Conducting a Cost Effectiveness Analysis of an Early Numeracy Intervention

Authors

Ben Clarke, Gulcan Cil, Keith Smolkowski, Marah Sutherland, Jessica Turtura, Christian T. Doabler, Hank Fien, Scott K. Baker

Publication History

School Psychology Review, Received 09 Apr 2020, Accepted 20 Apr 2020, Published online: 31 Dec 2020

Full citation

Clarke, B., Cil, G., Smolkowski, K., Sutherland, M., Turtura, J., Doabler, C. T., Fien, H., & Baker, S. K. (2020). Conducting a cost effectiveness analysis of an early numeracy intervention. *School Psychology Review, 40*(4), 359–373. doi: https://doi.org/10.1080/2372966X.2020.1761236

Funding

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through **Grant number R324A120304** to the Center on Teaching and Learning. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Conducting a Cost Effectiveness Analysis of an Early Numeracy Intervention

Abstract

This study examined the cost effectiveness of a 50 lesson mathematics intervention program focused on whole number concepts for at-risk kindergarten students, ROOTS. The study utilized a randomized block design with 1,251 at-risk students within 138 classrooms randomly assigned to one of two active treatment conditions (small groups of either two or five students) or control condition. Proximal and distal measures were collected in the fall (pretest) and spring (posttest). The costs per group per effect-size unit change varied from $216 to $736 depending on differing district scenarios and group size. The cost-effectiveness ratios per student varied from $267 to $3,201 depending on district scenario, group size, and the measure. Implications for conducting cost effectiveness evaluations and public policy are discussed.

Keywords: Mathematics, Learning Disabilities, Cost Effectiveness

Examining the Cost Effectiveness of a Kindergarten Mathematics: Implications for Practice and Policy

Increasingly, cost analyses and cost-effectiveness (CE) analyses are called for as a critical aspect of evaluating educational programs. Recent federal requests for grant proposals include specific directions to include both a cost analysis and a CE analysis in order to be considered for funding (IES RFP; US Dept. of Education, 2018). Although the research base is available to guide spending decisions toward programs and interventions that are effective, often little is known about the economic costs of implementing these programs that are deemed effective. This is especially important when the decision involves a choice between programs targeting similar outcomes. With CE analyses which provide a means of estimating the costs of two or more alternatives relative to the effectiveness of each alternative in producing a common outcome, the decision maker can make an informed decision and choose the alternative with the lowest cost for any given level of educational effectiveness (Levin and McEwan, 2002). The purpose of this manuscript is to conduct a CE analyses of a kindergarten mathematics intervention.

Levin and Belfield (2015) argue that, considering the substantial cost of the education sector, there needs to be greater scrutiny on how societal resources can be used more efficiently in education. In addition, Levin and Belfield note, that although rigorous CE analyses have been widely used in many government sectors including the military, health, transportation, and criminal justice since the 1960s, the education sector has been slow in adopting these analyses, and even when CE analyses have been conducted, the evaluations have been far from rigorous. Despite that overall concern, there are examples of rigorous CE analyses of early interventions For example, Borman, Hewes, Overman, and Brown (2002) performed a cost analysis on the early childhood development program Success for All and compared the program's effectiveness-cost ratios for reading and math achievement with that of three other model early interventions, namely, Perry Preschool Program, Abecedarian Project, and Tennessee STAR class-size reduction effort. The authors argued that their findings suggested that Success for All program is deserving of similar recognition as the other three interventions with a superior effect per $1,000 ratio for reading, and a comparable ratio for math achievement. In another study, Hollands et al. (2016) compared two early reading interventions, Corrective Reading and Wilson Reading System, calculating cost-effectiveness ratios for these programs using existing effectiveness data on common outcome measures. Although both programs have similar effect sizes (0.22 and 0.33), the findings of their CE analysis revealed that the program with the slightly smaller effect size is also the one that is more costly, making it a potentially less attractive option for the particular outcome in question.

In addition to facilitating comparisons between distinct programs targeting similar outcomes, CE analyses can also inform decision making about the implementation approach. For example, Knight et al. (2016) examined the cost-effectiveness of alternative approaches to enhancing Head Start, namely, instructional coaching for preschool teachers, family coaching for the parents, or a combination of the two. They found that the family coaching only was the most cost-effective option In another study, Hunter, DiPerna, Hart, and Crowley (2018) examined the cost effectiveness of a universal social emotional learning program, the Social Skills Improvement System – Class-wide Intervention Program (SSIS-CIP). They compared the cost-effectiveness across first and second grade classrooms and found that implementation in second grade was the more cost-effective option for SSIS-CIP. The authors note that, although this information is potentially useful in decisions regarding implementation of this program, richer interpretations of their cost-effectiveness findings and comparisons with other programs would be possible if CE analyses for other universal social emotional learning programs were available in the literature.

These studies exemplify the value of conducting CE analyses to enable more fine-tuned decision making when educators, schools, districts, or states are considering how best to allocate scarce resources. However, there were no CE analyses focused exclusively in the area of early mathematics. Yet the importance of early mathematics is gaining increased attention by educators and policy makers (Frye et al., 2013). Over two decades ago, calls were made for advancing the field's understanding of mathematics development and the corresponding instruction provided to our nation's students in mathematics (National Research Council, 2001). Particular attention was paid to the transition period between the informal mathematics encountered outside of school environments and the formal mathematics students are expected to learn when they enter school (Gersten & Chard, 1999). The call for focusing on this transition time was reinforced by emerging longitudinal studies documenting the relationship between early and later mathematics understanding (Duncan et al., 2007) with findings indicating that mathematics difficulty was relatively stable (Morgan, Farkas, & Wu, 2009) , that early risk in mathematics was highly predictive of long term mathematics difficulty (MD; Morgan, Hillemeier, Farkas, & Maczuga, 2014) , and that gaps between on-track and at-risk learners grew significantly over time (Morgan et al., 2019). Despite those troubling findings and the relatively low and stagnating performance of students across the nation (National Assessment of Educational Progress, 2019) some positive trends were also noted. Morgan and colleagues (2014) documented that during kindergarten long term trajectories could be altered while other research found that growth in mathematics across the informal to formal transition period was a stronger predictor of high school mathematics than initial skill at the start of the transition period (Watts, Duncan, Siegler, & Davis-Kean, 2014).

Given the importance of this time period, researchers developed and tested the efficacy of intervention programs targeting number sense and whole number understanding (e.g. Clarke, Doabler, Smolkowski, Baker, et al., 2016; Sood & Jitendra, 2013). For example, Dyson and colleagues (2013) developed a 24 lesson program targeting number sense and key concepts of whole number understanding including counting and cardinality, number magnitude, composing and decomposing numbers, addition and subtraction principles and strategies, and initial base 10 understanding. Lessons were approximately thirty minutes and utilized a systematic and explicit instructional design architecture (Archer & Hughes, 2011). Lessons consisted of multiple activities to build conceptual understanding and procedural fluency with tailored review on key concepts integrated throughout the lesson structure. Results indicated generally positive impact on two of three measures at post-test and one of three at delayed post-test (eight weeks after the end of the intervention). Similar results have been found for early mathematics interventions targeting whole number understanding (e.g. Clarke et al., 2014; Fuchs et al., 2005; Gersten et al., 2012) leading to a small but emerging research base.

As the field of mathematics intervention work has advanced, increased attention has been paid to gaining a greater understanding of more nuanced aspects of research (Ochsendorf, 2016). Important considerations for this second wave of research on mathematics interventions include understanding for whom and under what conditions interventions work (Miller, Vaughn, & Freund, 2014) and an increased recognition on the importance of replicating results across geographically and demographical diverse samples (Coyne, Cook, & Therrien, 2016). However, not as much focus has been on considering how programs fit within current multi-tier service deliver models, including Response to Intervention (RTI) and Multi-Tier Systems of Support (MTSS). Although there are consistent challenges in implementing RTI and MTSS models, they have been long called for as a mechanism to increase at-risk student outcomes (Vaughn & Fuchs, 2003) and are relatively prominent in some iteration (Balu et al., 2015). A defining feature of RTI and MTSS is the increasing intensity of services provided to students at each successive tier (NASDE, 2006). Yet little research has been conducted to date, that examines questions related to treatment intensity within these service delivery models (Codding et al., 2016) .

One mechanism to potentially increase intensity is to provide instruction in small groups with group size decreasing as the needs of students increase (Gersten et al., 2009). Investigations related to group size are limited and with mixed findings. In a series of group design randomized control trials, Vaughn and colleagues (Elbaum, Vaughn, Hughes, Moody, & Schumm, 2000; Vaughn & Fuchs, 2003) found generally stronger results for smaller small groups or individualized instruction in the area of elementary reading. However, group size was not directly manipulated. One study in which group size was directly manipulated contrasted groups with teacher student ratios of 1:1, 1:3, and 1:10. Significant differences on student outcomes were found between the two smaller groups and the larger group but no differences were found between the 1:1 and 1:3 teacher student ratio small groups (Vaughn et al., 2011). Similarly, Begeny and colleagues (Begeny, Hawkins, Krouse, & Laugle, 2011; Begeny, Levy, & Field, 2018; Begeny, Yeager, & Martínez, 2012; Ross & Begeny, 2015) using single case design methodology studied the effects of small group versus individual intervention size on the oral reading fluency of students who were struggling with reading and found that individualized intervention is not necessarily more effective than small group intervention. In a summary of existing research on interventions that target reading fluency, Begeny and colleagues found that the majority of participants improved as a result of receiving small group intervention, but that of the studies examining comparable small group and one-on-one interventions, 79% of students performed equally well from both interventions (Begeny et al., 2018).

Our research group conducted a large scale randomized control trial to investigate the impact of a kindergarten mathematics intervention, ROOTS, and to examine whether that impact varied by group size (Clarke, Gersten, Dimino, & Rolfhus, 2011). A review of the literature indicated no other studies in the area of mathematics interventions that included a systematic manipulation of group size. Implementation of the ROOTS intervention occurred across three school years (2012-2015) with four cohorts of students at two different research sites: Oregon and Massachusetts. A partially nested randomized controlled trial was employed (Baldwin, Bauer, Stice, & Rohde, 2011), randomly assigning kindergarten students within classrooms to one of three conditions: (2:1 ROOTS group, 5:1 ROOTS group, and a no-treatment control condition). Previous studies presented efficacy results for specific cohorts (Clarke, Doabler, Smolkowski, Kurtz Nelson, et al., 2016; Doabler et al., 2016), differential outcomes by small group (Clarke et al., 2017; Doabler, Clarke, Kosty, Kurtz-Nelson, et al., 2019), interaction by initial skill (Clarke et al., 2019) and English Language Learners (Doabler, Clarke, Kosty, Smolkowski, et al., 2019). Overall study results from Clarke et al. (in press) are summarized in Table1. Additional detail on the effect sizes used in this manuscript are found in the methods section.

The design of study allows for evaluating cost effectiveness within the framework of treatment versus control but also to examine variations of treatment (i.e. group size) and how those costs are associated with student outcomes. Given the growing but still relatively limited amount of research in early numeracy and mathematics intervention programs, the manuscript will offer insight into the costs needed to improve student outcomes in an area of critical importance for the field (Frye et al., 2013) and, perhaps of greater importance, the contrast of treatments options (i.e. comparing two group sizes) will offer insight into questions related to treatment intensity. Since RTI and MTSS models rest upon the idea of increasing intensity as students' progress through tiers and that decreasing of group size is hypothesized and utilized as a mechanism to increase intensity, the CE analysis will enable examining the increased costs of a smaller small group relative to student outcomes. Findings from the work have potential implications for policy makers, state and district leaders, and schools designing RTI and MTSS service delivery systems.

Method

This study presents the results of CE analyses on data collected during the federally funded ROOTS Efficacy Project (Clarke, Doabler, Fien, Baker, & Smolkowski, 2012). For a more detailed overview of study procedures see Clarke et al. (in press).

Participants

Twenty-three schools from four Oregon school districts and two Massachusetts school districts participated. The two Massachusetts districts were located in close proximity to Boston. Three of the Oregon districts were located in rural and suburban areas of western Oregon, while one district was located near Portland. A total of 138 kindergarten classrooms participated in the study, with the majority (57%) providing half-day kindergarten programs. The 138 classrooms were taught by 75 certified kindergarten teachers, of which 48 teachers participated for two consecutive years. All students with parental consent from the 138 classrooms were screened in the late fall of their kindergarten year. The screening process included the Assessing Student Proficiency in Early Number Sense (ASPENS; Clarke et al., 2011) and the Number Sense Brief (NSB; Jordan, Glutting, & Ramineni, 2010). Students were eligible for the ROOTS intervention if they received an NSB score of 20 or less and an ASPENS' score in the strategic or intensive ranges. Eligible student's raw NSB and ASPENS scores were converted into norm-referenced standard scores and then summed to form an overall composite score for each student. Within each classroom students were rank ordered, and the 10 ROOTS-eligible students with the lowest composite scores were randomly assigned to one of three conditions: (a) 2:1 ROOTS group, (b) 5:1 ROOTS group, or (c) a no-treatment control condition (not included in the present study). Of 3,130 screened students, 1,251 met eligibility criteria and were randomly assigned to the 2:1 group condition (n = 258), the 5:1 group condition (n = 622), or the no-treatment control condition (n = 371). All data management were conducted by project's independent evaluator.

ROOTS Intervention

The ROOTS intervention was delivered by district employees, instructional assistants, and interventionists hired specifically for the efficacy trial. The ROOTS interventionists participated in two five-hour professional development workshops that were delivered by project staff. All interventionists also received between two and four coaching visits from ROOTS coaches during intervention implementation to boost implementation fidelity and enhance instructional quality. The coaching visits consisted of direct observations of lesson delivery followed by feedback on instructional quality and fidelity of intervention implementation.

ROOTS. ROOTS is a 50-lesson, Tier 2 mathematics program designed to build students' proficiency in whole number concepts and skills. The ROOTS intervention was delivered in 20-minute small group sessions (2:1 or 5:1) 5 days per week for approximately 10 weeks. Instruction for all students began in the late fall and ended in the spring, and this start date was selected to provide students with the opportunity to respond to initial core mathematics instruction and to therefore minimize the identification of typically achieving students. ROOTS was designed to supplement core mathematics instruction and thus was delivered at times that did not conflict with students' core instruction in mathematics.

Emphasizing the Counting and Cardinality, Operations and Algebraic Thinking, and Number and Operations in Base Ten domains of the CCSS-M, ROOTS targets critical whole number mathematics skills, in line with expert recommendations for early mathematics intervention programs (Gersten et al., 2009). Importantly, ROOTS focuses on supporting the transition from informal to formal mathematics through the use of concrete and representational mathematics models to build understanding of abstract mathematical concepts (Agrawal & Morin, 2016; Witzel, Mercer, & Miller, 2003). For example, ROOTS uses finger models, base ten blocks, and teddy bear counters to provide physical representations of number. Students are then exposed to visual representations of number, such as ten frames or tally marks. These types of representations also have the advantage of fostering initial base ten understanding. Repeated work with number models eventually builds to students working with the abstract numerals themselves.

Though ROOTS provides an introduction to number names and the count sequence up to 100, it places a strong focus on mastery of whole number concepts and skill from 0 to 20. This is especially important given the difficulties that many at-risk students face with teen numbers, the first two-digit numbers that students encounter (National Research Council, 2001). ROOTS activities emphasize building conceptual understanding of the meaning of numerals (e.g., the numeral “8” represents 8 things or entities), relations among numbers (e.g., understanding that each successive number in the count sequence represents a quantity that is one more, comparing number magnitudes), and place value understanding (e.g., composing and decomposing the teen numbers into tens and ones). ROOTS places an equal emphasis on building procedural fluency and automaticity through repeated practice opportunities and systematic review within and across lessons. The ROOTS instructional approach is drawn from principles of explicit and systematic mathematics instruction (Coyne, Kame'enui, & Carnine, 2011; Gersten et al., 2009) including explicit teacher modeling, deliberate practice, visual representations of mathematics, and academic feedback. Frequent opportunities for students to verbalize their mathematical thinking and discuss problem solving methods are also embedded throughout the program’s lessons.

Outcome Measures

CE analyses were completed for a range of student outcome measures including a proximal measure, a measure of number sense, and a distal generalized measure of early mathematics understanding.

ROOTS Assessment of Early Numeracy Skills (RAENS; Doabler, Clarke, & Fien, 2012) is a researcher-developed, individually administered measure that consists of 32 items. The measure is aligned to the CCSS whole number standards for kindergarten. Items assess aspects of counting and cardinality, number operations, and the base-10 system. In an untimed setting, students are asked to count and compare groups of objects, write, order, and compare numbers, label visual models (e.g. ten-frames), and write and solve single digit addition expressions and equations. RAENS' predictive validity ranges from .68 to .83 for the TEMA-3 and the NSB (Clarke, Doabler, Smolkowski, et al., 2016)

Assessing Student Proficiency in Early Number Sense (ASPENS; Gersten et al., 2012) is a set of three 1-minute fluency-based measure that each assess an important aspect of early numeracy proficiency, including number identification, magnitude comparison, and missing number. Test-retest reliabilities of kindergarten ASPENS measures range from .74 to .85. Predictive validity of fall scores on the kindergarten ASPENS measures with spring scores on the TerraNova 3 is reported as ranging from .45 to .52.

Test of Early Mathematics Ability – Third Edition (TEMA-3; Ginsburg & Baroody, 2003) is a standardized, norm-referenced, individually administered measure of beginning mathematical ability. The TEMA-3 assesses whole number understanding for children ranging in age from 3 to 8 years 11 months. Alternate-form and test-retest reliabilities of the TEMA-3 are .97 and .93. The TEMA-3 has concurrent validity with other mathematics measures ranging from .54 to .91.

Source of Effect Sizes

The estimation of CE ratios requires a measure of effect size. The primary source for Hedges' *g* values in Table 1 was Clarke et al. (in press). In this study, the authors accommodated the partially nested design with an analysis that accounted for the nesting of students within small groups in the intervention condition and no clustering of students in the control condition

(Baldwin et al., 2011; Bauer, Sterba, & Hallfors, 2008). For a detailed description of the analysis approach specific to these studies, see Clarke, Doabler, Smolkowski, et al. (2016).

Hedges' *g* values were calculated with the procedures in What Works Clearinghouse (2017). In this paper, we presented additional details and supplemented the effect sizes with confidence intervals. The effect size estimates specifically for small groups compared to the control sample or large groups compared to the control sample were not reported in Clarke et al. (in press), which reported differences in effects between small and large groups. Clarke et al. (in press) reported moderation effects for the posttest TEMA using pretest TEMA as the moderator, but it did not report moderation effects for the RAENS or ASPENS or any effect sizes for percentiles. Finally, Clarke et al. (in press) did not report confidence intervals for effect sizes. We provide those herein to express the precision of estimates given we have not included the full model results.

Cost of ROOTS Implementation

We estimate costs of ROOTS using the "ingredients method" (see Levin, McEwan, Belfield, Bowden, & Shand, 2017) treating ROOTS as an add-on program. Ingredients method involves identifying resource requirements of a program and subsequently assigning values to these resources. Determining the value of each ingredient used for the program requires considering both direct costs of these ingredients and opportunity costs associated with the use of them for program-related activities. Budgeted expenditures tend to underestimate the true economic costs because they reflect only the direct costs (Levin et al., 2017) which are typically more straightforward to calculate as they often involve direct payments. Contrarily, opportunity costs refer to the value of the forgone alternative use for an existing resource and are often more complex to pinpoint.

The list of ingredients used in the ROOTS program are given in Table 2 broken down by the key stages of implementation. The quantities and unit prices associated with each ingredient are given in the last two columns. We exclude all costs associated with research activities, and price ingredients using national averages rather than local estimates.

Ingredients and prices: curriculum, training, coaching. The principal ingredient used in implementation of ROOTS program is the intervention kit. Each intervention kit includes curriculum books and student worksheets for one five-student group. A ROOTS instructor teaching more than one group can use the same curriculum books for multiple groups but needs to make copies of the student material for each additional group. The manipulatives needed for the program are not included with the intervention kit but can be purchased separately. The price of intervention kit and manipulatives are from the developers who publish the materials (https://dibels.uoregon.edu/market/movingup/kfoundation#pricing). The time requirements and fees charged for training and coaching are based on the reports of the developers and the fees charged by individual ROOTS trainers. As noted earlier, coaching visits consist of direct observations of lesson delivery and a brief feedback session on instructional quality and fidelity of implementation. In calculating instructor's time requirement for coaching, we include only the feedback session of coaching visits to avoid double counting of instructor's time spent on delivery.

Ingredients and prices: screening. We treat screening as a part of the ROOTS implementation, and accordingly, we determine the costs associated with screening of all kindergarten students in schools and include them in total cost calculation. It is likely, however, that schools already use different screening tools to assess early mathematics proficiency in their kindergarten classroom. The results of these tests can be used to determine at risk students appropriate for ROOTS intervention. Alternatively, schools may choose to rely on teachers' judgement based on their observation of students' math proficiency in selecting students to receive intervention. In those cases, the costs of screening do not apply, and thus the total cost of implementation is lower.

Ingredients and prices: instructor. As mentioned earlier, the ROOTS intervention includes 50 lessons each designed to be delivered in 20 minutes. We add a 10-minute instructor preparation time for each group session, bringing the time requirement for delivery to 30 minutes per session. We assume that there were no significant changes in instruction in the classroom from which the intervention students were pulled out. We also assume that the instructional assistants teaching ROOTS have not discontinued another activity that affects student outcomes.

Instructional assistant's time dedicated to all program activities is priced at the total of wages and benefits. For wages, we converted national median annual "teacher assistant" salary provided in the Occupational Outlook Handbook (US Bureau of Labor Statistics, 2019b) into hourly wage assuming 2080 work hours per year. The benefit rates for employees in public schools are from BLS Employer Costs for Employee Compensation (US Bureau of Labor Statistics, 2019a) which indicates that in 2018 benefits represent 33% of total compensation for primary school teachers.

Ingredients and prices: facilities. ROOTS trainings take place in spaces available for meetings and trainings in district offices, whereas for the delivery of the program and the feedback session, any available space in school building such as unused classrooms or meeting rooms, or library when available can be used. We follow the approach outlined in Levin et al. (2017) to calculate the costs associated with use of these facilities for ROOTS activities. First, we use the median construction costs of an office building and an elementary school building as suggested by CostOut, the CBCSE Cost Tool Kit. These construction costs are uprated by 21% to include furniture, furnishing, fees, and site preparation as per School Planning and Management magazine. We then annualize these building costs over 30 years using the conventional 3% interest rate to obtain the cost of per square foot district office and per square foot school space per year. The size of the space needed is assumed to be 900sqft for trainings, and 200sqft for delivery and feedback session.

Ingredients and Prices: Units. As noted in the third column of Table 2, some costs are borne at the district level, while others are at the school, instructor, or even group level. Calculating total costs at district level or converting the total costs into per group costs requires making assumptions about the number of schools in the district, number of instructors trained, and number of groups per instructor. We provide cost estimates for an "average" school district that have ten elementary schools with three kindergarten classes per school. We assume there are three groups per school (one group per class), and one instructor in each school leading all three groups.

All prices are converted into 2018 dollars. ROOTS is a one-year program where there is only a short period of time (less than a year) between when the costs are borne and when the effectiveness is measured, and thus, no adjustment is necessary for discounting of costs or returns.

Sensitivity of Costs to District Size. The cost of ROOTS program depends highly on the number of schools implementing the program as well as the number of instructors delivering the program in each school. In addition to the cost estimates for an “average” school district, we provide estimates for larger- or smaller-than-average districts showing sensitivity of our cost estimates to district size, number of instructors in trained in district, and number of groups per school.

Sensitivity of costs to “resource reallocation”. Our cost estimation treats ROOTS as an add on program, and all resources used in ROOTS implementation, including instructors time, as having a cost. We assume schools purchase more instructor time for ROOTS activities. In general, even when an instructor dedicates time to program activities within working hours without receiving extra payment, the value of their time spent on program-related work is accounted for in cost calculations. The idea behind this is that time spent on the program is in addition to time spent on their usual duties.

For a school with full-time instructional assistants in kindergarten classrooms assisting teachers in support of at-risk students, there may be room for resource reallocation. As argued in the context of other programs (e.g. Borman et al., 2002; Odden & Archibald, 2000), resources used especially in delivery of the program can be covered by reallocating existing staff and facility. Accordingly, we provide cost estimates with full resource reallocation of instructors' time and facility costs in program delivery in a school where kindergarten mathematics screening is routinely performed. We still include instructor time and facilities needed in training and coaching as part of the cost estimate because existing staff will still need to be trained and coached to implement a new program.

Cost Effectiveness Analysis

CE ratio represents the cost per unit of change in the outcome of interest. It is calculated by dividing the cost by the effect size where both the cost and the effect size are measured in the same unit such as per student. We divide the per-student costs calculated as described above by the effect sizes calculated elsewhere (Clarke et al., in press). We are confident that that the effectiveness data and data on resource requirements match as they were collected simultaneously.

In calculating costs, we consider resource use above and beyond the resources used in the status quo with the assumption that students continue to receive all other services as part of their regular instruction in school, and that ROOTS is not being implemented as replacement of another program. Therefore, the CE ratios we report can be considered as incremental CE ratios.

If ROOTS replaces another add-on program targeting similar math outcomes, it is likely that resources that are diverted from that program will imply further cost reductions for ROOTS. However, in that case, the CE ratio should use the incremental effect size which reflect the improvement in targeted outcomes over what the replaced program achieved. Similarly, if ROOTS replaces a program targeting other outcomes in other subject areas (e.g. reading), then a careful consideration should be given to the changes in those outcomes, relative efficiency of the two programs, and prioritization based on student needs. Therefore, we do not estimate CE ratio for the resource reallocation scenario discussed earlier. Lastly, we provide CE ratios for the cost scenarios for varying district sizes using the same effect sizes for each of these scenarios with the assumption that the effect sizes are not sensitive to these variables.

Results

Cost of ROOTS

Table 3 displays per-group cost estimates for each ingredient used in ROOTS implementation, along with total costs per group and per student. For an average size district that has ten elementary schools with one instructor and three groups in each school (Scenario A, column 2), the total per group cost is $1,192. Of these costs, 67.3% are labor costs that include staff time spent on delivery of the program as well as other program activities. Program curriculum, training and coaching fees, and material for screening account for another 28.7%. The remaining 5% is the facility costs.

As evident from Table 3, some ROOTS costs are fixed at district level (e.g. training fee), while others vary at the instructor (e.g. intervention kit) or group level (e.g. delivery). To demonstrate the sensitivity of these cost estimates to district size and choices regarding number of instructors, we present cost estimates for four other scenarios. For a larger district that has 20 elementary schools with four groups and two instructors per school (Scenario B) the per-group cost of the program is slightly higher at $1,268. However, schools in this scenario can reduce their per-group costs nearly %15 to $1,082 by having only one instructor per school teaching all four groups in school (Scenario C). The savings result primarily from the fact that fewer instructors mean less instructor time spent in training and coaching, and fewer intervention kits needed (i.e. instructor using the same kit for more groups).

In a smaller district that has only five elementary schools with two groups and one instructor per school (Scenario D), the per-group costs go up to $1,472. Main reason for this increase is that the fixed training costs at the district level and other instructor level costs (such as time cost of training and cot of intervention kit) are distributed across a smaller number of groups. Savings, similar to those outlined in Scenario C, can be achieved in a small district through resource sharing. Specifically, a small district may choose to train one instructor to teach all ROOTS groups across all schools in the district. Scenario E given in the last column of Table 3 assumes there is one ROOTS instructor in the district teaching all groups in all five schools. This results in a 20% decrease in per group costs from $1,472 to $1,174.

In each scenario, we divide these per-group costs by group size to achieve per student costs in both small and large groups. Per student cost in an average district is $238 for a five-student large group, and $596 for a two-student small group. Depending on district size and choice of number of instructors per school, these per-student costs vary between $216.31 and

$294 for large group, and between $541 and $736 for small groups.

For a school in the full resource reallocation scenario where the resources used in ROOTS delivery are covered by reallocating existing staff and facility from other activities and the school has routine kindergarten math screening, the delivery and screening costs will not be a part of incremental cost of ROOTS implementation. In that case, per-group cost of ROOTS in an average district will be $407, and per-student costs will be $81 and $203 for large and small groups, respectively.

Cost Effectiveness of ROOTS

We utilized the effect sizes reported in Table 1 to calculate CE ratios. When contrasting the treatment sample (both ROOTS conditions) to the control condition, Hedges' g effect sizes ranged from .23 on the TEMA to .81 on the ASPENS indicating the ROOTS intervention showed an educationally meaningful impact on a range of proximal and distal measures (see Table 1). Effect sizes were roughly equivalent across the two treatment conditions (e.g. .47 for the small group of 2 students and .50 for the large group of 5 students). Table 1 also reports out effect sizes by initial skill (collapsed across treatment conditions). Two of the three measures (the TEMA and RAENS) show stronger results for students who had lower initial skills (e.g. the effect size for initial skill at the 5th percentile of the sample was 1.11 on the RAENS compared to an effect size of .54 at the 95th percentile).

We calculated CE ratios with the full-sample effect sizes from three measures: RAENS (Hedges' g = 0.81), ASPENS (0.49), and TEMA (0.23). These measures represent differing degrees along the proximal-to-distal continuum. The effect sizes can also be interpreted as representative examples of effect sizes in Table 1, which range from 0.05 to 1.11. Table 4 presents CE ratios for different district scenarios discussed above and whether students are taught in small groups of two or large groups of five.

The CE ratio is calculated as the cost divided by the effect size, which represents the costs per-unit change in effect size for a given cost scenario. For example, in District A, the CE ratio of $294 for the RAENS in row 1 equals the costs per student for large groups, $238, divided by the effect size, 0.81. This implies that it costs approximately $294 to achieve a 1.0 effect size gain in student performance over conditions similar to those in the comparison condition. Critically, because the cost of the two student small group is roughly double that of the five student group ($596 to $238) the CE ratio for the five student small group is $736 compared to a CE ratio of $294.

Discussion

The relative lack of CE literature for academic programs and intervention programs, including the lack of CE work in the area of early mathematics intervention speaks to the need for conducting CE analyses (Levin & Belfield, 2015). While conducting CE analyses is being advocated for as a standard part of educational research (IES RFP; US Dept. of Education, 2018), the process is nuanced and interpreting results from CE analysis is complex. A quick glance at our results illustrates this point. The costs per group per effect-size unit change varies from $216 to $736 depending on district scenario and group size, and the CE ratios per student vary from $267 to $3,201 depending on district scenario, group size, and the measure (see Table 4, average district and district size sensitivity analysis). Despite this complexity, a few distinct points emerge related to the findings from this particular study. First, both treatment conditions showed a moderate to strong impact on a range of measures with effect sizes meeting a threshold for being educationally meaningful. Based on what we know related to early mathematics development, a focus by schools to provide mathematics intervention as students transition from informal to formal settings seems to be a worthwhile investment. Second, the roughly equivalent impact found across treatment conditions (small and large group delivery) but at substantively higher costs for the small group indicate that the cost of providing the ROOTS intervention is not worth the additional costs. Collectively, these two findings have implications for practice and policy. The continued advocation for and provision of early mathematics interventions should remain a top priority. However, serious discussion is needed related to default assumptions in RTI and MTSS models in regard to group size. The generally accepted notion that smaller group size is a mechanism to increase intensity, and thus student outcomes, both within and when moving across tiers (e.g. from Tier 2 to Tier 3) needs to be questioned. Our results suggest the need to examine how services are delivered in RTI and MTSS models and associated practice `and policy that is currently in place. Such issues should be considered with a substantial degree of caution as the findings from this study are specific to the intervention evaluated and results should not be overly generalized across intervention programs and settings.

Given the infancy of the research and availability of CE data for intervention programs, we focus the remainder of the discussion on key factors that researchers, program developers, and practitioners should consider when deriving or utilizing CE data. When conducting a CE analysis there are a range of potential costs to include and those costs vary across settings. Should facilities costs be included or excluded? How should we consider interventions which are often accessed through screening procedures that may already be in place within a school or district? And what about support costs including professional development and coaching? The result is that deriving an estimate of cost when conducting a CE analysis is not a simple or straightforward process. One proposed solution is for CE analyses to include a thorough list of components included. Comprehensive component lists would allow districts to select the costs that are relevant (Blonigen et al., 2008). For example, if a district has in place screening procedures, instructional assistants to deliver interventions during a pre-specified period, and sites to conduct the intervention, then they can consider the cost side of the equation to consist of only materials costs and derive a CE ratio.

A second factor to consider is the outcome measure used to calculate the CE ratio and the breadth of the intervention. Outcome measures that are proximal to the intervention are likely to result in greater effect sizes (Clarke et al., in press) and thus lower CE ratios. Whereas a distal measure that may focus on generalizing knowledge and skills to a broader measure will likely result in lower effect sizes and a higher CE ratio. But even using the terms proximal and distal can be fraught with complications. In the present study, for example, we utilized a proximal measure, RAENS, yet the content on that measure comprehensively covers the CCSS-M whole number domains for kindergarten whereas our distal measure, the TEMA-3, covers content spanning preschool to second grade and includes items outside of the intervention's focus on whole number (e.g. geometry items). Is the result on the RAENS or TEMA-3 a more valid representation of student mathematics development? At a minimum, measurement nets in CE analyses should include proximal and distal measures that should be evaluated on the importance of the outcomes they measure.

Consideration of what the outcome measure and associated effect sizes represent is inextricably linked to the breadth of the intervention. For example, consider the challenge posed by interventions focused on a singular narrow aspect of mathematics, like basic facts or computational fluency, in contrast to more comprehensive interventions focused on conceptual understanding of number systems (e.g. whole number). Several interventions exist that address procedural fluency (e.g. Cover, Copy, Compare) that are relatively narrow within the broader scope of what students with math difficulties need to learn and understand in mathematics. An appeal of these interventions is that they are also relatively easy to implement and have low costs in comparison to interventions that require direct teaching and associated professional development and materials costs. Further complicating a CE analyses is that these interventions are often evaluated with measures that are both narrow and proximal to the intervention content. It is likely that such an approach would result in a low CE ratio (i.e. it costs very little to achieve a high impact or effect size). However, does a low CE ratio matter if it doesn't result in a meaningful change in the student's knowledge? While fluency is a critical construct in mathematics (Clarke, Nelson, & Shanley, 2016), a sole focus on fluency is insufficient for addressing the needs of students with significant difficulties in mathematics (Gersten et al., 2009). There needs to be caution in interpreting CE ratio without strongly considering what the effect size represents in terms of the change in student knowledge.

A third factor to consider is who is receiving the intervention and analyzing for whom the intervention works (Miller et al., 2014). An analysis of whether initial skill moderated intervention outcomes of ROOTS (Clarke et al., 2019) found the greatest impact for students with the most severe deficits but no impact for students that were at the high end of the at-risk sample. Thus, as we reported in the results section, there are different CE ratios depending on the initial skills status of the student with low CE ratios for students with the most severe needs (i.e. high return on investment) but high CE ratios for students with stronger initial skills (i.e. low return on investment). From a tiered service delivery model perspective, an intensive program like ROOTS may be more suitable for tier 3 students with significant gaps in their conceptual knowledge whereas a less intensive approach like Cover, Copy, Compare may be a more economical choice for a student with less severe needs. Districts should consider treatment intensity and the intensity of student needs (Codding & Lane, 2015) when evaluating the CE ratios and the suitability of different programs to fit within service delivery models and address the range of student needs.

A final factor is recognizing that most CE ratios are likely to be derived from research studies. Given the research to practice gap (Cook & Odom, 2013), it is likely that there will be variance in implementation between a research study that is highly supported by external resources and school implementation that will be dependent on available resources (Onken, Carroll, Shoham, Cuthbert, & Riddle, 2014). As such, when evaluating and using CE data thought must be given to how much support was given to obtain the results. It is not uncommon in intervention research studies to have graduate students serve as interventionists and implement programs with high degrees of fidelity. Whereas in practice, instructional assistants may deliver intervention programs with lower degrees of support (e.g. coaching). While the general goal of replication studies is to derive estimates of effects across varying sites and conditions (Coyne et al., 2016), replication studies that include CE analyses are also critical to provide a range of costs, effects, and CE ratios. Analyzing implementation features found in research studies and replication studies may serve to provide schools with an estimate of what effect size they may expect or the amount of support, and associated costs, they may need to provide in order to approximate the effect found in a research study. For example, in this particular work, we assumed effect sizes would not vary by district size. However, in practice it would be reasonable to hypothesize that effect sizes might be lower in larger districts due to the complexity of operating with a larger system/district or alternately to hypothesize that larger districts would obtain greater impacts if they had already existing resources to support implementation fidelity.

Lastly, we would be remiss not to address an unspoken goal of CE work which is to enable the comparison of programs. The assumption is that a district or school choosing between programs could utilize CE ratios to guide that selection. As the field moves forward, we think it would be an unfortunate outcome if CE ratios were exclusively analyzed within a vacuum to compare programs without full consideration of a multitude of critical factors. We encourage the field to not only conduct CE research but also to offer a full accounting related to the costs of the intervention, the outcome measures used, the breadth of the intervention, the students for whom the intervention is intended, and implementation support. If these factors are fully considered, discussed, and debated, CE analyses have the potential to offer greater insight into effective and economical ways to address the learning needs of all students.

References

Agrawal, J., & Morin, L. L. (2016). Evidence-based practices: Applications of concrete representational abstract framework across math concepts for students with mathematics disabilities. *Learning Disabilities Research & Practice, 31*(1), 34–44. doi: 10.1111/ldrp.12093

Archer, A. L., & Hughes, C. A. (2011). Exploring the foundations of explicit instruction. In A. L. Archer & C. A. Hughes (Eds.), *Explicit instruction: Effective and efficient teaching* (pp. 1–22). New York, NY: The Guilford Press.

Baldwin, S. A., Bauer, D. J., Stice, E., & Rohde, P. (2011). Evaluating models for partially clustered designs. *Psychological Methods, 16*, 149–165. doi: 10.1037/a0023464

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. M. (2015). *Evaluation of response to intervention practices for elementary school reading* (NCEE No. 2016-4000). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved from http://files.eric.ed.gov/fulltext/ED560820.pdf

Bauer, D. J., Sterba, S. K., & Hallfors, D. D. (2008). Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavioral Research, 43*, 210–236. doi: 10.1080/00273170802034810

Begeny, J. C., Hawkins, A. L., Krouse, H. E., & Laugle, K. M. (2011). Altering instructional delivery options to improve intervention outcomes: Does increased instructional intensity also increase instructional effectiveness? *Psychology in the Schools, 48*(8), 769–785. doi: 10.1002/pits.20591

Begeny, J. C., Levy, R. A., & Field, S. A. (2018). Using small-group instruction to improve students' reading fluency: An evaluation of the existing research. *Journal of Applied School Psychology, 34*(1), 36–64. doi: 10.1080/15377903.2017.1328628

Begeny, J. C., Yeager, A., & Martínez, R. S. (2012). Effects of small-group and one-on-one reading fluency interventions with second grade, low-performing spanish readers. *Journal of Behavioral Education, 21*(1), 58–79. doi: 10.1007/s10864-011-9141-x

Blonigen, B. A., Harbaugh, W. T., Singell, L. D., Horner, R. H., Irvin, L. K., & Smolkowski, K. S. (2008). Application of economic analysis to school-wide positive behavior support (swpbs) programs. *Journal of Positive Behavior Interventions, 10*, 5–19. doi: 10.1177/1098300707311366

Borman, G., Hewes, G., Overman, L., & Brown, S. (2002). *Comprehensive school reform and student achievement: A meta-analysis* (Report No. CRESPAR-R-59). Baltimore, MD: CRESPAR/Johns Hopkins University. Retrieved from http://files.eric.ed.gov/fulltext/ED472569.pdf

Clarke, B., Doabler, C., Smolkowski, K., Kurtz Nelson, E., Fien, H., Baker, S. K., & Kosty, D. (2016). Testing the immediate and long-term efficacy of a tier 2 kindergarten mathematics intervention. *Journal of Research on Educational Effectiveness, 9*, 607–634. doi: 10.1080/19345747.2015.1116034

Clarke, B., Doabler, C., Turtura, J., Smolkowski, K., Kosty, D., Sutherland, M., . . . Baker, S. K. (in press). Examining the efficacy of a kindergarten mathematics intervention by group size and initial skill: Implications for practice and policy. *Elementary School Journal*.

Clarke, B., Doabler, C. T., Fien, H., Baker, S. K., & Smolkowski, K. (2012). *A randomized control trial of a tier 2 kindergarten mathematics intervention* (Project ROOTS). (USDE; Institute of Education Sciences; Special Education Research, CFDA Num: 84.324A, 2012-2016, Funding Number: R324A120304, awarded $3,338,552).

Clarke, B., Doabler, C. T., Kosty, D., Kurtz Nelson, E., Smolkowski, K., Fien, H., & Turtura, J. (2017). Testing the efficacy of a kindergarten mathematics intervention by small group size. *AERA open, 3*(2), 1–16. doi: 10.1177/2332858417706899

Clarke, B., Doabler, C. T., Smolkowski, K., Baker, S. K., Fien, H., & Strand Cary, M. (2016). Examining the efficacy of a tier 2 kindergarten mathematics intervention. *Journal of Learning Disabilities, 49*, 152–165. doi: 10.1177/0022219414538514

Clarke, B., Doabler, C. T., Smolkowski, K., Turtura, J., Kosty, D., Kurtz-Nelson, E., . . . Baker, S. K. (2019). Exploring the relationship between initial mathematics skill and a kindergarten mathematics intervention. *Exceptional Children, 85*(2), 129–146. doi: 10.1177/0014402918799503

Clarke, B., Doabler, C. T., Strand Cary, M., Kosty, D. B., Baker, S. K., Fien, H., & Smolkowski, K. (2014). Preliminary evaluation of a tier-2 mathematics intervention for first grade students: Utilizing a theory of change to guide formative evaluation activities. *School Psychology Review, 43*, 160–177.

Clarke, B., Gersten, R. M., Dimino, J., & Rolfhus, E. (2011). *Assessing student proficiency of number sense (ASPENS)*. Longmont, CO: Cambium Learning Group, Sopris Learning.

Clarke, B., Nelson, N. J., & Shanley, L. (2016). Mathematics fluency—more than the weekly timed test. In K. Cummings & Y. Petscher (Eds.), *The fluency construct* (pp. 67–89). New York: Springer Science+Business Media. doi: 10.1007/978-1-4939-2803-3_3.

Codding, R. S., & Lane, K. L. (2015). A spotlight on treatment intensity: An important and often overlooked component of intervention inquiry. *Journal of Behavioral Education, 24*, 1–10. doi: 10.1007/s10864-014-9210-z

Codding, R. S., VanDerHeyden, A. M., Martin, R. J., Desai, S., Allard, N., & Perrault, L. (2016). Manipulating treatment dose: Evaluating the frequency of a small group intervention targeting whole number operations. *Learning Disabilities Research & Practice, 31*(4), 208–220. doi: 10.1111/ldrp.12120

Cook, B. G., & Odom, S. L. (2013). Evidence-based practices and implementation science in special education. *Exceptional Children, 79*(3), 135–144. doi: 10.1177/001440291307900201

Coyne, M. D., Cook, B. G., & Therrien, W. J. (2016). Recommendations for replication research in special education: A framework of systematic, conceptual replications. *Remedial and Special Education, 37*, 244–253. doi: 10.1177/0741932516648463

Coyne, M. D., Kame'enui, E. J., & Carnine, D. (2011). *Effective teaching strategies that accommodate diverse learners* (4th ed.). Upper Saddle River, NJ: Pearson Education.

Doabler, C. T., Clarke, B., & Fien, H. (2012). *Roots assessment of early numeracy skills (raens)*. Unpublished measurement instrument. Center on Teaching and Learning, University of Oregon. Eugene, OR.

Doabler, C. T., Clarke, B., Kosty, D., Kurtz-Nelson, E., Fien, H., Smolkowski, K., & Baker, S. K. (2019). Examining the impact of group size on the treatment intensity of a tier 2 mathematics intervention within a systematic framework of replication. *Journal of Learning Disabilities, 52*, 168–180. doi: 10.1177/0022219418789376

Doabler, C. T., Clarke, B., Kosty, D., Smolkowski, K., Kurtz Nelson, E., Fien, H., & Baker, S. K. (2019). Building number sense among english learners: A multisite randomized controlled trial of a tier 2 kindergarten mathematics intervention. *Early Childhood Research Quarterly, 47*, 432-444. doi: 10.1016/j.ecresq.2018.08.004

Doabler, C. T., Clarke, B., Kosty, D. B., Kurtz-Nelson, E., Fien, H., Smolkowski, K., & Baker, S. K. (2016). Testing the efficacy of a tier 2 mathematics intervention: A conceptual replication study. *Exceptional Children, 83*, 92–110. doi: 10.1177/0014402916660084

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428–1446. doi: 10.1037/0012-1649.43.6.1428

Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilities, 46*, 166–181. doi: 10.1177/0022219411410233

Elbaum, B., Vaughn, S., Hughes, M. T., Moody, S. W., & Schumm, J. S. (2000). How reading outcomes of students with disabilities are related to instructional grouping formats: A meta-analytic review. In R. M. Gersten, E. P. Schiller & S. Vaughn (Eds.), *Contemporary special education research: Syntheses of the knowledge base on critical instructional issues* (pp. 105–136). Mahwah, NJ.: Lawrence Erlbaum Associates.

Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide* (Practice Guide No. NCEE 2014-4005). Washington, DC: National Center for Education and Regional Assistance (NCEE), Institute of Education Sciences, US Department of Education. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/early_math_pg_111313.pdf

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*, 493–513. doi: 10.1037/0022-0663.97.3.493

Gersten, R. M., Beckmann, S., Clarke, B., Foegen, A., March, L., Star, J. R., & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to intervention (rti) for elementary and middle schools* (Practice Guide Report No. NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/rti_math_pg_042109.pdf

Gersten, R. M., & Chard, D. J. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education, 33*, 18–28. doi: 10.1177/002246699903300102

Gersten, R. M., Clarke, B., Jordan, N., Newman-Gonchar, R., Haymond, K., & Wilkins, C. (2012). Universal screening in mathematics for the primary grades: Beginnings of a research base. *Exceptional Children, 78*, 423–445. Retrieved from http://cec.metapress.com/content/B75U2072576416T7

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of early mathematics ability- third edition (tema-3)*. Austin, TX: ProEd.

Hollands, F. M., Kieffer, M. J., Shand, R., Pan, Y., Cheng, H., & Levin, H. M. (2016). Cost-effectiveness analysis of early reading programs: A demonstration with recommendations for future research. *Journal of Research on Educational Effectiveness, 9*(1), 30–53. doi: 10.1080/19345747.2015.1055639

Hunter, L. J., DiPerna, J. C., Hart, S. C., & Crowley, M. (2018). At what cost? Examining the cost effectiveness of a universal social–emotional learning program. *School Psychology Quarterly, 33*(1), 147–154. doi: 10.1037/spq0000232

Jordan, N. C., Glutting, J., & Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learning and individual differences, 20*, 82–88. doi: 10.1016/j.lindif.2009.07.004

Knight, D. S., Landry, S., Zucker, T. A., Williams, J. M., Merz, E. C., & Taylor, H. B. (2016). *Cost-effectiveness of early childhood interventions to enhance preschool: Evidence from a randomized experiment in head start centers enrolling historically underserved populations* (CERPS Working Paper No. 2016-3). EL Paso, TX: Center for Education Research and Policy Studies, University of Texas at El Paso DigitalCommons@UTEP. Retrieved from https://digitalcommons.utep.edu/cgi/viewcontent.cgi?article=1002&context=cerps_wp

Levin, H. M., & Belfield, C. (2015). Guiding the development and use of cost-effectiveness analysis in education. *Journal of Research on Educational Effectiveness, 8*(3), 400–418. doi: 10.1080/19345747.2014.915604

Levin, H. M., & McEwan, P. J. (2002). *Cost-effectiveness analysis: Methods and applications*. Thousand Oaks, Calif.: Sage Publications. Retrieved from http://catalog.hathitrust.org/api/volumes/oclc/44066885.html

Levin, H. M., McEwan, P. J., Belfield, C., Bowden, A. B., & Shand, R. (2017). *Economic evaluation in education: Cost-effectiveness and benefit-cost analysis*: SAGE Publications. Retrieved from https://books.google.com/books?id=CRgnDwAAQBAJ

Miller, B., Vaughn, S., & Freund, L. (2014). Learning disabilities research studies: Findings from nichd funded projects. *Journal of Research on Educational Effectiveness, 7*, 225–231. doi: 10.1080/19345747.2014.927251

Morgan, P. L., Farkas, G., Wang, Y., Hillemeier, M. M., Oh, Y., & Maczuga, S. (2019). Executive function deficits in kindergarten predict repeated academic difficulties across elementary school. *Early Childhood Research Quarterly, 46*, 20–32. doi: https://doi.org/10.1016/j.ecresq.2018.06.009

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42*, 306–321. doi: 10.1177/0022219408331037

Morgan, P. L., Hillemeier, M. M., Farkas, G., & Maczuga, S. (2014). Racial/ethnic disparities in adhd diagnosis by kindergarten entry. *Journal of Child Psychology and Psychiatry, 55*, 905–913. doi: 10.1111/jcpp.12204

National Assessment Educational Progress. (2019). *Results from the 2019 mathematics and reading assessments* (NCES No. 2020012). Washington, DC: National Center for Education Statistics, U.S. Department of Education, Institute of Education Sciences. Retrieved from https://www.nationsreportcard.gov/mathematics?grade=4

National Association of State Directors of Special Education, & Council of Adminsitrators of Special Education. (2006). *Response to intervention. A joint paper by the national association of state directors of special education and the council of administrators of special education* (Joint paper. Washington, DC: National Association of State Directors of Special Education, Council of Adminsitrators of Special Education

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: Mathematics Learning Study Committee

Ochsendorf, R. (2016). Advancing understanding of mathematics development and intervention: Findings from ncser-funded efficacy studies. *Journal of Research on Educational Effectiveness, 9*, 570–576. doi: 10.1080/19345747.2016.1222144

Odden, A., & Archibald, S. (2000). *Reallocation of resources: How to boost student achievement without asking for more*. Thousand Oaks, CA: Corwin.

Onken, L. S., Carroll, K. M., Shoham, V., Cuthbert, B. N., & Riddle, M. (2014). Reenvisioning clinical science: Unifying the discipline to improve the public health. *Clinical psychological science : a journal of the Association for Psychological Science, 2*(1), 22–34. doi: 10.1177/2167702613497932

Ross, S. G., & Begeny, J. C. (2015). An examination of treatment intensity with an oral reading fluency intervention: Do intervention duration and student–teacher instructional ratios impact intervention effectiveness? *Journal of Behavioral Education, 24*(1), 11–32. doi: 10.1007/s10864-014-9202-z

Sood, S., & Jitendra, A. K. (2013). An exploratory study of a number sense program to develop kindergarten students' number proficiency. *Journal of Learning Disabilities, 46*, 328–346. doi: 10.1177/0022219411422380

US Bureau of Labor Statistics. (2019a). *Employer costs for employee compensation* (News Release USDL No. 19-1649). Washington, DC: US Bureau of Labor Statistics. Retrieved from https://www.bls.gov/news.release/pdf/ecec.pdf

US Bureau of Labor Statistics. (2019b). *Occupation outlook handbook*. Washington, DC US Bureau of Labor Statistics. Retrieved from https://www.bls.gov/ooh/home.htm?view_full

US Dept. of Education. (2018). *Rfa: Special education research grants* (CFDA No. 84.324A). Washington, DC: Institute of Education Sciences. Retrieved from https://ies.ed.gov/funding/pdf/2019_84324A.pdf

Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice, 18*, 137–146. doi: 10.1111/1540-5826.00070

Vaughn, S., Klingner, J. K., Swanson, E. A., Boardman, A. G., Roberts, G., Mohammed, S. S., & Stillman-Spisak, S. J. (2011). Efficacy of collaborative strategic reading with middle school students. *American Educational Research Journal, 48*(4), 938–964. doi: 10.3102/0002831211410305

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue:Relations between early mathematics knowledge and high school achievement. *Educational Researcher, 43*(7), 352–360. doi: 10.3102/0013189x14553660

What Works Clearinghouse. (2017). *Procedures handbook version 4.0*. Washington, DC: Institute of Education Science, U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Witzel, B. S., Mercer, C. D., & Miller, M. D. (2003). Teaching algebra to students with learning difficulties: An investigation of an explicit instruction model. *Learning Disabilities Research & Practice, 18*(2), 121–131. doi: 10.1111/1540-5826.00068

Table 1

ROOTS Effect Sizes (Hedges' g) for the Full Sample, Three Subsamples, and Four Quantiles of the TEMA for Three Outcomes Measures

	RAENS	ASPENS	TEMA
Full Sample	0.81 [0.69, 0.93]	0.49 [0.38, 0.60]	0.23 [0.14, 0.32]
Small Groups	0.80 [0.65, 0.94]	0.47 [0.32, 0.61]	0.26 [0.13, 0.38]
Large Groups	0.80 [0.66, 0.93]	0.50 [0.37, 0.62]	0.20 [0.11, 0.30]
Moderation by Pretest TEMA			
5^{th} percentile	1.11 [0.89, 1.32]	0.47 [0.26, 0.68]	0.45 [0.28, 0.62]
25^{th} percentile	0.94 [0.80, 1.07]	0.47 [0.34, 0.60]	0.33 [0.22, 0.43]
50^{th} percentile	0.81 [0.70, 0.93]	0.47 [0.36, 0.58]	0.24 [0.15, 0.33]
75^{th} percentile	0.71 [0.58, 0.85]	0.48 [0.35, 0.60]	0.17 [0.06, 0.27]
95^{th} percentile	0.54 [0.33, 0.75]	0.49 [0.28, 0.68]	0.05 [-0.12, 0.21]

Note. Effect size estimates were taken from results from Clarke et al. (in press) and were estimated according to procedures recommended by the What Works Clearinghouse (2017). Large-group and small-group effects, each compared to the control sample, have not been explicitly reported previously (see text for details). We found moderation effects for the RAENS and TEMA but not the ASPENS. Percentiles represent the sample percentiles (not normed percentiles). Confidence intervals not reported previously and provided here to characterize the precision of the effect size estimates.

Table 2.

List of ROOTS Ingredients, and Associated Quantities and Unit Costs

Major Components	Ingredients	Quantity	Unit Costs
Curriculum	Intervention kit	1 per instructor (curriculum books, Practice worksheets for 5 students)	$399+5% shipping
	Worksheet copy	50 pages per intervention student	$0.10 per page
	Manipulatives	1 per group	$55.00
Training	Trainer fee	1 per district (two 5-hour training sessions for all instructors in district)	$2,200.00
	Instructor time	10 person-hours	$25.82/hour[1]
	Facility	900sqft space in district offices for 10 hours	$13.23/hour
Coaching	Coaching fee	1 per instructor (four coaching sessions)	$50.00
	Instructor time	0.7 person-hours	$25.82/hour [1]
	Facility	200sqft school space for 10 min. per coaching session	$1.99/hour
Screening	Staff time	2.5 person-hour per class	$25.82/hour[1]
	Test materials	1 per student	$1.00/per student
	Facility	200sqft school space for 2.5 hours per class	$1.99/hour
Delivery	Instructor time	25 person-hours per group	$25.82/hour[1]
	Facility	200sqft school space for 25 hours per group	$1.99/hour

[1]$17.30 in wages +$8.52 in benefits. $17.20/hour in wages corresponds to annual wage of $26,970 which is the median annual "teacher assistant" salary in 2018 (Bureau of Labor Statistics, 2019a). Benefits represent 33% of total compensation in 2018 (Bureau of Labor Statistics, 2019b). Annual wage converted into hourly rate assuming 2080 work hours per year.

Table 3.

Per Student and Per Group Cost for ROOTS Implementation by Five District Scenarios

District or Major Cost Component	Items	Average District (A)	Larger District (B)	Larger District with More Groups per Instructor (C)	Smaller District (D)	Smaller District with Shared Instructor (E)
		District Scenarios				
Schools		10	20	20	5	5
ROOTS groups per school[1]		3	4	4	2	2
Instructors per school		1	2	1	1	1/District
		Cost Estimates				
Curriculum	Intervention kit	$139.67	$209.50	$104.75	$209.50	$41.90
	Worksheet copy	$25.00	$25.00	$25.00	$25.00	$25.00
	Manipulatives	$55.00	$55.00	$55.00	$55.00	$55.00
Training	Trainer fee	$73.33	$27.50	$27.50	$220.00	$220.00
	Instructor time	$86.07	$129.10	$64.55	$129.10	$25.82
	Facility	$4.41	$1.65	$1.65	$13.23	$13.23
Coaching	Coaching fee	$16.67	$25.00	$12.50	$25.00	$5.00
	Instructor time	$6.02	$9.04	$4.52	$9.04	$1.81
	Facility	$0.46	$0.70	$0.35	$0.70	$0.14
Screening	Staff time	$64.55	$64.55	$64.55	$64.55	$64.55
	Test materials	$21.00	$21.00	$21.00	$21.00	$21.00
	Facility	$4.97	$4.97	$4.97	$4.97	$4.97
Delivery	Instructor time	$645.52	$645.52	$645.52	$645.52	$645.52
	Facility	$49.69	$49.69	$49.69	$49.69	$49.69
Total per group cost		$1,192	$1,268	$1,082	$1,472	$1,174
Large group per student cost		$238	$254	$216	$294	$235
Small group per student cost		$596	$634	$541	$736	$587

[1]The number of groups does not account for group size; adjustments for group size are shown in the last two rows.

Table 4.

Cost-Effectiveness Ratios for Average District and Four Additional District Scenarios

Sensitivity Analysis Scenario	Group Size	Cost per Student	RAENS	ASPENS	TEMA-3
Average District (A)	Large Group	$238	$294	$487	$1,037
	Small Group	$596	$736	$1,217	$2,592
	Sensitivity Analysis 1—District Size				
Larger district (B)	Large Group	$254	$313	$518	$1,103
	Small Group	$634	$783	$1,294	$2,757
Larger district with more groups per instructor (C)	Large Group	$216	$267	$441	$940
	Small Group	$541	$668	$1,104	$2,351
Smaller district (D)	Large Group	$294	$364	$601	$1,280
	Small Group	$736	$909	$1,502	$3,201
Smaller district with shared instructor (E)	Large Group	$235	$290	$479	$1,021
	Small Group	$587	$724	$1,198	$2,551
Effect sizes			0.81	0.49	0.23

Note. Effect sizes represent the effects for the full sample and were chosen to represent the range of intervention effects reported in Table 1. See text for the assumptions incorporated in the cost-effectiveness ratios.