

Symposium Justification

Accelerating the Academic Achievement of Students with Learning Disabilities Research Initiative: On the Importance of Moderator Analysis in Intervention Research

The purpose of this symposium is to bring together researchers with interest in learning disabilities to explore learner characteristics associated with inadequate response to generally efficacious interventions in reading and mathematics. Presenters in this symposium were supported by the Accelerating the Academic Achievement of Students with Learning Disabilities Research Initiative (A3 Initiative) funded by the National Center for Special Education Research (NCSER) in the Institute of Education Sciences and all contributed to an upcoming special issue in *Exceptional Children*.

Most schools recognize that their core instruction does not help all learners. Recognition that poorly-achieving students with and without disabilities require a more appropriate education is implicit in many schools' support of Response to Intervention (RTI), or Multi-Tiered Systems of Support (MTSS); specifically, multiple tiers of instruction that (in principle) offer the structure and opportunity to apply programs of greater intensity when those of lesser intensity prove ineffective. Scores of published evaluations have documented that at-risk students participating in these programs have performed statistically significantly more strongly than comparable controls.

Nevertheless, a caveat is in order. None of the intensive reading and mathematics programs with which we are familiar claim robust effects. In each and every efficacy study of these programs, there were students nonresponsive to treatment. We estimate that such students represent between 10% and 25% of study samples. Moreover, most study authors did not explore why these children failed to benefit from programs that seemed to advantage many others. As a consequence, few interventions have been developed for them.

The purpose of this symposium is to explore learner characteristics associated with inadequate response to generally efficacious interventions in reading and mathematics through moderation analyses. One plausible explanation, or hypothesis, for inadequate response is the severity of students' pre-intervention academic deficits. For example, students with more severe initial math deficits may profit less from targeted intervention than those whose initial math deficits are less profound. Understanding the veracity of this *academic deficit hypothesis* is important for gauging the robustness of interventions, which is to say the degree to which they address a full range of at-risk students' performance; for identifying which interventions are more and less appropriate for students with most severe initial academic deficits; and for gaining insight into the dimensions or aspects of intervention that promote a relatively positive response among such students.

This symposium includes four intervention researchers who have selected one of their previously published intervention studies on which to conduct moderation analysis to examine whether at-risk students' pretreatment performance moderated their response to intervention. Christian Doabler from the University of Texas at Austin, and Lynn Fuchs from Vanderbilt University focused on mathematics; Mike Coyne from the University of Connecticut and Doug Fuchs from Vanderbilt University focused on reading. The symposium will be introduced by Sarah Brasiel, a NCSER project officer and will include a discussant, Greg Roberts, an expert on intervention research and moderation analyses from the University of Texas at Austin.

We hope this symposium inspires others to explore learner characteristics that moderate or interact with response to Tier 2 and Tier 3 interventions and strengthens researchers' curiosity about whether children's cognitive, behavioral, and social characteristics might eventually shape instructional programs that more successfully address the needs of students with very significant learning needs.

Abstract 1

Using Moderator Analysis to Identify the First-Grade Children Who Benefit More and Less from a Reading Comprehension Program: A Step towards Aptitude-by-Treatment Interaction

Doug Fuchs, Vanderbilt University

Background/Context:

Like many program developers, we typically set out to produce instructional regimens with robust effects, effects that benefit virtually all children of a certain age- or grade level and in an educational setting like the general classroom. Use of moderator analysis reflects recognition that, despite instructional researchers' hopes and ambitions, their programs can vary in value for different subgroups. The programs can provide substantial help to some while offering little benefit to others. In this study, we investigated word reading as a potential moderator because of its acknowledged role in beginning reading and because many children struggle at the word level.

Purpose:

We explored the efficacy of The Nashville Early Reading Program with two related purposes in mind. First, we wanted to know the value of its reading comprehension component. Towards this end, we conducted a component analysis of the program's decoding/fluency (DF) dimension and reading comprehension (COMP) dimension. That is, we contrasted a DF treatment condition against a DF+COMP condition to parse the added value of COMP. We also compared the two conditions when combined against controls. Our second purpose was to explore whether the efficacy of the DF and DF+COMP conditions interacted with our sample's pretreatment word reading performance. Put differently, our second study purpose was to use pretreatment word reading as a moderator to explore for whom our program was beneficial.

Setting/Participants

Research Design:

We recruited students from 73 first-grade classrooms and 13 schools in the Metro-Nashville Public Schools. At our request, teachers nominated their lowest readers ($n = 532$). We screened those for whom we obtained consents ($n = 389$) on reading measures creating a factor score for each child. On this score, we rank ordered the 389 children and eliminated the top 40%. This resulted in 223 study-eligible children who were randomly assigned to the larger reading study ($n = 132$), a smaller reading study ($n = 51$), and a wait list ($n = 40$) serving the two studies. The 132 students in the present reading study were then randomly assigned to DF, DF+COMP, or controls. There were 63 teachers of the 125 children in the final sample.

Intervention/Program/Practice:

Intervention were delivered through 1:1 tutoring outside the classroom, three times per week, for 21 weeks (63 sessions). Each DF and DF+COMP session was designed to last 45 min. The DF curriculum included 10 short narrative stories and scripted lessons. Each lesson had six parts: sight words, sound-symbol correspondence, decodable words, spelling, fluency building, and supplemental activities. The COMP part of each DF+COMP lesson included four activities: vocabulary, big idea, wrapping up, and supplemental activities. Vocabulary instruction involved the tutor and child previewing the meanings of unusual or otherwise difficult words prior to reading the story for the day. Big Idea included identifying: “the most important person” and the “most important thing that happened.” In Wrapping Up, students answered four question types based on the day’s story: factual questions, main-ideas questions, inference questions, and vocabulary questions.

Data Collection and Analysis:

To create more robust representations of domains of interest, we calculated either factor or composite scores to index Word Reading, Non-Word Reading, and Reading Comprehension performance at pre- and posttreatment across measures.

To examine the efficacy of the two treatment conditions and account for school and classroom clustering, we used multilevel models. Prior research and theory suggest that the efficacy of reading interventions differ depending on students’ pretreatment word reading. Thus, the final models are described as “moderation” models. The outcomes of interest were posttreatment Word Reading, Non-Word Reading, and Reading Comprehension. Our models had random and fixed components. Because the study sample came from different schools and classrooms, we used three-level models in all analyses in which students (at Level 1, $n = 125$) were nested within classrooms (at Level 2, $j = 56$), which were nested within schools (at Level 3, $k = 26$). Residual terms for the intercept were allowed at Levels 2 and 3. Independent residual terms were permitted for the three conditions and were retained where likelihood ratio tests indicated such terms were necessary compared to a mixed χ^2 distribution.

For the fixed effects, we examined treatment effects, moderated treatment effects with pretreatment Word Reading as a moderator, and two covariates in each model: (a) the relevant pretreatment score (e.g., controlling for pretreatment Non-Word Reading factor score when evaluating the posttreatment Non-Word Reading factor score), and (b) pretreatment Word Reading (the moderator of interest). Treatment effects were estimated by two orthogonal contrasts. The first compared both treatments together against controls. The second compared DF+COMP to DF. Due to our relatively small sample size, we used restricted maximum likelihood estimation with the Kenward-Roger degrees of freedom adjustment.

Findings/Results:

The general purpose of this randomized control trial was to explore the efficacy of the Nashville Early Reading Program on the word reading, non-word reading, and reading comprehension of first graders with poor reading skills. Following a 21-week intervention, the combined performances of the children in the two active treatment conditions, DF and DF+COMP, were statistically significantly stronger than controls on all outcomes including reading comprehension. Moderate-to-strong effect sizes accompanied these findings.

Whereas our first study aim was to explore the Nashville Early Reading Program’s efficacy and whether a COMP component added value to DF, a second purpose was to determine if the program affected first-grade participants *uniformly* or *differentially* such that one subgroup

of the sample benefitted more than another. We found evidence of the latter, although inconsistent evidence, perhaps because of relatively low power to detect such relations. Generally, children in DF and DF+COMP with weaker word reading outperformed similarly low-performing controls to a greater extent than treatment children with stronger word reading outperformed comparable controls.

Abstract 2

Racing Against the Vocabulary Gap: Matthew Effects in Early Vocabulary Instruction and Intervention

Michael Coyne, University of Connecticut

Background/Context:

Stanovich (1986) describe the Matthew effect as the “facilitation of further learning by a previously existing knowledge base that is rich and elaborated. A person with more expertise has a larger knowledge base, and the large knowledge base allows that person to acquire even greater expertise at a faster rate.” For vocabulary learning, the Matthew effect suggests that individual differences in overall vocabulary knowledge cause differential efficiency in acquiring new vocabulary during learning opportunities, and this differential vocabulary learning in turn causes further individual differences in vocabulary knowledge

Purpose

The purpose of this study was to consider the Matthew effect in students’ response to vocabulary intervention. We conducted a secondary analyses of data from two randomized control trials, an efficacy and replication study, evaluating the effects of an effective supplemental kindergarten vocabulary intervention to investigate whether individual differences in overall receptive vocabulary knowledge measured at the beginning of the year moderated the impact of the vocabulary intervention. A second purpose was to examine whether moderation would offset the benefits of providing Tier 2 vocabulary intervention within a MTSS, or RTI framework.

Setting/Participants:

Research Design

Our efficacy and replication studies took place in 48 elementary schools located in a mix of urban, suburban, and rural districts in the eastern and northwestern United States. All 284 kindergarten classes in these schools participated in the studies. Participants included students from two previous studies identified as at risk for language and learning difficulties who were randomly assigned in clusters to receive small group vocabulary intervention in addition to classroom vocabulary instruction (n=825) or to receive classroom vocabulary instruction only (n=781). A group of not-at-risk students (n=741) who received classroom vocabulary instruction served as a reference group.

Early in the school year we administered the Peabody Picture Vocabulary Test (PPVT) to all kindergarten students in participating classrooms (N=6360) to determine their entry level receptive vocabulary knowledge. Students who scored below a standard score of 92 (30th percentile) on the PPVT were considered at-risk for language and learning difficulties and eligible to participate in the study. We identified clusters of at-risk students within each

kindergarten classroom (n=6-8) and created subclusters of 3-4 students within each cluster that were matched on initial PPVT scores. Then, within each cluster (classroom), one subcluster was randomly assigned to the treatment group and the other was assigned to the control group. We also identified 3-4 students from each classroom who we considered typical achievers (n=741) to serve as a not-at-risk reference group. These students were chosen based on an initial PPVT score that fell between standard scores of 95-105 (37th and 67th percentiles).

Intervention:

To standardize Tier 1 classroom instruction, teachers in participating classrooms implemented the Elements of Reading Vocabulary program, a commercially available curriculum with evidence of efficacy which provided teachers with a set of 24 weekly vocabulary lessons and accompanying materials. Lessons were delivered in a 5-day sequence for about 20 minutes each day.

Students who were assigned to the intervention group received the Tier 2 supplementary vocabulary intervention in addition Tier 1 classroom instruction. The vocabulary intervention was implemented in small groups of 3-4 students outside of the classroom for 30 minutes per day, 4 days per week over the course of approximately 22 weeks. The vocabulary intervention was developed to align with the Tier 1 classroom vocabulary lessons and to emphasize features of effective instruction that have been demonstrated to enhance students' vocabulary learning. Interventionists provided explicit instruction with extensive teacher modeling and multiple opportunities for students to practice using the target vocabulary words introduced during Tier 1 classroom instruction.

Data Analysis:

To examine whether PPVT moderated the treatment effect, we fit a series of 4-level multilevel models on selected outcome variables of interest: the researcher developed Expressive Target Word Measure and Listening Comprehension Measure, and post-test PPVT. We treated students (level-1) as nested within sub-clusters (level-2), which were nested within clusters, which were nested within schools (level-4). The full model included treatment (at level-2), group mean centered fall PPVT score (at level-1), and the cross-level interaction between PPVT and treatment. PPVT was group mean centered at the student level (level-1). Therefore, to preserve the between cluster information contained in the original variable, PPVT was included at the higher levels as well, group mean centered at levels 2 and 3 and grand mean centered at level 4. For all models, we allowed the intercept to randomly vary across subclusters, clusters, and schools; however, we did not allow the any of the PPVT slopes to randomly vary across subclusters, clusters, or schools.

Findings/Results:

Results indicated that individual differences in overall receptive vocabulary knowledge measured at pretest on the PPVT did moderate the impact of our Tier 2 intervention on experimenter developed measures of expressive vocabulary learning and listening comprehension favoring students with higher initial vocabulary knowledge. However, the intervention was powerful enough to boost the target word learning of at-risk students, regardless of initial vocabulary knowledge, to levels well above the control group and comparable to their not-at-risk peers. In contrast, the ability of the Tier 2 intervention to close the gap between at-risk students and their not-at-risk peers on the listening comprehension measure also depended on

their pretest PPVT scores.

Conclusions:

Stanovich's (1986) conceptualization of the Matthew Effect continues to provide a powerful framework for not only understanding vocabulary development but also response to vocabulary instruction and intervention. Consistent with the Matthew Effect, results of this study provide additional evidence that individual differences in overall receptive vocabulary knowledge measured at the beginning of the school year are a strong predictor of response to vocabulary instruction and moderate the impact of supplemental vocabulary intervention.

Although we may not be able to design universal vocabulary instruction that will benefit all students equally, MTSS frameworks offer a promising approach for providing different levels of language and vocabulary support to students based on their level of risk. By screening kindergarten students at the beginning of the year and identifying those students who we predicted would be less responsive to universal, Tier 1 classroom vocabulary instruction, we were able to provide students with timely, intensive, small group intervention that helped to close gaps between them and their not-at-risk peers.

Abstract 3

Exploring the Relationship between Initial Math Skill and the Impact of a Kindergarten Math Intervention on Student Math Outcomes.

Christian Doabler, University of Texas at Austin

Background/Context:

Although significant advances have been made in the development and validation of early mathematics intervention programs, challenges remain in ensuring that the learning needs of all students are met. Emerging evidence suggests that not all students will respond to a generally effective intervention program. Within a Response to Intervention (RTI) or Multi-Tier System of Support (MTSS) model of service delivery, it is assumed that as students exhibit non-response instruction is altered to provide a more targeted intensive experience. One proposed mechanism to accomplish this goal is to gain a better understanding of what student level variables, including academic, cognitive, and behavioral, are associated with unresponsiveness to generally efficacious Tier 2 interventions

Purpose:

While the field's understanding of factors related to intervention response is growing, continued research is needed to expand our understanding of specific student-level predictors of responsiveness, such as initial skill, and how initial skill status interact with approaches, such as modifying group size, to increase instructional intensity. The purpose of this study was to expand previous work by conducting secondary analyses of a randomized control trial which investigated the efficacy of a Tier 2 kindergarten mathematics intervention, with a focus on two specific research questions: 1. Did students benefit differentially from the ROOTS intervention by initial early mathematics skill? 2. Did students benefit differentially from the two treatment conditions, large group versus small group, by initial mathematics skill?

Setting/Participants

Research Design:

This study analyzed data collected from two cohorts of a federally-funded efficacy project. A partially nested randomized controlled trial was employed randomly assigning kindergarten students within classrooms to one of three conditions: (2:1 group, 5:1 group, and a no-treatment control condition). Fourteen elementary schools and 69 classrooms participated in the study. In each participating classroom, all students with parental consent were screened in the late fall of their kindergarten year on early mathematics assessments. Composite scores on screening measures within each classroom were then rank ordered, and the 10 students with the lowest composite scores were randomly assigned to one of three conditions: (a) 2:1 intervention group, (b) 5:1 intervention group, or (c) a no-treatment control condition. A total of 1,550 kindergarten students were screened for eligibility. Of these students, 592 met eligibility criteria and were randomly assigned within classrooms to the two-student group condition ($n = 120$), the five-student group condition ($n = 295$), or the no-treatment control condition ($n = 177$).

Intervention/Program/Practice:

The Tier 2 mathematics program was designed to build students' proficiency in whole number concepts and skills. The 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. The intervention was designed to supplement core mathematics instruction and thus was delivered at times that did not conflict with students' core instruction in mathematics. The intervention is aligned with Common Core State Standards for Mathematics and recommendations from expert panels to focus on whole number concepts and skills. Specifically, instruction emphasizes concepts from the Counting and Cardinality and Operations and Algebraic Thinking domains of the CCSS-M. The instructional approach is drawn from principles of explicit and systematic mathematics instruction 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.

Data Collection and Analysis:

Students were administered five measures of whole number sense at pretest and posttest. These measures included a proximal assessment of whole number understanding that measured skills taught during the intervention, two distal measures of whole number sense, and a set of curriculum-based measures of discrete early number sense skills. A distal outcome measure was administered six months into students' first grade year.

We conducted two sets of analyses to address our research questions about differential response to the mathematics intervention and the group size based on initial screening scores. Previously, we examined overall effects of the intervention on mathematics achievement and the overall effects of group size with an analysis designed to account for students partially nested within small groups. Because the ROOTS groups, but not the unclustered controls, required a group-level variance, the analyses accounted for the potential heterogeneity among variances across conditions. In this study, we examined whether initial mathematics achievement based on mathematics screening scores predicted differential response to the ROOTS intervention or to the

different group sizes. Our statistical model included the pretest as a predictor of differential response and its interaction with the condition effect, either intervention versus control or small versus large groups. For the analysis of the group-size condition, the models used a standard analysis for nested data, rather than the partially nested models.

Findings/Results:

The purpose of this study was to explore the moderating role of initial skill on student outcomes. Results related to this research question indicated that students with lower initial skills showed greater benefit from the intervention on two out of six kindergarten outcome measures. We did not find a difference in this relationship by group size. Based on our previous work in mathematics and with the intervention, the results matched our study hypotheses. Overall, most students differed on the pretest by less than one point, yet by posttest, substantially greater numbers of students in the intervention condition scored within the average range.

Conclusions:

Findings from this study add to a number of studies that have attempted to examine the role of initial skill status and intervention response. Results from this study showed a general pattern of differential effect for students with initial lower skill on a general measure of mathematics achievement. We believe a continued focus on questions related to mediators and moderators of interventions, aspects of intervention intensity, and modifications to RTI service delivery models warrant additional empirical investigation and discussions amongst leaders and researchers in the field of special education.

Abstract 4

Does the Severity of Students' Pre-Intervention Math Deficits Affect Responsiveness to Generally Effective First-Grade Intervention?

Lynn Fuchs, Vanderbilt University

Background/Context:

When a randomized control trial (RCT) produces statistically significant effects favoring the learning outcomes of students who receive intervention over those who do not, that intervention is deemed validated. Validation suggests most students respond to the intervention, but few if any standard (non-individualized) interventions achieve universal response. Some students require adjustments to make intervention more intensive. Little is known about student characteristics that explain responsiveness.

Purpose:

The purpose of this study was to conduct a secondary analyses of a randomized control trial to assess whether effects of first-grade mathematics intervention apply across the range of at-risk learners' initial skill levels. We evaluated whether pre-intervention mathematics performance moderated intervention efficacy using a first-grade whole-number intervention designed to improve children's arithmetic.

Setting/Participants

Research Design:

Across four cohorts in four consecutive years, we worked with 40 schools and 227 first-grade classes. We relied on a latent class approach to screen the first cohort of children for at-risk and not-at-risk status by combining scores across math applications, concepts, calculations, and word-reading screeners into a single latent factor, used to designate risk status. For remaining cohorts, we used the first year cut-points for consistency. We enrolled into the study 608 at-risk (below the 40th percentile on the latent factor score) and 325 not-at-risk (above the 40% percentile) students from 227 classes. Then we randomly assigned at-risk students at the individual level, stratifying by pre-intervention math scores and classrooms, to three conditions: control, conceptual arithmetic intervention with speeded practice (A+SP), and conceptual arithmetic intervention with non-speeded practice (A+NSP). Of the 608 at-risk students, 213 were in the control group, 199 in the A+SP condition, and 196 in the A+NSP conditions.

Intervention/Program/Practice:

Intervention, which addresses the conceptual and procedural bases for emerging competence with arithmetic, occurred three times per week, 30 minutes per session, for 16 weeks. The program is organized in a manual with materials and guides that provide each lesson's structure, content, and language of explanation. Each lesson includes a 25-minute segment on the conceptual bases for arithmetic and 5 minutes of practice to support accurate arithmetic skill. Content and activities in the 25-minute segment were the same in the two practice conditions. Lessons are organized in five units. Unit 1 addresses basic number knowledge; Unit 2, arithmetic doubles; Unit 3, arithmetic sets 5 through 12 (e.g., the 5 set includes all problems with sums or minuends of 5); and Unit 4, 10s concepts. Students who advance quickly through most lessons also complete Unit 5, a review set of lessons. Instruction incorporates manipulatives and number lines.

Data Collection and Analysis:

For each outcome, the corresponding pre-intervention score was treated as the moderator of the intervention effect. Our data structure incorporated 3 levels: students (level 1) cross-classified by classrooms and teachers (level 2) and classrooms and teachers nested in schools (level 3). For each outcome, we ran unconditional multilevel models including a random effect for classrooms, teachers, and schools. In the final model, we retained non-zero random effects. For the 3 moderator models, pre-intervention variables were grand mean-centered before generating interaction variables, by multiplying the centered pre-intervention score by each contrast code.

Findings/Results:

Moderation analysis indicated no significant interactions between at-risk children's pre-intervention mathematics skill and either contrast on any outcome. Across pre-intervention math skill, effects favored intervention over control on arithmetic and transfer to double-digit calculations and number knowledge and favored speeded over non-speeded practice

Conclusions:

Findings suggest that first, efficacy for this 1st-grade math intervention, when conventionally framed as stronger outcomes for at-risk intervention students compared to at-risk control students, is strong. Findings also suggest the need for continued focus on moderator effects,

considering pre-intervention math skill and other student-level variables theoretically connected to the design of interventions. The eventual goal is to forecast which students will and will not respond to intervention. Until such methods are available, the focus turns to distinguishing students who have and have not adequately responded at the end of intervention. Our final conclusion is that research is needed to provide schools with technically strong post-intervention benchmarks for identifying students who are not adequately prepared to exit intervention and instead require more sustained, intensive services to avoid long-term failure.