Abstract Title Page

Title: From Efficacy Trial to Large Scale Effectiveness Trial: A Tier 2 Mathematics Intervention for First Graders with Difficulties in Mathematics.

Authors and Affiliations:

Eric Rolfhus
Russell Gersten
Instructional Research Group

Ben Clarke
University of Oregon

Lauren E. Decker
Chuck Wilkins
Edvance Research, Inc.

Joseph Dimino
Instructional Research Group

Abstract Body

Background / Context: Large scale longitudinal research (Morgan, Farkas, & Wu, 2009) and a meta-analysis (Duncan et al., 2007) have found that early mathematics achievement is a strong predictor of later mathematics achievement. In fact, end of Kindergarten and end of grade 1 mathematics achievement on ECLS–K and similar mathematics proficiency measures tends to be a stronger predictor than early measures of reading or reading-related skills such as phonemic awareness.

Yet, despite increasing interest, there is little research on the effectiveness of recommended best practices in Response to Intervention Models (RtI) in mathematics (Gersten et al. 2009). A recent literature review of grades K–3 mathematics interventions suitable for use in Tier 2 revealed just nine relevant studies (Newman-Gonchar, Clarke, and Gersten 2009), with just one that was a rigorous evaluation of an intervention and that used a randomized controlled trial (RCT) design (Fuchs et al. 2005). The Fuchs et al. (2005) study examined the impact of *Number Rockets*, a small-group intervention for grade 1 students at risk for mathematics difficulties and found statistically significant positive effects on several measures of mathematics proficiency. However, that study was an efficacy trial (one implemented under ideal conditions), involved considerable monitoring and support for interventionists, and was conducted in only a single district.

Purpose / Objective / Research Question / Focus of Study: This study (Rolfhus, et al. 2012) replicates the Fuchs et al. (2005) study as the first large-scale effectiveness trial (one intended to approximate real-world implementation) of *Number Rockets*. Key differences between the studies, that represent adaptations to implementation to facilitate scale-up are detailed in Table 1.While the Fuchs et al. study used interventionists experienced with at-risk students, the current study employed interventionists with a range of experience who were selected from the local community. While the Fuchs et al. study provided interventionists with substantial monitoring and support, the current study provided professional development and a support program similar to those provided by publishers of curriculum products (Agodini et al. 2009). Different measures were used to identify at-risk students, and measure outcomes. Finally, the district in the Fuchs et al. study used just one curriculum; each of the four urban districts in the current study used a different one, which may have provided a more heterogeneous instructional context.

The current study addresses the following confirmatory research question:

- Do grade 1 students at risk in mathematics who participate in the intervention perform better than at-risk control students on the Test of Early Mathematics Ability–Third Edition (TEMA–3; Ginsburg and Baroody 2003)?

The study also investigated three exploratory research questions:

- Does the intervention have a differential impact on grade 1 students at risk in mathematics, based on baseline mathematics proficiency?
- Do grade 1 students who participate in the intervention score differently than control students on the Woodcock-Johnson—Third Edition Letter/Word (WJ–III Letter/Word; Woodcock, McGrew, and Mather 2001) subtest?
- Is there a relationship between the level of implementation of the intervention, as measured by the average number of sessions, and the effect of the intervention within each matched-school pair on student TEMA–3 performance?

(please insert Table 1 here)

Setting: The study was conducted in 76 elementary schools in 4 urban school districts, one from each of four states in the south-central United States. Schools had student free-reduced lunch eligibility of 40% or higher, and at least three first grade teachers per school.

Population / Participants / Subjects: The sample consisted of *n*=2719 students who were initially screened for the study, and *n*=994 students identified as at-risk because they performed on the lowest 35% study-wide on the screening assessment (*n*=615 treatment, *n*=379 control). The at-risk student sample was 48.5% female; the sample was 44% Black, 46% Hispanic, 8% White, while the remaining students (≈2%) were from other several other race/ethnicity categories. Approximately 34% of students were eligible for free/reduced lunch and approximately 8% had an IEP. See Table 2 for more detail. (please insert Table 2 here)

Intervention / Program / Practice: *Number Rockets* was implemented as a supplemental intervention with a group of two or three students, typically meeting three or four times per week from November to May. The program covers 17 topics, each divided into three to six lessons, not all of which are required. If the entire group of students met the mastery criteria for a topic during a required lesson, the additional lessons for the topic were skipped. In this study, students received an average of 45 lessons, one each day the student groups met.

In this study, the intervention was delivered during regular school hours in pullout sessions conducted by part-time interventionists. Students meet with the interventionist, usually around a small table, for about 40 minutes per session. The interventionist followed instructions and read text aloud from a lesson script that includes highly prescribed feedback and prompting procedures to use with students as they perform various individual and group activities. For the last 10 minutes of the session, the interventionist worked with the students on mathematics fact practice using flashcards. The interventionist prepared a deck of flashcards for each student prior to each lesson based on the student's current skill level with addition and subtraction facts.[*] The difficulty of the flashcards increased with the skill of individual student and was independent of the group progression through the lessons.

Throughout the lesson, interventionists also used a behavior management system, representing an established protocol of interventionist behaviors intended to maintain student attention on, or redirect student attention to, *Number Rockets* tasks. The system used positive reinforcement practices, rewarding points for both accomplishment and reaching mastery criteria and at various randomly selected points during the lesson, for student engagement (defined as "listening carefully, working hard, and following directions"). When a student accumulated a predetermined number of points, she or he received a small reward. Most students earned a reward approximately every two sessions.

Research Design: Schools were randomly assigned to a condition using a matched-pair design, which increased the probability of baseline equivalence of schools—and the targeted at-risk students within those schools—in both conditions. This option was chosen primarily because a Tier 2 intervention would typically be implemented at the school level. Schools were matched within district on a composite score calculated from a mean school mathematics achievement score and the percentage of students receiving free or reduced-price lunch (FRPL). One school in

[*] Examples of addition and subtraction facts included on the flashcards are 1+1, 2+1, 3+1, or 5-0, 4-0, 3-0.

each pair was then randomly assigned to the intervention condition and the other to the control condition.

At each school, students whose parents signed a consent form were screened using an individually administered screener. A simple composite was computed based on the average *z*-scores across the screener's six subtests. Next, a cutscore for identifying students at risk for mathematics difficulties that corresponded to the 35th percentile of sample students was determined; the 35th percentile cutscore was used because it is consistent with others in the literature.

Data Collection and Analysis: Data collection was conducted from October 2008 to May 2009. Student Screener. Prior to the intervention, the student screener was administered by trained study staff in October and November, 2008. The screener consisted of six individually administered subtests, given before the intervention to determine student eligibility. Three were used in the Fuchs et al. (2005) study: *Curriculum-Based Measurement–Computation* (Fuchs, Hamlett, and Fuchs 1990)*, First-Grade Concepts/Applications* (Fuchs, Hamlett, and Fuchs 1990) and a revised version of *Story Problems* (Jordan and Hanich, 2007). Three others were selected from recent research on valid screening measures in mathematics for grade 1 students: *The Number Knowledge Test* (Baker et al. 2006), *Quantity Discrimination* (Clarke et al. 2006), and *Digit-Span Backward* (Geary 1993).

Fidelity Measures. Three types of fidelity measures were collected during implementation. *Lesson fidelity checklists* were coded from audio recordings of tutoring sessions. *Instructional logs* were used to track administrative information about each tutoring group session. All lessons were captured by interventionists on audio and session logs kept from December 2008 through May 2009, though coded by study staff in June 2009. In addition to these two fidelity of implementation measures, *Classroom instruction checklists* were used to measure the fidelity of schools' adherence to the developer's instruction to use *Number Rockets* as a strictly supplemental mathematics program. This was collected by interventionists in May 2009.

Student Outcome. The Test of Early Mathematics Ability–Third Edition (TEMA−3; Ginsburg and Baroody 2003), an individually administered mathematics test, was used as the primary outcome measure. Student outcomes were collected in May 2009 by trained study staff, after the intervention. The TEMA–3 assesses a broader set of mathematics skills than those represented in the pretest screener measures. Given that state mathematics assessments were not available as grade 1 outcome measures, the TEMA–3 was selected because, as an individually administered test, it was appropriate for the grade level of the students and measured mathematics achievement broadly, as state accountability measures do. The TEMA-3 test measures both formal and informal mathematics skills, and is designed to be consistent with typical grade level curricula taught in schools (Ginsburg and Baroody 2003). The reliability of the measure is reported (alpha = 0.95; test-retest = 0.82–0.93), and norms are based on a sample weighted to be nationally representative and scaled to a mean of 100 and a standard deviation of 15. Test administration takes about 30 minutes.

Analytic Model. Impacts of the intervention were tested with a three-level HLM model; students nested within schools, nested within school-pairs. Multiple imputation was used when attrition led to missing post-test scores. There were no differences between the experimental groups at baseline on student demographics or screener scores (see Table 3); therefore, no covariates were included in the model.

Findings / Results: The main finding of this replication is that students at risk for difficulties in grade 1mathematics benefited by participating in *Number Rockets*. Participation had a statistically significant difference *(g*= 0.34) on the TEMA–3 scores favoring the intervention group over the control group ($p < .001$). See Table 3 for details. (please insert Tables 3 here)

The current study's effect size of 0.34 standard deviations is smaller than the effect sizes for all four outcome measures demonstrating statistically significant results in the Fuchs et al. (2005) study (statistically significant effect sizes ranged from 0.40–0.70). This was expected given the current study's emphasis on implementing *Number Rockets* in conditions more closely resembling what urban school districts experience in their day-to-day instructional environment when implementing interventions. The observed lower levels of fidelity of implementation (85% vs. 94.6% in Fuchs et al. 2005) are consistent with this expectation.

Three exploratory analyses were also conducted:

(1) Results indicate that there was no statistically significant interaction between baseline mathematics proficiency (from the screener) and the impact of *Number Rockets* (effect size = 0.08, $p = .564$). Therefore, *Number Rockets* had no statistically significant differential effect on TEMA–3 scores by baseline mathematics proficiency for the sample of at-risk grade 1 students participating in this study.

(2) The classroom instruction checklist was used to record a one-week sample of instructional activities missed by students while they were participating in the intervention. Various reading activities recorded on the classroom instruction checklist were combined, they accounted for up to 33.8 percent of the classroom activities missed by students. Participating in *Number Rockets* could have reduced the amount of classroom reading instruction received and, consequently, affected reading achievement. However there was no statistically significant relationship between participation in *Number Rockets* and performance on the WJ–III Letter/Word subtest (effect size = –0.01, $p = .913$).

(3) There was no statistically significant relationship between the school-average number of *Number Rockets* sessions delivered to *Number Rockets* tutoring groups in each intervention school and the school-pair level intervention effect on student TEMA–3 performance (effect size = 0.07, $p = .667$). Therefore, higher levels of implementation of *Number Rockets* were not associated with larger school-pair level impacts on TEMA-3 performance.

Conclusions: This study is the first effectiveness evaluation (and first replication) of *Number Rockets* and builds on the positive findings of the Fuchs et al. (2005) efficacy study. When implemented under more typical LEA conditions at scale, the intervention was still found to effective; however, impact estimates were lower than observed in Fuchs et al. (2005). This may be due to several possibilities, including: (1) the use of a more general measure of mathematics achievement (TEMA-3) in the current study, (2) less support and supervision provided to interventionists, (3) lower observed implementation levels, (4) the use of different screening tools to identify at-risk students, (5) different or more heterogenous student demographics, or (6) increased variability in the number of lessons delivered across schools and student groups. However, exploratory analyses found no relationship between implementation quality or lesson dosage and student outcomes.

Appendices

Appendix A. References

Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., Novak, T., Murphy, R., et al. (2009). *Achievement effects of four early elementary school math curricula: Findings from first graders in 39 schools*. (NCEE 2009-4052). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved June 8, 2010, from http://ies.ed.gov/ncee/pubs/20094052/pdf/20094052.pdf

Duncan, G., Dowsett, C., Claessens, A., Magnuson, K., Huston, A., Klebanov, et al. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1446.

Fuchs, L., Compton, D., Fuchs, D., Paulsen, K., Bryant, J., and Hamlett, C. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97,* 493–513.

Fuchs, L., Hamlett, C., and Fuchs, D. (1990). *Curriculum-based math computation and concepts/applications*.

Fuchs, L. S., Hamlett, C. L., & Powell, S. R. (2003). Fact fluency assessment. (Available from L. S. Fuchs, 328 Peabody, Vanderbilt University, Nashville, TN 37203).

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J., et al. (2009). *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools* (NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved June 8, 2010, from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Ginsburg, H., and Baroody, A. (2003). *Test of Early Mathematics Ability–Third Edition*. Austin, TX: Pro-Ed.

Jordan, N., and Hanich, L. (2000). Mathematical thinking in second grade children with different forms of LD. *Journal of Learning Disabilities, 23*(6), 567–578.

Morgan, P., Farkas, G., and Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities*, *42*(4), 306–321.

Newman-Gonchar, R., Clarke, B., and Gersten, R. (2009). *A summary of nine key studies: Multitier intervention and response to interventions for struggling students in mathematics.* Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Rolfhus, E., Gersten, R., Clarke, B., Decker, L., Wilkins, C., and Dimino, J. (2012). *An evaluation of Number Rockets: A Tier 2 intervention for grade 1 students at risk for difficulties in mathematics.* (NCEE 2012-4007). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Woodcock, R., McGrew, K., and Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

Appendix B. Tables and Figures

Table 1. Key differences between the Fuchs et al. (2005) study and the current study

Characteristic		*Fuchs et al. (2005)*	*Current study*
Study	Type	Efficacy trial	Effectiveness trial
	Design	Student-level random assignment within classroom	School-level random assignment, schools paired within district
Consent	Type	Active parent consent	Active parent consent
	Rate	89 percent consent granted [a]	70.6 percent consent granted for intervention students 52.5 percent consent granted for control students
Sample	Districts	One district	Four districts across four states
	Schools	10 urban public elementary schools	76 urban public elementary schools
		6 Title I schools, 4 non-Title I schools	73 Title 1 schools, 3 non-Title 1 schools
	Students	667 screened	2,719 screened
		139 identified as at-risk	994 identified as at-risk
		70 received the intervention	615 received the intervention
		69 served as controls	379 served as controls
Screening	Procedure	Two stages: (1) A 15-minute screener comprised of four mathematics tests (2) Response to classroom instruction measured by limited progress on weekly CBM[b] measures after 4 weeks Students rank-ordered by factor score	One stage: A 25-minute screener comprised of six mathematics tests Students rank-ordered by composite score
	At-risk rate	Lowest 21 percent of students screened	Lowest 35 percent of students screened
Teacher involvement		Trained the regular classroom teacher to administer weekly CBM measures to whole class Teachers provided with student progress monitoring reports and classroom instructional strategies by research team every 2 weeks	Only teachers in intervention schools knew students' at-risk status No progress monitoring data were collected Teachers received no information about students' progress

Characteristic		*Fuchs et al. (2005)*	*Current study*
Interventionists	Total number	12	86
	Qualifications	10 Master's-level graduate students 1 Ph.D. researcher 1 experienced interventionist	100 percent with a bachelor's degree (at minimum) Wide range in teaching experience: 6 months to 38 years Locally recruited from retired teachers and substitute teacher pool
	Training and coaching	One-day training followed by additional practice and two follow-up sessions[c] Weekly coaching sessions throughout the intervention[c]	One day (8 hour) training Two 2-hour follow-up trainings with question and sessions Questions submitted and answered via email or telephone, as received
Delivery of mathematics fact practice[d]		Mathematics fact practice delivered via a computer program titled Math Flash	Mathematics fact practice delivered via paper flash cards
Number of lessons		48 lessons delivered	45 lessons targeted for delivery
Primary outcome measure(s)		(1) First Grade Concepts/Applications[e] (2) CBM—Computation[e] (3) Addition Fact Fluency[f] (4) Subtraction Fact Fluency[f] (5) Woodcock-Johnson Third Edition (WJ–III) —Calculation[g] (6) WJ–III—Applied problems[g] (7) Story Problems[h]	Test of Early Mathematics Ability–Third Edition (TEMA–3; Ginsburg and Baroody 2003)

Note: CBM is Curriculum-Based Measurement.
a. Did not report consent rate by experimental condition; parents provided consent prior to random assignment.
b. CBM—Computation is a one-page set of 25 grade 1 computation items group-administered to all students weekly in the Fuchs et al. (2005) study for purposes of progress monitoring.
c. Training was provided for one day followed by additional practice over two weeks; a second training session, on how to deliver mathematics fact practice; a final review session prior to start of intervention; and weekly coaching meetings. The number of hours involved in the training sessions, the additional practice, and the weekly coaching meetings was not specified.
d. Due to the lack of available computers in study schools, the *Number Rockets* developers adapted Math Flash to a parallel paper format.
e. Fuchs, Hamlett, and Fuchs (1990).
f. Fuchs, Hamlett, and Powell (2003).
g.Woodcock, McGrew, and Mather (2001).
h. Jordan and Hanich (2000).
Source: Fuchs et al. 2005; authors' analysis of data collected October 2007–May 2009.

Table 2. Demographic characteristics and mean screener composite scores for students with TEMA–3 scores and for students missing TEMA–3 scores

	TEMA–3 score status			
	Complete	*Missing*	χ^2	p
Characteristic	*(*n = *881)*	*(*n = *113)*		
Sex				
Female	49.0	42.3	0.92	.338
Race/ethnicity[a]			3.63	.057
American Indian/Asian/Other	1.0	--[d]		
Black	44.4	40.7		
Hispanic	46.1	43.4		
White	8.5	--[d]		
FRPL				
Yes	34.9	31.0	0.67	.415
English language learner				
Yes	12.0	9.7	0.51	.476
IEP status				
Yes	8.1	7.1	0.13	.717
Screener composite	*Mean (SD)*[c]	*Mean (SD)*[c]		
	–0.86 (0.38)	–0.91 (0.40)	–1.45	.149

FRPL is free or reduced-price lunch program; IEP is Individualized Education Program; *p* is the probability level associated with the level of the χ^2 statistic.

Note: Demographic characteristics of the students for whom TEMA–3 scores were available are reported in percentages; all χ^2 results are Mantel-Haenszel Chi-Square.

a. Districts reported race/ethnicity in six categories: American Indian, Asian, Black, Hispanic, Other, and White. A multiracial category was not included, as districts did not report these data. Due to small sample sizes, the American Indian, Asian, and Other categories have been collapsed in this table. Unless otherwise noted, Black includes African American, Hispanic includes Latino, Asian includes Native Hawaiian or Other Pacific Islander, and American Indian includes Alaska Native. Percentages may not sum to 100 because of rounding.

b. TEMA–3 scores were missing for 11.4 percent of students in the analytic sample (113 of 994), including 9.8 percent of intervention students (60 of 615) and 14.0 percent of control students (53 of 379). Missing TEMA–3 scores were imputed as described in text. A two-tailed *z*- test of the difference in attrition proportions for each experimental group was conducted with alpha = 0.05 and was not statistically significant ($z = 1.933$; $p = .053$).

c. Screener composite standard deviations reported here are not adjusted for clustering.

d. These two cells suppressed because one of the cells represented 3 or fewer cases.

Source: Authors' analysis of study team data collected April 2009–May 2009.

Table 3. Impact of *Number Rockets* on mathematics achievement of grade 1 students as measured by the TEMA–3, by assigned condition

	Intervention (n = 615)		*Control (n = 379)*		*Estimated intent-to-treat impact*			
Outcome measure	*Mean*	*Standard deviation*	*Mean*	*Standard deviation*	*Estimated impact*	*Standard error*	p	*Effect size*[a]
TEMA–3[b]	88.32	(12.64)	84.04	(12.74)	4.28	0.82	< .001	0.34

Note: All statistics are based on the analysis of five multiply imputed datasets using a three-level hierarchical linear model, which accounts for clustering of data (students clustered within schools, which are in turn clustered within pairs of schools) and controls for baseline screener score. Means presented here are the unadjusted means for both groups.

a. Computed by dividing the estimated impact by the pooled within-group standard deviation of the TEMA–3.

b. Scores are scaled with a mean of 100 and a standard deviation of 15.

Source: TEMA–3 data collected April 2009–May 2009.