***	

[a] The unconditional model is a two-level model with students (level-1) nested in schools (level-2) and only an intercept term on the right-hand side of the model.
Note: * p < 0.05, ** p < 0.01, *** p < 0.001

Table 3. Treatment Effects by Subgroup

	Year 1						Year 2					
	Estimate	**SE**	**Effect Size**		**Variance**	**SD**	**Estimate**	**SE**	**Effect Size**		**Variance**	**SD**
Subgroup: Grade												
Treatment	0.12	0.66		School	1.06	0.65	-2.35 ***	0.67		School	1.20	0.69
High Grade	0.25	0.59		Student	195.00	4.12	-2.15 ***	0.59		Residual	196.96	4.14
Treatment * High Grade	1.49	0.83	0.24				2.70 ***	0.83	0.39			
Subgroup: Sex												
Treatment	0.93	0.65		School	1.04	0.65	-0.60	0.66		School	1.10	0.67
Female	2.19	0.59		Student	194.24	4.11	2.13 ***	0.59		Residual	196.87	4.14
Treatment * Female	-0.08	0.83	-0.01				-0.84	0.83	-0.12			
Subgroup: Race – Black												
Treatment	1.27	0.72		School	0.99	0.64	-1.07	0.75		School	1.18	0.69
Black	-3.34	0.99		Student	194.25	4.11	-3.33 ***	1.02		Residual	197.57	4.16
Treatment * Black	-0.73	1.00	-0.12				0.11	1.01	0.02			
Subgroup: Race – Hispanic												
Treatment	0.80	0.52		School	1.05	0.65	-1.04	0.54		School	1.21	0.70
Hispanic	-5.19	1.39		Student	194.21	4.11	-3.36 *	1.39		Residual	197.56	4.16
Treatment * Hispanic	1.56	1.89	0.25				0.38	1.85	0.05			
Subgroup: FRPL												
Treatment	2.92	0.89		School	1.12	0.67	-0.74	0.90		School	1.24	0.72
FRPL	-2.41	0.75		Student	195.00	4.13	-3.59 ***	0.76		Residual	197.53	4.16
Treatment * FRPL	-2.90	1.03					-0.38	1.02	-0.05			
Subgroup: Pretest Continuous												
Treatment	5.21	7.33		School	1.03	0.64	10.46	9.78		School	1.09	0.67
Pretest	0.84	0.07		Student	194.23	4.11	1.20 ***	0.09		Residual	197.57	4.16
Treatment * Pretest	-0.05	0.08	-0.01				-0.13	0.11	-0.02			