Examining the Efficacy of a Tier 2 Kindergarten Intervention

Ben Clarke
Christian Doabler
University of Oregon
Keith Smolkowski
Oregon Research Institute
Scott Baker
Hank Fien
Mari Strand Cary
University of Oregon

Technical Report #1104
Center on Teaching and Learning
University of Oregon
Introduction

The low level of mathematics performance of U.S. students in relation to national standards and in international comparisons has concerned educators and policy makers and garnered increasing attention in the past decade (National Mathematics Advisory Panel [NMAP], 2008; National Research Council [NRC], 2001; Schmidt, Houang, & Cogan, 2002). Results from the 2011 National Assessment for Educational Progress (NAEP) indicate that only 40% of 4th graders were deemed at or above proficient in mathematics and 18% were below basic (National Center for Education Statistics, 2011). Difficulties in mathematics achievement are particularly severe for students from low income and minority backgrounds and those with learning disabilities. Signs of long-term difficulty in mathematics appear early. Specifically, significant differences in student knowledge can be reliably assessed at school entry on measures ranging from counting principles and number knowledge to more complex understandings of quantities, operations, and problem solving (Griffin, Case, & Siegler, 1994; Jordan, Kaplan, & Locuniak, 2007). Longitudinal research suggests that students who perform poorly in mathematics at the end of kindergarten are likely to continue to struggle throughout elementary school (Bodovski & Farkas, 2007; Duncan et al., 2007; Hanich, Jordan, Kaplan, & Dick, 2001; Morgan, Farkas, & Wu, 2009). Using a nationally representative sample of students from the Early Childhood Longitudinal Study-Kindergarten Cohort (ECLS-K), Morgan et al. (2009) found that students who were in the lowest 10th percentile at entrance and exit from kindergarten (considered an indicator of a learning disability in mathematics – MLD) had a 70% chance of remaining in the lowest 10th percentile five years later. Their overall mathematics achievement in fifth grade remained
2 standard deviation units below students who did not demonstrate a MLD profile in kindergarten. These data as well as Duncan et al.’s (2007) recent meta-analysis of longitudinal research on the development of mathematics proficiency argue strongly for the importance of a successful start in mathematics. Unless these differences are addressed at entry into kindergarten, they are likely to persist and become more difficult to remediate over time as students exit kindergarten without a solid foundation upon which to build an increasing complex understanding of mathematics (Geary, 1993; Jordan, Kaplan, & Hanich, 2002; Lyon et al., 2001; Morgan et al., 2009).

Recognizing that mathematics trajectories are established early in school and kindergarten represents a critical transition from informal to formal mathematics, our research group developed and evaluated ROOTS, a 50-lesson (Tier 2) kindergarten math intervention. The purpose of this technical report is to summarize results from an initial randomized controlled trial (RCT) study on the efficacy and efficiency of ROOTS. In this report, we present an overview of the ROOTS curriculum, study design, methodology and results focusing on the impact of ROOTS on kindergarten student math outcomes. Our research team had previously developed and evaluated the *Early Learning in Mathematics* (ELM) kindergarten core curriculum (Davis & Jungjohann, 2009). ELM consists of 120 lessons and focuses on four key mathematics strands: (a) Number and Operations, (b) Geometry, (c) Measurement, and (d) Vocabulary. The first three of these map directly onto the three content domains contained in the National Council of Teachers of Mathematics Curriculum Focal Points (NCTM, 2006), and the fourth (vocabulary) is addressed in the NCTM Process Standards (2000). We tested the efficacy of ELM in a RCT, randomly assigning 66 kindergartens classrooms to treatment and
control conditions. Students in the treatment condition, ELM, outperformed their control classroom peers on two distal measures of math proficiency. ELM students significantly outperformed control students on both the Test of Early Mathematics Ability – 3rd Edition (TEMA; \( t = 2.41, p = .02 \)) and Early Numeracy Curriculum Based Measurement (EN-CBM; \( t = 1.99, p = .05 \)). Overall Hedges g effect sizes were .13 on the TEMA and .14 on EN-CBM (Clarke et al., 2011). In condition by risk status analyses, at-risk students (defined as performing below the 40th percentile on the TEMA at pretest) demonstrated the greatest treatment benefit (Clarke et al., 2011). At-risk treatment students significantly outperformed at-risk control students on both the TEMA (\( t = 3.29, p = .0017, g = .24 \)) and EN-CBM total score (\( t = 2.54, p = .0138, g = .22 \)).

The pattern of findings is important for two reasons. First, differential impact favoring the at-risk students was aligned with our theoretical framework. In developing ELM our objective was to create a core mathematics programs (i.e., Tier 1) that would address the needs of the average- and high-performing students (in the analysis, students above the 40th percentile at pretest performed the same at posttest in both ELM and control conditions) and substantially increase the mathematics achievement of students at risk for math difficulties (which occurred). Second, although the ELM curriculum was beneficial to at-risk students in particular, the performance of at-risk students at the end of the year did not match the performance of students not at risk (at pretest). In other words, although there was differential impact by risk status, the program did not fully eliminate the gap between at-risk and average-achieving students.

Spurred by our findings that while ELM helped reduce but not fully eliminate the achievement gap between at-risk students and their on track peers, we developed a Tier 2
intervention, ROOTS, to be used as a supplemental program with ELM. The ROOTS intervention was designed to focus exclusively on Number and Operations because an in-depth understanding of the whole number system is a critical step in achieving proficiency in more sophisticated mathematics, such as rational numbers and algebra (Gersten et al., 2009a; NCTM, 2006; NRC, 2001, 2009). For example, authors of the IES practice guide on effective mathematics instruction and intervention for at-risk students observed that “individuals knowledgeable in instruction and mathematics look for [and develop] interventions that focus on whole numbers extensively in kindergarten through grade 5” (Gersten et al., p. 18; 2009a). Concurring, the Common Core State Standards for Mathematics (CCSS, 2010) recommends extensive coverage of whole number concepts and skills during kindergarten (i.e., counting and cardinality, operations and algebraic thinking, and number and operations in base ten).

In the second study of the ELM efficacy trial (Baker et al., 2008), we explored the value added of the intervention program, ROOTS, to accelerate the learning of students at risk for MLD. All students randomly assigned to the treatment condition received core instruction in ELM and at-risk students received the combination of ELM and the ROOTS intervention. Instructional assistants delivered ROOTS to small groups of 4-5 at-risk students during the second half of the kindergarten school year. This technical report summarizes findings from that study.

**ROOTS Intervention Overview**

ROOTS is a 50-lesson kindergarten intervention program designed to develop procedural fluency with and conceptual understanding of whole number concepts. ROOTS is delivered by instructional assistants to small groups consisting of 4-5 students,
2 to 3 times per week, for 20 weeks during the second half of the school year. Each ROOTS lesson is approximately 20 minutes in duration and includes 4 to 5 brief math activities that center on whole number concepts and skills. ROOTS provides in-depth instruction in whole number concepts by linking the informal mathematics developed prior to kindergarten to the formal mathematics of kindergarten. Specifically, ROOTS focuses on three key areas of whole number understanding (a) Counting and Cardinality (b) Number Operations and (c) Base 10/Place Value. Students advance from objectives centered on counting objects and identifying numerals to applying their knowledge of foundational principles to number operations and the base 10 system. The specific focus on whole number aligns with calls for more focused and coherent curricula (NCTM, 2006; NMAP, 2008), and intervention programs designed to meet the needs of students at-risk for MLD (Gersten et al., 2009a). ROOTS Objectives and the alignment overview between ROOTS and the Common Core State Standards (CCSS, 2010) are listed in Appendix A (pp. A1-A10).

A central feature of the ROOTS program, established through a set of research-based instructional design principles (Doabler et al., in press), is an explicit and systematic approach to instruction. Mathematics intervention studies consistently demonstrate that students at-risk for MLD learn more from explicit math instruction compared to other approaches to instruction (Baker, Gersten, & Lee, 2002; Gersten, et al., 2009b; Haas, 2005; Kroesbergen & Luit, 2003). Explicit and systematic instruction is a method for teaching the “essential skills in the most effective and efficient manner possible” (Carnine, Silbert, Kame’enui, & Tarver, 2004; p. 5). This approach incorporates the instructional design principles that (a) facilitate overt and conspicuous
instructional interactions among teachers and students targeting key math content and (b) increase intervention intensity required to accelerate learning for kindergartners at-risk for MLD. For example, the lessons provide teachers with guidelines for modeling and demonstrating what they want students to learn, and providing specific academic feedback to students as they engage in learning activities. The program also provides students with frequent and structured opportunities to practice learning key mathematics concepts and content.

**Methodology**

**Design**

A randomized controlled trial was the research design used. Math achievement data were collected from individual students, and random assignment and instructional delivery took place at the classroom level. Classrooms were randomly assigned to treatment or control conditions, blocking on school. Our primary analysis framework is a group-randomized trial (Murray, 1998) with students nested within classrooms and classrooms nested within condition. Fourteen classrooms were in the treatment condition and 15 classrooms were in the control condition.

**Participants**

Full day kindergarten classrooms who participated in the ELM study were randomly assigned to treatment or control conditions, blocking on teachers’ ELM experience (one year or none) and school. Fourteen classrooms were in the treatment condition (ELM+Roots) and 15 classrooms were in the control condition (ELM-only). Time was controlled so that treatment and control classrooms provided the same amount of daily mathematics instruction. This was accomplished by delivering the Roots
instruction during the individual, worksheet-based “math practice” portion of ELM in treatment classrooms. Classroom teachers (treatment and control) provided whole class ELM instruction throughout the year. Roots instruction began in January and was provided by trained instructional assistants.

**Instructional Assistants.** Fourteen instructional assistants (IAs) participated in the study, of which thirteen were female and all identified themselves as White. Three of the IAs had college degrees and two held current teacher certifications. In this sample, nine of the IAs had four or more years experience as an instructional assistant and all but four of the IAs had completed college level coursework in mathematics.

**Students.** A list of students by classroom who had scored below the 40th percentile on the TEMA pre-test was provided to participating teachers (n = 29) who were asked to select 5 students from the list who they believed would benefit from a small group math intervention. In treatment classrooms (ELM+Roots), these “Roots students” (n=69) received all of the whole-class ELM instruction, however, three days per week, instead of practicing that day’s ELM topics independently (i.e. math practice worksheets), they received Roots instruction. Nominated students in control classrooms (n = 74) served as “matched pairs.” Given that Roots was not offered in their classrooms, they participated only in whole class ELM instruction, five days per week.

**Fidelity of Implementation**

**Professional development.** Participating IAs attended three PD workshops focused on the ROOTS curriculum. The initial PD workshop focused on the instructional objectives related to Lessons 1-25, the critical content of kindergarten mathematics, small-group management techniques, and the instructional practices that have been
empirically validated to increase student math achievement (e.g., teacher provided academic feedback). In the second and third workshops the same format was followed as in workshop 1 but with a focus on the second half of the ROOTS curriculum, Lessons 26-50. Workshops were 4 hours in length and organized around three principles: (a) active participation (b) content focused, and (c) coherence. Although the empirical literature on PD in math instruction is thin, a burgeoning base of evidence indicates these principles can lead to improved outcomes for teachers and students (Blank & de las Alas, 2009; Fixsen et al., 2005; Joyce & Showers, 2002; Scher & O’Reilly, 2009). On at least three occasions, IAs also received onsite coaching from two expert teachers. Some IAs received more than three coaching visits if the IA or coach felt more support was warranted (e.g., when there were particularly pervasive student behavior problems or the IA struggled with lesson implementation).

**Fidelity.** Online logs completed by the 14 instructional assistants (IA) who delivered the ROOTS intervention revealed that groups generally completed all 50 ROOTS lessons during the year. Research staff observed ROOTS instruction 1-3 times over the course of the study and rated fidelity of implementation using a 3-point rating scale where 3 represented “full implementation,” 2 represented “partial implementation,” and 1 was “not taught.” IAs demonstrated high fidelity scores ($M = 2.92, SD = 0.06$) for the activities prescribed. Surveys were completed by the 14 ROOTS IAs as well as the 14 classroom teachers who provided whole class instruction to ROOTS students. IAs (100%) and teachers (87.5%) reported that it was possible to provide 20 minutes of ROOTS instruction on a regular basis. Using a 7-point Likert scale where 1 represented “greatly benefited,” 4 represented “moderately benefited,” and 7 represented “did not
benefit.” IAs and classroom teachers reported that ROOTS students benefited significantly from ROOTS instruction (respectively, $M = 2.0, SD = 1.15$; $M = 2.0, SD = 1.19$).

**Student Measures**

**Test of Early Mathematics Ability-Third Edition (TEMA-3; Pro Ed, 2007).** The TEMA-3 is a norm-referenced individually-administered measure of early mathematics for children ages 3 to 8 years 11 months. The TEMA-3 is designed to identify student strengths and weaknesses in specific areas of mathematics. The TEMA-3 measures both formal mathematics and informal mathematics including skills related to counting, number facts and calculations, and related mathematical concepts. Test authors report alternate-form reliability of .97 and test-retest reliability ranges from .82 to .93. Concurrent validity with other criterion measures of mathematics is reported as ranging from .54 to .91.

**Early Numeracy-Curriculum-Based Measurement measures (EN-CBM; Clarke & Shinn, 2004).** EN-CBM is a set of four measures based on principles of curriculum-based measurement (Shinn, 1989). For our purposes, these measures map onto the ROOTS objective of developing procedural fluency. Each 1-minute fluency-based measure assesses an important aspect of early numeracy development including magnitude comparisons and strategic counting. The EN-CBM measures have been validated for use with kindergarten students (Chard et al., 2005; Clarke et al., 2008). **Oral Counting (OC).** The OC measure requires students to orally rote count as high as possible without making an error. Concurrent and predictive validity range from 46 to .72 with other published standardized measures. **Number Identification Measure (NI).** The NI
measure requires students to orally identify numbers between 0 and 10 when presented with a set of printed number symbols. Concurrent and predictive validity range from .62 to .65 with other published standardized measures. **Quantity Discrimination Measure (QD).** The QD measure requires students to name which of two visually presented numbers between 0 and 10 is greater. Concurrent and predictive validities range from .64 to .72. **Missing Number Measure (MN).** The MN measure requires students to name the missing number from a string of numbers (0-10). Students are given strings of three numbers with the first, middle, or last number of the string missing. Concurrent and predictive validities range from .46 to .63.

**Data Collection**

All measures were individually administered to students. Trained staff with extensive experience in collecting educational data for research projects administered all student measures. All data collectors were required to obtain inter-rater reliability coefficients of .90 prior to collecting data with students. Follow-up trainings were conducted prior to each data collection period to ensure continued reliable data collection.

**Analysis**

We assessed intervention effects on each of the primary outcomes with a mixed model (multilevel) time by condition analysis (Murray, 1998). This tests differences between conditions on change in outcomes from the fall of kindergarten (T1) to the spring (T2). The specific model tests time coded 0 at T1 and 1 at T2, condition coded 0 for control and 1 for ROOTS, and the interaction between the two. With 29 schools, tests of time by condition used 27 degrees of freedom. The analyses included students
identified as at risk for math difficulties across both conditions. The analysis included all available data—whether or not students’ scores were present at both time points—to estimate differences between assessment times and between conditions, which minimizes the potential for bias due to missing data. The nested time by condition analysis accounts the intraclass correlation associated with multiple students nested within the same schools. As a test of net differences, it also provides an unbiased and straightforward interpretation of the results (Cribbie & Jamieson, 2000, Fitzmaurice, Laird, & Ware, 2004).

Model estimation. We fit models to our data with SAS PROC MIXED version 9.1 (SAS Institute, 2009) using restricted maximum likelihood (REML), generally recommended for multilevel models (Hox, 2002). From each model, we estimated fixed effects and variance components. Maximum likelihood estimation for the time by condition analysis allows the use of all available data and provides potentially less biased results even in the face of substantial attrition, provided the missing data were missing at random (Schafer & Graham, 2002). In the present study, we did not believe that attrition or other missing data represented a meaningful departure from the missing at random assumption, meaning that missing data did not likely depend on unobserved determinants of the outcomes of interest (Little & Rubin, 2002).

The models assume independent and normally distributed observations. We addressed the first, more important assumption (van Belle, 2008) by explicitly modeling the multilevel nature of the data. Regression methods have also been found quite robust to violations of normality and outliers have a limited influence on the results in a variety of multilevel modeling scenarios (Bloom, Bos, & Lee, 1999; Donner & Klar, 1996;
Fitzmaurice et al., 2004; Maas & Hox, 2004; 2004b; Murray et al., 2006). This feature of multilevel models also eases concerns about the use of different scoring methods used for different measures in the analyses (e.g., raw scores, scaled scores, standard scores).

**Effect sizes.** To ease interpretation of effects we computed an effect size, Hedges’ $g$ (Hedges, 1981), for each fixed effect. Hedges’ $g$ represents an individual-level effect size comparable to Cohen’s $d$ (Rosenthal, Rosnow, & Rubin, 2000), except that Cohen’s $d$ uses the sample standard deviation while Hedges’ $g$ uses the population standard deviation (Rosenthal & Rosnow, 2007).

**Results**

Table 1 presents means, standard deviations, and sample sizes for the TEMA-3 and CBM outcomes by assessment time and condition. We addressed our research hypotheses with the nested time by condition model defined above, focusing on tests of gains by condition. The results indicated that students in the ROOTS condition outperformed students in the control condition on the TEMA scaled score ($t = 2.193, df = 27, p = .0371$). We did not find differences on the EN-CBM measure ($t = 1.354, df = 27, p = .1870$). One significant difference was found on the EN-CBM Quantity discrimination measure ($t = 2.18, df = 27, p = .038$). Hedges’s $g$ effect sizes were .37 (TEMA scaled score), .30 (EN-CBM) and .41 (Quantity Discrimination).
Table 1. Means, standard deviations and sample sizes pre and post treatment.

<table>
<thead>
<tr>
<th></th>
<th>T1</th>
<th>ROOTS</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TEMA Scaled</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>67.8 (9.6)</td>
<td>70.3 (9.9)</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>85.3 (9.5)</td>
<td>84.2 (12.1)</td>
<td></td>
</tr>
<tr>
<td><strong>CBM Total</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>14.9 (15.8)</td>
<td>21.4 (22.5)</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>106.3 (39.0)</td>
<td>100.0 (40.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Quantity Discrimination</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>1.4 (3.5)</td>
<td>1.8 (3.7)</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>19.8 (9.4)</td>
<td>16.4 (8.8)</td>
<td></td>
</tr>
<tr>
<td><strong>TEMA sample size</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>56</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>63</td>
<td>62</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Standard deviations (SDs) presented in parentheses. Sample sizes for CBM and TEMA data differed by one to three subjects per cell—minimum sample sizes shown.
References


