Article

Hammill Institute on Disabilities

Journal of Learning Disabilities
2015, Vol. 48(3) 255–270

DOI: 10.1177/0022219413497097
journaloflearningdisabilities.sagepub.com
SAGE

# Examining the Effects of Linking Student Performance and Progression in a Tier 2 Kindergarten Reading Intervention

**Deborah C. Simmons, PhD[1], Minjung Kim, PhD[1], Oi-man Kwok, PhD[1], Michael D. Coyne, PhD[2], Leslie E. Simmons, MEd[1], Eric Oslund, PhD[1], Melissa Fogarty, PhD[1], Shanna Hagan-Burke, PhD[1], Mary E. Little, PhD[3], and D'Ann Rawlinson, MEd[3]**

**Abstract**

Despite the emerging evidence base on response to intervention, there is limited research regarding how to effectively use progress-monitoring data to adjust instruction for students in Tier 2 intervention. In this study, we analyzed extant data from a series of randomized experimental studies of a kindergarten supplemental reading intervention to determine whether linking performance on formative assessments to curriculum progression improved kindergarten reading outcomes over standard implementation. We were interested in whether specific progression adjustments would enhance the effects of supplemental reading intervention. Growth-mixture modeling using data from kindergarteners ($n$ = 136) whose intervention progression (e.g. repeat lessons, skip lessons) was adjusted every 4 weeks based on mastery data identified four latent classes characterized by unique profiles of curriculum progression adjustments. Multilevel analyses comparing the performance of students in the four classes with that of propensity matched groups whose intervention was not adjusted ($n$ = 101) indicated positive effects of curriculum progression for (a) students whose formative assessment performance exceeded 90% and received early and sustained lesson acceleration and (b) students who initially performed below 70% on assessments and who repeated early lessons and progressed to conventional implementation. Effects of curriculum adjustments for the two smallest groups were less clear.

**Keywords**

at risk/prevention, intervention, response to intervention

Response to intervention (RTI) in reading is an instructional framework that enables schools to provide the level and type of early intervention necessary to meet the needs of students with early academic risk. Grounded in prevention science and the evidentiary knowledge base in early reading, RTI in the primary grades provides a critical opportunity for early intervention, compared to later intervention, to differentially accelerate reading growth for children with early reading risk (Simmons et al., 2007; Vellutino, Scanlon, Small, & Fanuele, 2006; Vellutino, Scanlon, & Tanzman, 1998). Since RTI was authorized in 2004 (Individuals with Disabilities Education Improvement Act [IDEIA], 2004), schools have increasingly embraced RTI practices. Specifically, a recent report on implementation of IDEIA indicated that more than 70% of elementary schools from a nationally representative sample use a version of RTI in reading/language arts (Bradley et al., 2011). Typical features of RTI include (a) providing comprehensive high-quality classroom instruction, (b) screening all students to identify those who are at risk for future reading difficulties, (c) implementing evidence-based supplemental intervention for students who need additional support, and (d) intensifying the level of support for students who do not respond to supplemental intervention based on progress monitoring data (Bradley, Danielson, & Doolittle, 2005; National Center on Response to Intervention, 2010).

In their review of RTI as implemented over the past decade, Fuchs and Vaughn (2012) concluded that although some RTI components have been incorporated into schools' practice (e.g. screening), other essential features, including

[1]Texas A&M University, College Station, USA
[2]University of Connecticut, Storrs, USA
[3]University of Central Florida, Orlando, USA

**Corresponding Author:**
Deborah C. Simmons, Texas A&M University, College of Education and Human Development, College Station, TX 77845, USA.
Email: dsimmons@tamu.edu

progress monitoring, have failed to become widely adopted. In addition, these authors identified pressing issues that need to be addressed including research examining how to use progress monitoring to optimally inform instructional decisions.

The need for research to complement the existing RTI knowledge base was also highlighted in the RTI practice guide by Gersten et al. (2009). In this guide, an expert panel concluded that there is a strong research base to support providing "intensive, systematic [reading] instruction . . . in small groups who score below the benchmark score on universal screening" (p. 6). The efficacy of Tier 2 intervention notwithstanding, a significant number of children do not demonstrate adequate response and require Tier 3 intervention. Of relevance to this study was how to intensify the effects of Tier 2 intervention to reduce the number of students who may require Tier 3 intervention.

The purpose of the current study was to extend our understanding of the effects of using progress-monitoring mastery data to adjust the rate at which students progressed through a Tier 2 intervention program. Specifically, we conducted secondary analyses with extant data from a series of randomized controlled trials to discern whether there was an overall effect of adjusting curriculum progression based on student performance and to identify which unique student performance/curriculum progression adjustments were most effective. In this study, we were interested in determining whether different student performance/curriculum progression profiles (e.g., accelerated pacing, repeated lessons) enhanced the effects of a Tier 2 kindergarten reading intervention implemented conventionally (i.e., without adjustments). We hypothesized that examining the effects of specific adjustments would enable more effective engineering of interventions for students who do not respond adequately to standard implementation.

## *Conceptual and Empirical Framework for Using Data to Adjust Reading Interventions*

Conceptually, using student performance data to determine the pedagogical, instructional, and organizational variables that need to be manipulated is central to RTI (Lembke, McMaster, & Stecker, 2010; Stecker, Fuchs, & Fuchs, 2008). Empirically, however, few studies have examined the effects of data-adjusted interventions. The RTI practice guide recommends using curriculum-embedded mastery checks to monitor student progress (Gersten et al., 2009) because the experimental research reviewed used these types of assessments to monitor progress.

Curriculum-embedded measures (CEMs) are different from commonly used progress-monitoring materials that fall under the rubric of curriculum-based measurement (CBM). Unlike CBM probes, which assess student growth on a curriculum-independent general outcome (e.g., oral reading fluency), CEMs typically assess content that has been recently taught and measure mastery of specific skills. Thus, grounded in the theory of mastery learning (Bloom, 1971), CEMs are formative assessments that provide information to guide instructional modifications for individual students. CEMs are designed to complement general-outcome curriculum-based assessments by providing instructors information to make adjustments based on an individual student's performance.

In a study of first-grade students, Denton et al. (2010) investigated the effects of a supplemental intervention, Responsive Reading Instruction (RRI; Denton & Hocker, 2006). In the RRI condition, interventionists assessed students weekly to determine curriculum mastery and identify instructional needs. Based on mastery data, interventionists then selected instructional activities and curriculum adjustments for students, including progression through the curriculum, from a menu of options. Findings indicated that students in the RRI outperformed students in a control group. Of students in the business as-usual condition, 43% received a different intervention and 57% received no additional intervention. Although this study provides evidence supporting RRI's overall approach, it is difficult to identify which activities and adjustments contributed to improved student outcomes because RRI was implemented as a package and compared to multiple variations of business as usual.

In a study investigating the effects of individualized reading intervention for middle school students, Vaughn et al. (2011) used student performance on biweekly teacher-developed measures to determine skill mastery and guide instruction. They also used standardized monthly progress-monitoring measures to adjust intervention. Details on mastery levels for adjusting instruction were not discussed. Findings indicated no statistical or practical differences between the performance of students whose instruction was individualized based on performance data and that of students who received a standardized intervention. Students in both conditions outperformed students who received no additional intervention.

To date, the converging body of research supporting tiered interventions is based on standard protocol reading interventions (Coyne, Kame'enui, Simmons, & Harn, 2004; Vadasy, Sanders, Peyton, & Jenkins, 2002; Vaughn et al., 2011), in which students typically receive the same instructional content at the same pace and with the same emphasis over a designated period. In their review of Tier 2 research, Fuchs and Vaughn (2012) concluded that the existing evidence supports standardized over problem-solving or individualized intervention approaches in reading. Advantages cited to support standardized approaches included better ability to document what students have been taught, more efficient use of resources, and increased opportunities to document implementation fidelity.

Problem-solving or individualized interventions that use student data to dynamically adjust instruction inherently add complexity to scientific inquiry because of the many moving parts involved. Prior studies (Denton et al., 2010; Vaughn et al., 2011) that have compared standardized to individualized approaches used intervention packages that differed across multiple dimensions (e.g., content of instruction, instructional emphasis, opportunity to learn). These studies are pragmatically important given the relatively few comparative experimental studies that have been conducted. We propose, however, that there is middle ground between standard and individualized intervention protocols consisting of using a consistent approach to adjust a student's progression in an otherwise standard intervention. To understand the effects of adjusting intervention based on student performance, we sought to hold the intervention constant and manipulate a critical feature: progression in the curriculum based on student performance on CEMs.

### *Previous Studies and Current Research Focus*

The study used extant data from a series of experimental studies to (a) identify whether unique performance/progression profiles emerged for different groups of students when interventionists adjusted student progression through a Tier 2 intervention based on student mastery data from formative assessments and (b) examine the effects of different performance/progression profiles (e.g., accelerating lessons for students who score 90% or greater, repeating lessons for students who score less than 70%) on kindergarten reading outcomes. Data were drawn from four randomized trials that examined the effects of the Early Reading Intervention (ERI; Pearson/Scott Foresman, 2004), a Tier 2 kindergarten supplemental beginning reading intervention. Holding the intervention constant across studies enabled us to examine the effects of data-based adjustments.

In the first two studies (Coyne et al., 2013; Simmons et al., 2011), researchers compared a conventional implementation of ERI to school-designed Tier 2 interventions. In the first study, ERI groups outperformed school-designed intervention groups on a majority of measures. In the second study conducted in different schools, no statistically significant differences were found between ERI and school-designed intervention. Follow-up analyses indicated that ERI groups across both studies made similar gains; however, in the second study, groups receiving strong business-as-usual interventions experienced similar outcomes.

We hypothesized that a conventional implementation, where students progressed through the intervention at the same pace, may have impeded some students' progress and that some students could have benefitted from repeated lessons and others from an accelerated pace. Therefore, in two subsequent studies, we examined whether adjusting intervention progression based on student performance would intensify ERI effects (Coyne et al., in press; Little et al., 2012). In the Coyne et al. (in press) study, we compared ERI that was implemented conventionally to ERI that was systematically adjusted over the year based on student performance. Between-group differences on a range of early reading measures including letter names, letter sounds, word identification, and oral reading fluency were statistically significant, with effect sizes ranging from 0.29 to 0.76 favoring ERI with ongoing adjustments. In the Little et al. (2012) study, we compared an adjusted ERI implementation version in a district to an organized RTI implementation. Results showed no statistically significant group differences on any outcome measures; however, effect sizes, ranging from 0.35 to 0.59, revealed substantively important differences (Valentine & Cooper, 2003) favoring ERI with ongoing adjustments on multiple measures.

Although these findings provided preliminary evidence that adjusting progression through the ERI program benefitted kindergarten students at early reading risk, these initial between-group analyses were unable to identify whether specific performance/progression adjustments were more effective than conventional implementation. That is, were overall differences attributable to specific groups of students whose progression was adjusted, such as students who did not demonstrate mastery on formative assessments and repeated targeted lessons? In the present study, we used data from students who received either ERI implemented conventionally or ERI implemented with ongoing adjustments from the four previously conducted experimental trials to better understand whether using formative assessment profiles to adapt curriculum-progression would differentially improve end-of-kindergarten achievement. We defined "curriculum progression" as (a) accelerations of specific lessons, (b) repetitions of targeted lessons, or (c) conventional implementation with targeted skill review. Decision-rule criteria for accelerating, maintaining, or repeating lessons are described later in this article. Through secondary analyses of data from four intervention cohorts, our research addressed two primary questions:

1. What student performance/curriculum progression profiles (latent classes) were identified from different groups of students when ERI formative assessment data were used to accelerate, maintain, or repeat lessons?
2. Which student performance/curriculum progression profiles enhanced end-of-kindergarten outcomes over the conventional ERI implementation?

## Method

### *Participants*

Students included 237 kindergarteners who participated in one of four randomized control trials that examined the effects of variations of the ERI (Coyne et al., 2013; Coyne et al., in press; Little et al., 2012; Simmons et al., 2011). Students were divided into two groups: students who received ERI with ongoing adjustments (ERI-A) and students who received ERI implemented conventionally (ERI–matched peers). Data from students ($n$ = 136) who received ERI-A (Studies 3 and 4) were used to identify profiles of curriculum progression. Data from students ($n$ = 101) who received conventional ERI (Studies 1, 2, and 3) were identified as propensity matched peers (the details of the propensity score matching procedure are described in the analysis section).

Participants were from three states (CT, FL, and TX) and 26 public schools. Percentages of students at participating schools receiving free and reduced-cost lunch ranged from 72% to 82% in Texas, from 3% to 81% in Connecticut, and from 31% to 92% in Florida. Enrollments ranged from 278 to 985 students in the participating Texas schools, from 266 to 749 in the Connecticut schools, and from 401 to 832 in the Florida schools.

Across all four studies, researchers consulted with school personnel to identify children who (a) were considered in need of supplemental reading instruction based on school-administered reading assessments, (b) were at least 5 years of age, and (c) received reading instruction in English. Across the four cohorts, a two-step process was used to select students to be included in the analyses (descriptions of measures are reported later). For the first step, student inclusion criteria were (a) a score at or below the 36th percentile on the *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS; Good & Kaminski, 2002) Letter Naming Fluency (LNF) measure (i.e., fewer than 6 letters correctly named in 1 min) and (b) a score at or below the 37th percentile on the *Comprehensive Test of Phonological Processing* (CTOPP; Wagner, Torgesen, & Rashotte, 1999) Sound Matching (SM) subtest. For students who qualified in the first step, a second set of criteria were applied that included performance (a) at or below the 16th percentile on the Rapid Object Naming (RON) measure from the CTOPP or (b) at or below the 16th percentile on the Letter Identification from the *Woodcock Reading Mastery Test–Revised/Normative Update* (WRMT-R/NU; Woodcock, 1987/1998). These criteria were based on an earlier study (Simmons et al., 2011) that accurately identified students most at risk of reading difficulty. Specifically, we found that students who performed below the 16th percentile on RON or letter identification in this second-stage screening were more likely to be true positives and to need Tier 2 intervention compared to students who performed above those percentile levels.

Student demographics for ERI-A and ERI–matched peers, including baseline scores, are presented in Table 1. As illustrated, of the full sample of children who participated in intervention, 45% were girls and 55% were boys. The mean age of participants at pretest was 5.47 years. The sample was ethnically diverse, and 18.6% of students were identified by teachers as being English language learners. Approximately 13.5% of students were eligible to receive special education services.

## Intervention Procedures

In the following, we summarize the ERI (see Simmons et al., 2011, for fuller description), the ERI-A (for more detail see Coyne et al., in press), and the standard components across conditions.

*Early Reading Intervention.* ERI (Pearson/Scott Foresman, 2004) is a supplemental small-group intervention program that explicitly teaches phonologic, alphabetic, decoding, spelling, and sentence-reading skills to kindergarten students at early reading risk. The program includes 126 lessons structured in four units that progress from early phonemic and alphabetic skills to more complex regular and irregular word reading, spelling, and multiple sentence-reading skills. A typical 30-min lesson consists of seven activities, each designed to last 3 to 5 min and to actively engage students. Lessons provide explicit scripting for introducing, reviewing, and providing corrective feedback. New content is taught for 3 days and systematically reviewed. Students practice the new skill with the teacher and then apply it to discrimination or generalization tasks. The program includes four assessments, one at the end of each unit, and an instructional pacing chart.

Students in the ERI condition were initially grouped according to similar entry-level skills based on pretest assessment data. Starting at Lesson 1, interventionists progressed through the program sequentially. Students generally stayed with their initial groups throughout the school year; however, in a very few cases, groups were adjusted slightly at midyear. Approximately every 8 weeks, members of the research team administered ERI end-of-unit assessments to evaluate students' mastery of taught skills. End-of-unit assessment results were given to interventionists; however, unlike in the experimental condition, research team members did not make instructional recommendations based on student performance.

*ERI Adjusted Condition (ERI-A).* ERI-A used the same instructional lessons and materials as the ERI condition, but CEMs were added at the midpoint of each unit (approximately every 4 weeks) to evaluate students' mastery of skills taught during that period. Research team members met with interventionists to make instructional adjustments based on the student performance data.

Assessment data were used to determine whether to repeat or accelerate lessons as well as to determine which skills to review. Our decision rules for adjusting curriculum progression were based generally on Engelmann's (1996) theory of

**Table 1.** Student Demographic and Baseline Scores by Condition.

| Variable | ERI-A (*n* = 136) | | ERI–Matched Peers (*n* = 101) | |
|---|---|---|---|---|
| | *n* | % | *n* | % |
| Gender | | | | |
| Male | 62 | 46.6 | 45 | 44.6 |
| Female | 74 | 54.4 | 56 | 55.4 |
| Ethnicity | | | | |
| American Indian or Alaska Native | 1 | 0.7 | 0 | 0.0 |
| Black or African American | 40 | 29.4 | 21 | 20.8 |
| Hispanic or Latino | 50 | 36.8 | 40 | 39.6 |
| White | 40 | 29.4 | 38 | 37.6 |
| Other | 5 | 3.7 | 2 | 2.0 |
| Identified for special education | 14 | 10.3 | 18 | 17.8 |
| English language learner | 22 | 16.1 | 22 | 21.8 |
| | *M* | *SD* | *M* | *SD* |
| Age | 5.44 | 0.31 | 5.50 | 0.36 |
| Letter ID | 80.65 | 8.19 | 79.60 | 8.83 |
| Phoneme segmentation fluency[a] | 1.82 | 6.02 | 2.04 | 7.21 |
| Rapid object naming[b] | 6.08 | 2.19 | 6.26 | 2.74 |
| Letter naming fluency[a] | 1.20 | 1.77 | 1.79 | 1.94 |
| Interventionist's years of experience[c] | 11.85 | 7.41 | 11.44 | 9.70 |

*Note.* ERI = Early Reading Intervention; ERI-A = ERI with ongoing adjustments.
[a]Raw score. [b]Standard score. [c]Ten teachers, who taught in both ERI-A and ERI conditions, were excluded for this mean difference test.

mastery and acceleration and Bloom's (1971) theory of mastery learning. Engelmann's theory specifies that students who master material should demonstrate 90% accuracy on assessments of previously taught material. In studies that subscribe to Bloom's theory of mastery learning, the criterion is 80% (e.g., Guskey, 2009). To our knowledge, research has not definitely established a criterion for mastery, although there are generally accepted criteria (e.g., 80%). Our criterion for acceleration (90% on two consecutive mastery assessments) was rigorous as we wanted to ensure that students were highly proficient before allowing them to skip lessons. Generally, acceleration schedules involved students completing two of every three lessons until they reached the final unit, in which they were taught each lesson.

Students with performance ranges of 70% to 89% continued with the typical lesson progression, which included targeted review of letter–sounds already incorporated into lessons. Targeted review was designed to enable students to progress through the intervention. Performance below 69% triggered instructional adjustments including reteaching targeted lessons (i.e., based on incorrect letter–sound knowledge) and adding a modified letter–sound review activity to lessons. After repeated targeted lessons, students resumed typical lesson progression until the next mastery assessment. Because lesson repetition could involve a significant reduction in content coverage, we restricted this modification to students who scored less than 70% on content that had been previously taught.

When performance did not fall within the performance band of the other group members (<70%, 70%–89%, >90%), students were regrouped. Because of the small number of students at each school, instructional regrouping was limited. At the first regrouping occasion, 33.1% of the students were regrouped. Following the initial regrouping, the average number of students who were regrouped became stable and ranged from 7.4% to 13.2% ($M$ = 11.5%) for the four subsequent occasions.

*Standard intervention components.* To increase comparability between ERI and ERI-A conditions, several instructional components were standardized. Specifically, (a) both conditions implemented the ERI program (previously described) as a supplemental reading intervention, (b) groups in both conditions were composed of three to five students, and (c) interventionists in both conditions were asked to meet with their groups for 30 min, 5 days per week, over the course of the intervention period for an equivalent total number of sessions.

## Measures and Assessment Procedures

*Screening and pretests.* Approximately 6 weeks into the school years, students were screened using the measures previously described (i.e., RON, Letter ID, LNF, and SM). Students who met the screening criteria were also administered the *Peabody Picture Vocabulary Test–III* (PPVT-III; Dunn & Dunn, 1997).

Rapid automatized naming was measured by the RON subtest from the CTOPP (Wagner et al., 1999). In RON, students are required to name five objects repeated multiple times in a random pattern on two stimulus pages. The score is the total time it takes the student to name all stimulus objects on both pages. Alternate-form reliability coefficients as reported by the CTOPP manual range from .79 to .82 for students 5 to 7 years old.

Receptive vocabulary was measured using the PPVT-III. This test requires the student to choose from four examples the picture that represents a word presented orally by the examiner. Cronbach's alpha reliability coefficient is .95 for 5- to 6-year-old students, and retest reliability estimates range from .92 to .93.

Alphabetic knowledge was measured using the Letter ID subtest from the WRMT-R/NU and LNF from DIBELS (Good & Kaminski, 2002). Letter ID requires the student to orally provide the names of upper- and lowercase letters in various fonts presented on a stimulus sheet. Split-half reliability in first grade is .94; it is not reported for kindergarten students. LNF, in turn, measures the student's ability to fluently name letters presented on a stimulus page. Students are scored on how many letters they can correctly name in 1 min. Kindergarten alternate-form reliability is .88.

*Curriculum-embedded mastery assessments.* ERI mastery assessments are criterion-referenced measures that evaluate students' mastery of skills taught in each part of the program (i.e., alphabetic, phonemic, word reading, spelling). Whereas students in the ERI-A condition were assessed at the midpoint and the end of each of the four units of the program (approximately every 3–4 weeks), students in the ERI condition were assessed at the end of each unit (approximately every 7–8 weeks, depending on the number of lessons in the unit). Midunit and end-of-unit measures sampled skills introduced in each respective part. Both assessed three to four major skills. For example, the mastery assessment at the end of Unit 1 included three subtests that assessed phonemic and alphabetic tasks. Letter Names and Sounds require the student to provide the correct letter name and corresponding sound for each of 11 letters. First and Last Sounds in Words requires the student to produce the beginning and ending sounds in words presented both orally and with an accompanying picture. Finally, for the Letter-Sound subtest, the child is given tiles with the letters *d, f, l, m, p, r, s*, or *t* printed on them. The examiner presents a page with six pictures and three empty boxes below the pictures. After the examiner points to a picture and reads it aloud, the student is asked to put the letter tiles for the first and last sounds into their corresponding (i.e., first and third) boxes. Scores derived from these measures included the percentage of total points earned and the percentage of points for each skill. Results of end-of-unit mastery assessments were shared with interventionists in both conditions. In addition, both the midunit and end-of-unit performance of students in the ERI-A condition were shared with interventionists and subsequently used to inform instructional modifications.

*Posttests.* The following normative and criterion-referenced assessments were administered in April and May of kindergarten within 2 weeks of intervention completion. Students were given measures of phonological awareness, decoding, letter knowledge, word reading, reading fluency, and spelling.

*Phonological awareness* was assessed using the Blending Words (BW) and SM subtests from the CTOPP and Phoneme Segmentation Fluency (PSF) from DIBELS. The BW subtest requires a student to blend orally presented phonemes into the correct word. Cronbach's alpha is .88 for 5-year-old students and .89 for 6-year-old students. In SM, the student is provided with a target word and asked to indicate one of three words that begins with the same sound as the target word. The alpha coefficients for SM are .93 for both 5- and 6-year-old students. The DIBELS PSF measures the fluency with which a student can segment words comprising three or four phonemes. Students are scored on the number of phonemes they can correctly provide in 1 min. Alternate-form reliability is .88 in kindergarten.

*Letter knowledge* was assessed using the Supplementary Letter Checklist (SLC) subtest from the WRMT-R/NU. The SLC assesses the ability to provide both the letter names and sounds of letters presented on a stimulus sheet. No technical adequacy information is provided in the examiner's manual.

*Decoding* was assessed using the Nonsense Word Fluency (NWF) subtest from DIBELS and the Word Attack (WA) subtest from the WRMT-R/NU. On NWF, students are asked to decode nonsense words presented on a stimulus page. Students are scored on the number of correct individual or word-level sounds they can provide in 1 min. Kindergarten alternate-form reliability is .88. On WA, pseudowords are presented on a stimulus page, and the student is required to provide the correct pronunciation for the word. The median split-half reliability is .87.

*Word reading* was assessed using the Word Identification (WI) subtest from the WRMT-R/NU. Students are asked to correctly read as many words as possible of increasing difficulty. The median split-half reliability is .97.

*Reading fluency* was assessed using a passage titled "Mac Gets Well" (Makar, 1995). Students are scored on how many words they can read correctly in 1 min on the highly decodable passage. The measure was initially designed for early first-grade administration; however, Vadasy, Sanders, and Peyton (2006) reported a Cronbach's reliability of .93 for kindergarten students.

*Spelling* was assessed using the *Test of Written Spelling–4* (Larsen, Hammill, & Moats, 2005). Students are required to spell correctly 10 words presented orally. For 6-year-old students, alternate-form reliability is .86 and Cronbach's alpha is .87.

## Assessment Procedures

All assessments were administered by trained data collectors, including graduate students and members of the research team. Training consisted of two 4-hr sessions that involved a review of general assessment procedures, modeling of the specific test protocols, paired practice, and supervised independent practice of each test. Each data collector met the criterion of 90% accuracy in recording scores for a modeled administration of each measure. After data collection, two trained individuals scored each testing protocol independently.

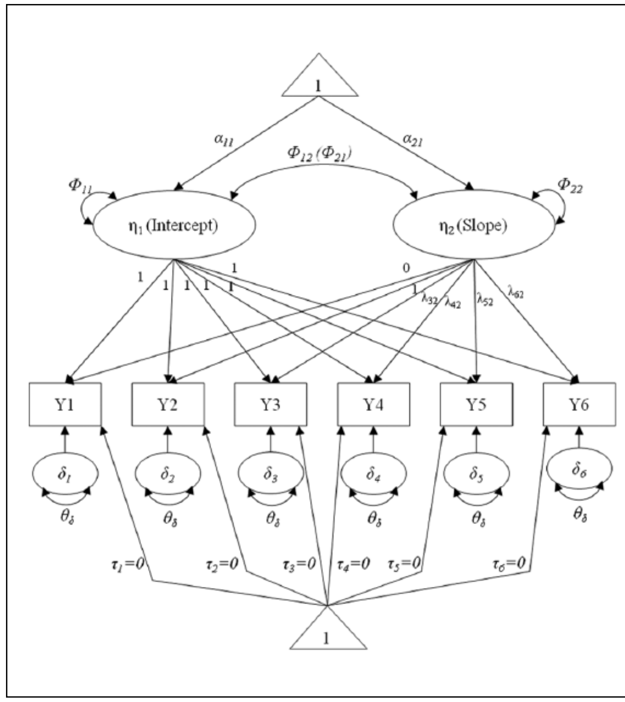


**Figure 1.** Latent growth model of the curriculum progression for Early Reading Intervention with ongoing adjustments students.

## Data-Analysis Procedures

We used a three-stage approach to analyze the data: the growth-mixture modeling stage, the propensity score matching stage, and the multilevel modeling stage. We first identified profiles of curriculum-progression adjustment and the subgroups of ERI-A students who fit each profile using growth-mixture modeling. We then identified matched groups of ERI students for each ERI-A subgroup using propensity score matching procedures. This enabled us to compare differences across end-of-kindergarten reading outcomes between each ERI-A subgroup and the corresponding matched ERI group using multilevel modeling. The details of these three stages are described below.

*Growth-mixture modeling.* We used growth-mixture modeling (GMM; B. Muthén, 2004) to identify profiles of curriculum-progression and the unobserved subgroups/latent classes of students in each profile. GMM is an exploratory approach that can flexibly estimate the potential of unobserved subpopulations of trajectories with class-specific parameter estimates, including the mean/overall average trajectory and interindividual and intraindividual differences in change for each of the unobserved subgroups/latent classes (Ram & Grimm, 2009).

In this study, GMM was used to identify the latent classes of curriculum-progression profiles for the ERI-A students, which were determined by the total number of lessons they accelerated (skipped) or repeated at the end of each curriculum part, with a total of six parts over the course of the study. The proposed latent growth model for the GMM analysis is presented in Figure 1. Unlike the simple linear growth model with two latent growth factors (i.e., intercept and slope factors) and fixed factor loadings, we estimated most of the factor loadings of the slope factor, which allows for estimation of a nonlinear growth model (Kwok, Luo, & West, 2010; Meredith & Tisak, 1990). Given the multilevel structure of the current data (i.e., repeated measures on curriculum progression nested within students and students nested within intervention groups), we adopted the "Type = Complex Mixture" routine in Mplus 6.11 (L. K. Muthén & Muthén, 1998–2011), which can adequately account for potential dependency by adjusting the standard errors of the parameter estimates. To compare different class solutions (Nylund, Asparouhov, & Muthén, 2007), the final number of latent classes was determined by using the generally recommended information criteria, namely, Bayesian information criterion (BIC) and Akaike information criterion (AIC).

*Propensity score matching.* After confirming the final number of latent class profiles of curriculum progression for the ERI-A students, we identified matched groups of ERI students for each of the ERI-A subgroups to be able to test the potential effect of adjusting the curriculum progression on end-of-kindergarten reading outcomes. Because data were

**Table 2.** Information Criteria for Various Class Solutions.

| | One-Class Model | Two-Class Model | Three-Class Model | Four-Class Model |
|---|---|---|---|---|
| AIC | 4032.03 | 3864.62 | 3864.90 | 3775.15 |
| BIC | 4075.72 | 3928.69 | 3949.37 | 3880.00 |
| ABIC | 4028.27 | 3859.10 | 3857.63 | 3766.12 |

*Note.* ABIC = sample-size-adjusted Bayesian information criterion; AIC = Akaike information criterion; BIC = Bayesian information criterion.

drawn from a series of studies that compared various interventions across several years, it was not possible to randomly assign students to each condition (i.e., not all the students in our data were randomly assigned to either ERI-A and ERI; in particular, for the first cohort, only ERI was available). The data set containing 136 ERI-A students was separated into four different classes/subgroups based on the growth-mixture analysis. Each of the four ERI-A subgroups was then separately combined with a data set of 101 ERI condition students to estimate the propensity scores. Propensity scores, defined as a student's probability of being in the corresponding ERI-A condition, were estimated based on the 14 covariates measured at the baseline, including both demographic variables (i.e., gender, age, limited English proficiency, ethnicity, special education status, and sites) and the baseline achievement scores (i.e., blending words, sound matching, rapid object naming, letter naming fluency, PPVT, letter identification, word identification, and word attack). By mimicking a randomized experimental design, which controls for any potential preexisting differences on these variables between the four different groups of ERI-A and ERI children, propensity matched samples enable us to estimate the true difference between ERI-A and ERI conditions on the target reading outcome measures.

The propensity scores were computed utilizing logistic regression analysis by using the above 14 covariates as predictors (see Note 1). Based on the propensity scores, the ERI-A and ERI students were then optimally matched within each of the four ERI-A conditions. We used the one to one matching scheme (Ming & Rosenbaum, 2001) with the PROC ASSIGN routine in SAS (v. 9.3) to create the optimally matched groups. We also used caliper distance of 0.25, indicating that any pair of matched groups did not differ by 0.25 standard deviations in their propensity scores. We then checked if the matching result provided good balance between the ERI-A and ERI groups. In other words, for two groups to be balanced, they should not differ on the 14 variables because they were matched based on having a similar probability of being assigned. We found that the four matched data sets had good balance.

*Multilevel modeling.* We adopted multilevel modeling to test the differences between each ERI-A subgroup and the corresponding matched ERI group on end-of-kindergarten reading outcomes. Because nonindependent observations were used because of the nesting structure in our data (i.e., students nested within different intervention groups), multilevel modeling (Hox, 2002) was applied to analyze the data, as it takes nonindependence into account. All multilevel models were analyzed using the MIXED routine in IBM SPSS V20.0. Restricted maximum likelihood was used to estimate all the models, and two-tailed tests were used to evaluate intervention effects. The corresponding equations of the multilevel model are presented in Appendix A.

In addition to the tests of significance (i.e., *t* test) of the difference between the ERI and the ERI-A conditions on the posttest measures, we used Hedges's (2007) proposed effect size ($\delta_w$) to evaluate the practical significance of differences in cluster-randomized designs. The equation of $\delta_w$ is presented in Appendix B. To control for the potential inflated Type I error rate because of the multiple comparisons, we used the Benjamini–Hochberg (BH) correction (Benjamini & Hochberg, 1995), as recommended by the What Works Clearinghouse (WWC, 2008). In accordance with the approach recommended by the WWC, we report comparisons that are statistically significant and also interpret treatment effects that are greater than .25 but not statistically significant as "substantively important" (p. 23).

## Results

### *Identifying the Number of Latent Class Profiles of Student Performance/Curriculum Progression*

To identify the optimal number of latent class profiles linking student performance and curriculum progression, we estimated one-class, two-class, three-class, four-class, and five-class models separately. We did not obtain the converged solution from the five-class model, which implied that the four-class solution was sufficient to account for heterogeneity in the curriculum progression trajectories (Enders & Tofighi, 2008; Nylund et al., 2007). The corresponding AIC and BIC values for different class solutions are presented in Table 2. By comparing it with other class solutions, the four-class solution resulted in the lowest values for both AIC and BIC, which also indicated that the four-class model outperformed other models and fit our data the best.

Figure 2 presents the four latent profiles based on the four-class solution across the six measurement points.

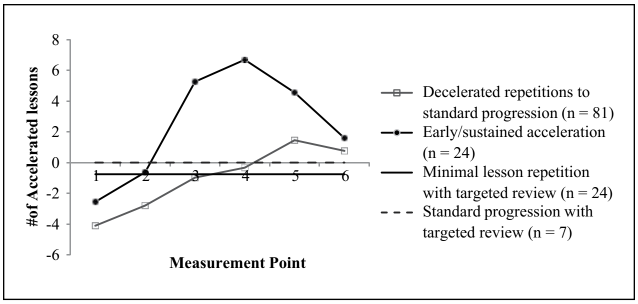


**Figure 2.** Four latent classes from growth-mixture modeling using the number of accelerations across six time points.

Among the total 136 students in the ERI-A condition, a majority ($n$ = 81) belonged to a class we labeled the "decelerated repetitions to standard progression" group. In this class, students repeated four lessons at the first assessment point. The pattern of repetitions steadily decreased, indicating improved performance on CEMs. At the fourth assessment point, students in this class were on track for typical lesson progression. A total of 24 students were in the "early/sustained acceleration" group. The performance/progression profile of this class indicated a few repetitions at the first assessment point, after which students consistently performed at 90% or higher on CEMs and accelerated as many as seven lessons per unit. Another 24 students belonged to the "minimal lesson repetition with targeted review" group, which repeated, on average, one lesson every measurement point. Finally, 7 students were in the "standard progression with targeted review" group, indicating they consistently scored between 70% and 89% on mastery measures and received the standard implementation schedule. Table 3 shows the corresponding parameter estimates (i.e., the means of the latent growth factors and the estimated factor loadings) of each of the four classes, as presented in Figure 2; Table 4 shows the means and standard deviations of the poststudy reading outcomes for all four classes respectively.

## *Matching Peers in the ERI Condition Using Propensity Score Analysis*

We used the propensity score matching procedure (Rosenbaum & Rubin, 1983) to select the comparison units (i.e., ERI students) that were similar in terms of the sample characteristics. To produce the propensity score, we used 14 covariates, including demographic variables (i.e., gender, age, English language learner status, special education status, ethnicity, and site) and baseline reading skills (i.e., blending words, sound matching, rapid object naming, letter naming fluency, vocabulary, letter identification, word identification, and word attack). Based on the propensity score, 75 ERI students were matched to the group of "decelerated repetitions to standard implementation" ($n$ = 81 students), whereas 22 ERI students were matched to the "early and sustained acceleration" group containing 24 ERI-A students. Similarly, 15 ERI students were matched to the "minimal lesson repetition with targeted review" group ($n$ = 24 ERI-A students), whereas two students in ERI were matched to the "standard progression with targeted review" group, which consisted of only 7 ERI-A students.

We conducted the sensitivity analyses (Robins, Rotnitzky, & Scharfstein, 1999) using $t$ tests and chi-square tests to check the quality of the matched groups by examining the possible initial group difference between the matched ERI and the corresponding ERI-A conditions on all the demographic variables, as well as the baseline reading skills. No statistically significant difference was found on any of these variables between the two conditions across all four classes. Hence, the matching procedure was successfully employed. As a result, if any significant difference on end-of-kindergarten outcomes was found, we could attribute it to being the result of the curriculum progression adjustments.

## *Testing Difference Between ERI-A and ERI on End-of-Kindergarten Reading Outcomes*

Table 5 presents the results of the multilevel analyses for testing the effect of performance/progression adjustments by the four latent class profiles separately, and the standardized effect size, $\delta_w$, for each comparison and outcome measure. As illustrated, within the comparisons between the ERI-A "decelerated repetitions to standard progression" group and the corresponding matched ERI group ($n$ = 156), the ERI-A group scored significantly higher than the comparison group on the following posttest measures after adjusting for the covariates and controlling for the comparison-wise Type I error rate using the BH correction: WRMT-R/NU SLC-Sound, PSF, Word ID, and Oral Reading Fluency. Effect sizes for these measures were also substantively important (effect sizes greater than .25) along with the effect size for WRMT-R/NU SLC-Name and CTOPP BW.

Within the comparisons between the ERI-A "early/sustained accelerations" group and the corresponding matched ERI group, the ERI-A scored significantly higher on five outcome measures (i.e., WRMT-R/NU SLC Letter and Sound, CTOPP SM, CTOPP BW, and Word ID) after controlling for the covariates and Type I error rates by BH correction. Results showed that the ERI-A students in general outperformed their matched ERI counterparts with substantively important effects on all measures ($\delta_w$ range from 0.36 to 1.25).

For the "minimal repetition with targeted review" group ($n$ = 35) who consistently performed between 70% and 89% on mastery assessments, no statistically significant between-group differences were found. Effect sizes varied greatly by

**Table 3.** Parameter Estimates for Growth-Mixture Analysis With Four-Class Solution.

| | Decelerated Repetitions to Standard Progression (*n* = 81) | | | Early/Sustained Accelerations (*n* = 24) | | | Minimal Lesson Repetition With Targeted Review (*n* = 24) | | | Standard Progression With Targeted Review (*n* = 7) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Estimate | *SE* | *p* Value | Estimate | *SE* | *p* Value | Estimate | *SE* | *p* Value | Estimate | *SE* | *p* Value |
| $\alpha_{11}$ | 4.102 | 0.600 | .000 | 2.553 | 0.592 | .000 | 0.756 | 0.418 | .071 | 0.932 | 0.816 | .254 |
| $\alpha_{21}$ | –1.309 | 0.541 | .016 | –1.922 | 0.851 | .024 | –0.294 | 0.322 | .362 | –0.169 | 0.225 | .455 |
| $\lambda_{32}$ | 2.387 | 0.984 | .015 | 4.067 | 1.858 | .029 | 2.550 | 2.394 | .287 | –9.961 | 5.816 | .087 |
| $\lambda_{42}$ | 2.884 | 1.223 | .018 | 4.811 | 1.972 | .015 | 13.626 | 7.458 | .068 | 2.077 | 0.857 | .015 |
| $\lambda_{52}$ | 4.247 | 1.852 | .022 | 3.701 | 1.530 | .016 | 4.795 | 2.421 | .048 | 3.001 | 1.410 | .033 |
| $\lambda_{62}$ | 3.720 | 1.608 | .021 | 2.159 | 0.942 | .022 | 3.921 | 1.772 | .027 | 3.286 | 1.582 | .038 |

outcome measure, with some favoring the ERI-A group (spelling $\delta_w$ = 0.50, blending words $\delta_w$ = 0.36) and others the ERI condition (word identification $\delta_w$ = 0.77, oral reading fluency $\delta_w$ = 1.16). The pattern of findings was similar for the "standard progression with targeted review" group, with no statistically significant differences. Letter-sound knowledge, phonemic segmentation fluency, blending words, and spelling effect sizes exceeded .25 and favored the ERI-A condition. Nonsense word fluency, word attack, and word identification effect sizes were greater in the ERI conventional implementation.

## Discussion

To date, few studies have examined the impact of intensifying interventions by adjusting students' progression through an intervention curriculum in response to their mastery of taught content (e.g., Denton et al., 2010). Although adjusting intervention in response to content mastery is recommended in research-to-practice guides and articles (Gersten et al., 2009; Mellard, McKnight, & Jordan, 2010), to our knowledge no studies have specifically compared an intervention implemented conventionally with the same intervention implemented with ongoing adjustments based on student response. The analyses presented in this study were motivated by findings from two randomized trials that provided encouraging evidence that adjusting curriculum progression based on student mastery data resulted in improved outcomes with effect sizes that ranged from 0.29 to 0.76. The primary findings from previous studies, however, did not identify which curriculum adjustments were associated with enhanced achievement.

Using GMM, we identified and examined student performance/curriculum progression profiles that emerged for different groups of students when mastery assessment data were used to make instructional adjustments. We compared subgroups of students from each latent profile to propensity matched peer groups to investigate whether the performance of students receiving the ERI kindergarten reading intervention with ongoing curriculum adjustments differed from that of students receiving the ERI intervention implemented conventionally. Findings indicated that disaggregating the overall effects of instructional adjustments from prior studies offers important information for intensifying interventions. Analyses identified four profiles that characterized students' progression through ERI-A. Two of these profiles indicated statistically significant advantages over conventional intervention implementation with effects sizes ranging 0.13 to 1.25. The pattern of performance in the two smallest profile groups was less clear.

### *The Effects of Distinctive Student Performance/Curriculum-Progression Profiles*

*Decelerated repetitions to standard implementation.* The largest class comprised 81 (60% of ERI-A participants) students who scored, on average, below 70% on early mastery checks (e.g., 68% on Point 1 and 76% on Point 3) followed by improved performance (e.g., 81%) at the fifth adjustment period. Results of multilevel analyses indicated that students in this profile outperformed 75 matched peers from the ERI condition on 4 of 10 outcomes with effect sizes ranging from 0.13 to 0.48. (Students in this class were the lowest performing of the four groups at posttest.) Findings from this profile suggest that adjusting instruction to develop mastery enhances learning outcomes over conventional progression through ERI; however, we consider the differences of moderate magnitude. On average, students in this group repeated four lessons at the first adjustment period with steady decreases in the number of lesson repetitions. We hypothesize that the initial repetitions were necessary for students to develop building-block skills (i.e., new letter–sound correspondences), which they then applied to more complex skills such as word identification ($\delta_w$ = 0.45) and oral reading fluency ($\delta_w$ = 0.48).

*Early and sustained acceleration.* A second class was the group of 24 students who received ERI-A with early and sustained lesson acceleration. Consistently, the average performance of this group was greater than 90% on all assessments, enabling interventionists to compress the curriculum (Reis, McCoach, Little, Muller, & Kaniskan, 2011) and skip

**Table 4.** Descriptive Statistics for Student Achievement Outcomes by Latent Class and Condition.

| | Decelerated Repetitions to Standard Progression | | Early/Sustained Accelerations | | Minimal Lesson Repetition With Targeted Review | | Standard Progression With Targeted Review | |
|---|---|---|---|---|---|---|---|---|
| Measure | ERI-A ($n$ = 81) | Matched Group ($n$ = 75) | ERI-A ($n$ = 24) | Matched Group ($n$ = 22) | ERI-A ($n$ = 24) | Matched Group ($n$ = 11) | ERI-A ($n$ = 7) | Matched Group ($n$ = 2) |
| Alphabet knowledge | | | | | | | | |
| WRMT-R/NU Letter Name Checklist[a] | 24.27 (4.88) | 22.48 (6.61) | 26.88 (2.35) | 22.09 (7.56) | 27.29 (1.99) | 22.64 (5.84) | 26.29 (1.70) | 27.00 (1.41) |
| Letter sound knowledge | | | | | | | | |
| WRMT-R/NU Letter Sound Checklist[a] | 24.00 (6.34) | 21.27 (8.38) | 28.54 (3.50) | 21.95 (9.25) | 27.71 (3.41) | 21.36 (9.23) | 28.00 (4.08) | 26.50 (0.71) |
| Phonemic awareness | | | | | | | | |
| DIBELS Phonemic Segmentation Fluency[a] | 39.15 (21.62) | 32.04 (19.85) | 44.00 (17.61) | 36.45 (18.71) | 48.42 (16.57) | 39.09 (17.52) | 24.86 (23.53) | 26.50 (19.09) |
| CTOPP Sound Matching[b] | 8.99 (2.49) | 8.83 (1.73) | 10.33 (2.01) | 9.09 (1.74) | 9.21 (2.25) | 8.73 (1.90) | 9.00 (2.08) | 9.00 (0.00) |
| CTOPP Blending Words[b] | 10.35 (2.67) | 9.70 (2.57) | 12.71 (1.65) | 10.36 (2.61) | 11.58 (2.12) | 10.36 (2.25) | 11.29 (2.29) | 9.50 (3.54) |
| Word attack | | | | | | | | |
| DIBELS Nonsense Word Fluency[a] | 27.80 (14.02) | 24.07 (14.87) | 34.50 (21.32) | 24.95 (16.13) | 29.83 (13.75) | 23.18 (19.57) | 25.71 (13.45) | 34.50 (3.54) |
| WRMT-R/NU Word Attack[b] | 105.47 (11.35) | 103.35 (11.22) | 112.67 (8.76) | 106.52 (9.91) | 110.75 (8.14) | 102.27 (11.26) | 105.14 (13.73) | 109.50 (7.78) |
| Word identification | | | | | | | | |
| WRMT-R/NU Word Identification[b] | 103.10 (12.89) | 98.15 (12.76) | 114.50 (8.23) | 101.27 (12.89) | 109.25 (11.47) | 101.73 (13.30) | 106.86 (10.12) | 108.00 (5.66) |
| Spelling | | | | | | | | |
| *Test of Written Spelling*[b] | 85.38 (6.09) | 84.28 (5.93) | 92.25 (9.92) | 87.36 (6.43) | 88.96 (7.11) | 84.09 (5.24) | 87.43 (9.09) | 83.00 (2.83) |
| Oral reading fluency | | | | | | | | |
| "Mac Gets Well"[a] | 9.51 (8.06) | 7.61 (6.36) | 15.33 (11.83) | 9.09 (7.48) | 13.75 (10.56) | 7.36 (7.76) | 8.14 (4.22) | 9.50 (6.36) |

*Note.* All values are mean scores with standard deviations in parentheses. CTOPP = *Comprehensive Test of Phonological Processing*; DIBELS = *Dynamic Indicators of Basic Early Literacy Skills*; ERI-A = Early Reading Intervention with ongoing adjustments; WRMT-R/NU = *Woodcock Reading Mastery Test–Revised/Normative Update*.
[a]Raw score. [b]Standard score.

**Table 5.** Group Difference Testing Between ERI-A and ERI.

| Measure | Decelerated Repetitions to Standard Progression ($n$ = 156)[a] | | | Early/Sustained Accelerations ($n$ = 46)[b] | | | Minimal Lesson Repetition With Targeted Review ($n$ = 35)[c] | | | Standard Progression With Targeted Review ($n$ = 9)[d,e] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\gamma^f$ | *p* Value | $\delta_w$ | $\gamma^f$ | *p* Value | $\delta_w$ | $\gamma^f$ | *p* Value | $\delta_w$ | $\gamma^f$ | *p* Value | $\delta_w$ |
| Alphabet knowledge | | | | | | | | | | | | |
| WRMT-R/NU Letter Name Checklist[g] | 1.55 | .11 | 0.34 | 3.55 | .04†† | 0.79 | 0.87 | .65 | 0.19 | –0.43 | .76 | –0.10 |
| Letter sound knowledge | | | | | | | | | | | | |
| WRMT-R/NU Letter Sound Checklist[g] | 2.68 | .03†† | 0.48 | 6.72 | .02†† | 1.21 | 0.19 | .96 | 0.03 | 2.84 | .28 | 0.51 |
| Phonemic awareness | | | | | | | | | | | | |
| DIBELS Phonemic Segmentation Fluency[g] | 12.11 | .00†† | 0.47 | 9.15 | .14 | 0.36 | 4.28 | .56 | 0.17 | 7.45 | .79 | 0.29 |
| CTOPP Sound Matching[h] | 0.25 | .51 | 0.13 | 1.84 | .02†† | 0.97 | –0.68 | .57 | –0.36 | 0.01 | 1.00 | 0.01 |
| CTOPP Blending Words[h] | 0.94 | .04 | 0.45 | 2.23 | .01†† | 1.05 | 1.50 | .17 | 0.71 | 2.12 | .13 | 1.00 |
| Word attack | | | | | | | | | | | | |
| DIBELS Nonsense Word Fluency[g] | 5.20 | .04 | 0.22 | 11.50 | .11 | 0.48 | –13.44 | .13 | –0.56 | –16.80 | .44 | –0.70 |
| WRMT-R/NU Word Attack[h] | 2.27 | .18 | 0.22 | 4.64 | .18 | 0.44 | –0.09 | .99 | –0.01 | –6.68 | .61 | –0.64 |
| Word identification | | | | | | | | | | | | |
| WRMT-R/NU Word Identification[h] | 5.09 | .04†† | 0.45 | 14.24 | .01†† | 1.25 | –8.82 | .11 | –0.77 | –2.96 | .72 | –0.26 |
| Spelling | | | | | | | | | | | | |
| *Test of Written Spelling*[h] | 1.48 | .15 | 0.23 | 5.30 | .09 | 0.81 | 3.25 | .35 | 0.50 | 4.70 | .51 | 0.72 |
| Oral reading fluency | | | | | | | | | | | | |
| "Mac Gets Well"[g] | 3.58 | .00†† | 0.48 | 7.98 | .06 | 1.06 | –8.70 | .12 | –1.16 | –0.42 | .91 | –0.06 |

*Note.* Reference group is ERI standard implementation group. Effect size was calculated as $\frac{\gamma}{\sqrt{\sigma^2}}$, where $\gamma$ is the parameter estimate for the intervention effect and $\sigma^2$ is the within-group variance from unconditional multilevel model with all 237 students. CTOPP = *Comprehensive Test of Phonological Processing*; DIBELS = *Dynamic Indicators of Basic Early Literacy Skills*; ERI = Early Reading Intervention; ERI-A = ERI with ongoing adjustments; WRMT-R/NU = *Woodcock Reading Mastery Test–Revised/Normative Update*.
[a]75 matched ERI group students. [b]22 matched ERI group students. [c]11 matched ERI group students. [d]2 matched ERI group students. [e]Given the small sample size ($n$ = 9), we controlled only for the students' baseline knowledge for the corresponding measure without any other covariates. [f]Gamma ($\gamma$) is the difference between ERI-A and ERI comparison groups on the posttest measure while holding the effect of other variables as constant. Positive values indicate that ERI-A group, on average, scored higher on the posttest measure than the comparison group. [g]Raw score. [h]Standard score.
†† Significant effect after Benjamini–Hochberg correction.

targeted lessons that reviewed and applied skills. Students in this group significantly outperformed their matched peers on five outcomes, and effect sizes ranged from 0.36 to 1.25 on the full complement of outcome measures.

A premise of acceleration is that by eliminating skills or lessons that do not need to be taught, more advanced content can be incorporated enabling students to acquire more complex skills. Because we worked with a fixed set of 126 lessons, we may inadvertently have constrained acceleration effects. The deceleration in the curriculum progression graph (see Figure 2) indicates that as students approached the end of the program, lesson acceleration decreased, an artifact of both the intervention (e.g., there were no more lessons to accelerate through) and the desire to control for total number of lessons that students received across conditions. Based on these findings, further research on the benefits of lesson acceleration for students in Tier 2 intervention is clearly warranted.

*Effects of minimal repetition and targeted lesson review.* Two profiles, 3 and 4, closely resembled standard implementation. The third profile reflected a curriculum progression that hovered slightly below zero, indicating a progression that averaged less than one repetition per curriculum unit with targeted skill review. Group performance averaged between 82% and 87% at all assessment points. The fourth trajectory included only seven students, and, on average, performance on part tests improved over time from 68% at the first point to 81% at the fifth checkpoint. Curriculum progression included repetitions, standard progression, and accelerations.

Findings from both trajectories indicated no statistically significant differences between the performance of adjusted groups and those of peers who received the conventional intervention. Effect sizes, however, revealed a range of performance differences that varied between ERI-A and ERI groups. These were the smallest profile groups, and further research is warranted to examine performance variability.

### *Limitations of the Study*

We want to acknowledge several limitations of this study. First, the study was based on studies of only one supplementation intervention (i.e., ERI) in one grade, kindergarten. To understand whether these types of mastery-based curriculum-progression adjustment profiles replicate and influence reading achievement, research is needed with different interventions across a range of grades. A second limitation is the method we used to operationalize RTI, leaving unanswered the following questions: Were the procedures and decision rules used to accelerate students too stringent? Was the decision to provide targeted review of letter–sound correspondences as opposed to other skills too restricted or the interval too broad?

Our cut points of 90% for acceleration and below 70% for lesson repetition, although instructionally defensible, also require further study. Moreover, the decision to provide targeted review and no lesson repetition for students performing in the range of 70% to 89% needs further examination. This is a broad range of performance, and our targeted review may need to be intensified or alternate methods of intensifying intervention explored. To our knowledge, however, there are no studies that provide definitive decision rules on how to adjust intervention in response to student performance.

A final limitation relates to balancing science and the realities of intervention. Our intent in the ERI-A condition was to group students with similar skill profiles and adjust lesson progression accordingly. Our efforts were, at best, reasonable approximations. In some schools, there were too few groups to appropriately place students with similar skills and needs. In other cases, when a student's performance fell between two groups, we may have erred by placing a student in a group that was too high or too low, thus influencing his or her performance.

## Conclusion and Implications

There is emerging but incomplete scientific evidence for how to intensify Tier 2 intervention for students at risk of reading difficulties (Gersten et al., 2009). Findings from the present study indicate that two types of progression adjustments based on student response to mastery assessments differentially improved some but not all reading and reading-related outcomes. The strongest benefits of curricular adjustments, based on statistical and practical effects, were for students who scored at or above 90% correct on mastery assessments and who were accelerated through an early reading intervention by skipping targeted lessons. Although their matched peers in the unadjusted condition also achieved positive outcomes, findings suggest that we can enhance intervention efficacy by using data to expedite lesson progression. Thus, we found a moderate benefit of curriculum adjustments for students who scored below 70% correct on early mastery checkpoints and repeated key lessons compared to matched peers in the unadjusted condition. Findings indicated a steady improvement on mastery assessments from 68% to 81% for students who repeated early lessons (covering foundational skills) with decelerating lesson repetition necessary as the program progressed. Performance profiles that most closely resembled standard implementation yielded non–statistically significant differences but mixed effect sizes.

Although significant strides have been made in addressing the needs of children who are likely to experience reading difficulty, findings from this study suggest the need for further research that addresses the intervention, grade, and

RTI limitations identified. Specifically, we are in need of answers to such questions as these: Will the effects of acceleration and lesson repetition based on student performance replicate in different intervention programs that use different forms of assessment? What are the most effective methods for improving the performance of students who achieve mastery of 70% to 89% of intervention content?

In conclusion, we are reminded of a quote from a coordinator with whom we worked in a statewide effort to improve reading instruction and achievement. Her advice seems particularly appropriate in this context: "You want to teach children as quickly as you can but as slowly as you must." Although research on intensifying intervention through performance/progression is relatively new, findings suggest the promising results of responsive intervention, that is, intervention that systematically uses student data to adjust instruction and enhance performance.

## Appendix A

We compared the performance of the Early Reading Intervention (ERI) with ongoing adjustments (ERI-A) students in each latent class to that of the corresponding propensity matched ERI students from the ERI condition using multilevel analysis (Hox, 2002) with the following set of models,

**Level 1 (Student-Level) Model**

$$\text{Posttest}_{ij} = \beta_{0j} + \beta_{1j}\text{Pretest}_{ij} + \beta_{2j}\text{PPVT}_{ij} + \beta_{3j}\text{Age}_{ij} + \beta_{4j}\text{Gender}_{ij} + \beta_{5j}\text{Hispanic}_{ij} + \beta_{6j}\text{African\_American}_{ij} + \beta_{7j}\text{Special\_ed}_{ij} + \beta_{8j}\text{Bilingual}_{ij} + \beta_{9j}\text{Last Lesson Number}_{ij} + \beta_{10j}\text{Number of Lessons}_{ij} + \beta_{11j}\text{Propensity Score}_{ij} + r_{ij} \quad \text{(A1)}$$

where i represents the i-th student and j represents the j-th group. $\text{Posttest}_{ijk}$ is the score of one of the posttest measures for the i-th student from the j-th group.

In this student-level model, covariates included the corresponding pretest score ($\text{Pretest}_{ijk}$), PPVT score ($\text{PPVT}_{ijk}$) as an indicator of the student's receptive language ability, plus demographic variables consisting of student age ($\text{Age}_{ijk}$), gender ($\text{Gender}_{ijk}$), ethnicity (represented by two dummy-coded variables: $\text{Hispanic}_{ijk}$ and $\text{African American}_{ijk}$), special education services ($\text{Special\_ed}_{ijk}$), bilingual status ($\text{Bilingual}_{ijk}$), and two dosage variables, the number of lessons students received ($\text{Number of Lessons}_{ijk}$) and the last lesson number ($\text{Last Lesson Number}_{ijk}$). We also included the propensity score generated from the logistic regression model as a covariate. The within-group random error is $r_{ij}$, and the corresponding variance, $V(r_{ij}) = \sigma^2$, captures the within-group variation.

For those poststudy-only measures (i.e., WRMT-R/NU Letter Name and Letter Sound Checklists, *Test of Written Spelling–4*, and oral reading fluency), we used the untimed letter ID pretest score as their pretest score given that untimed letter ID was highly correlated with all these poststudy reading measures.

The intervention group-level models were specified as shown below:

**Level 2 (Group-Level) Models**

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{ERI-A}_j + \gamma_{02}\text{Interventionist's Experience}_j + u_{0j}$$

and

$$\beta_{1j} = \gamma_{10};\ \beta_{2j} = \gamma_{20};\ \beta_{3j} = \gamma_{30};\ \beta_{4j} = \gamma_{40;}\ \beta_{5j} = \gamma_{50;}\ \beta_{6j} = \gamma_{60;}\ \beta_{7j} = \gamma_{70;}\ \beta_{8j} = \gamma_{80;}\ \beta_{9j} = \gamma_{90;}\ \beta_{10j} = \gamma_{100;}\ \beta_{11j} = \gamma_{110} \quad \text{(A2)}$$

In Equation A2, $\text{ERI-A}_j$ is a dummy-coded variable with 1 as ERI-A condition and 0 as ERI condition. The interventionist's years of teaching experience was also controlled at the group level. The between-group random effect is $u_{0j}$, and the corresponding variance, $V(u_{0j}) = \tau_{00}$, captures the between-group variation. Intraclass correlations (ICCs; i.e., between-group variation over the total variations) were calculated. The average ICC of all measures across four classes ranged from .24 to .48, a common range in educational research (Hox, 2002). The nonzero ICCs provided support for the dependency of observations in same clusters and the need to use multilevel models to adequately handle the nonindependency in our data.

## Appendix B

To evaluate the practical significance of the difference on the poststudy reading outcomes between ERI-A and the corresponding ERI condition (i.e., ERI), we used Hedges's (2007) proposed effect size (i.e., $\delta_w$), which is specifically designed for cluster-randomized studies. It is computed using the following equation,

$$\delta_w = \frac{\mu_{\bullet}^{T} - \mu_{\bullet}^{C}}{\sigma_w} = \frac{\gamma_{01}}{\sqrt{\sigma^2}} \quad \text{(B1)}$$

where $\gamma_{01}$ is the mean difference between ERI-A and ERI after adjusting for the demographical and pretest covariates. Within-group variance (i.e., $\sigma^2$) was adopted from the unconditional multilevel model in which no covariate/predictor was included. All 237 students were included for conducting the unconditional model.

## Acknowledgments

We thank Kimberly Williams for her support in manuscript preparation and the school districts, administrators, teachers, interventionists, and children for their support and participation in this research.

## Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Dr. Deborah Simmons is a coauthor of the Early Reading Intervention, the experimental program investigated in this article. Dr. Michael Coyne is a consulting author of the Early Reading Intervention, the experimental program investigated in this article. To address perceived conflicts of interest, independent researchers conducted data analyses and reviewed the method, analyses, results, and discussion sections of this article for objectivity and accuracy. All data analyses were conducted by statisticians who had no financial interest with the Early Reading Intervention. In addition, an external consultant with no financial affiliation with the Early Reading Intervention program independently reviewed the manuscript to ensure that (a) data analyses were appropriate, accurate, and objective, (b) the reported findings and discussion were accurate, and (c) the interpretations were consistent with data analysis.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported in part by the Grant R324E06067, National Center for Special Education Research, from the Institute of Education Sciences, U.S. Department of Education. The content and positions in this article do not necessarily represent those of the funding agency.

## Note

1. ERI-responsive group (0/1) = male + age + LEP + Ethnicity + Special_Ed + BW + SM + RON + LNF + PPVT + LI + WI + WA + site.

## References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, *57*, 289–300.

Bloom, B. S. (1971). Mastery learning. In J. H. Block (Ed.), *Mastery learning: Theory and practice* (pp. 47–63). New York, NY: Holt, Rinehart & Winston.

Bradley, M. C., Daley, T., Levin, M., O'Reilly, F., Parsad, A., Robertson, A., & Werner, A. (2011). *IDEA national assessment implementation study: A final report* (NCEE 2011-4027). Washington, DC: Institute of Educational Sciences. Retrieved from http://ies.ed.gov/ncee/pubs/20114026/pdf/20114027.pdf

Bradley, R., Danielson, L., & Doolittle, J. (2005). Response to intervention. *Journal of Learning Disabilities*, *38*, 485–486. doi:10.1177/00222194050380060201

Coyne, M. D., Kame'enui, E. J., Simmons, D. C., & Harn, B. A. (2004). Beginning reading intervention as inoculation or insulin: First-grade reading performance of strong responders to kindergarten intervention. *Journal of Learning Disabilities*, *37*, 90–104. doi:10.1177/00222194040370020101

Coyne, M. D., Little, M., Rawlinson, D., Simmons, D., Kwok, O., Kim, M., & Civetelli, C. (2013). Replicating the impact of a supplemental beginning reading intervention: The role of instructional context. *Journal of Research on Educational Effectiveness*, *6*, 1–23. doi:10.1080/19345747.2012.706694

Coyne, M. D., Simmons, D. C., Hagan-Burke, S., Simmons, L. E., Kwok, O., Kim, M., & Rawlinson, D. M. (in press). Adjusting beginning reading intervention based on student performance: An experimental evaluation. *Exceptional Children*.

Denton, C. A., & Hocker, J. L. (2006). *Responsive Reading Instruction: Flexible intervention for struggling readers in the early grades*. Longmont, CO: Sopris West.

Denton, C. A., Nimon, K., Mathes, P. G., Swanson, E., Kethley, C., Kurz, T., & Shih, M. (2010). Effectiveness of a supplemental early reading intervention scaled up in multiple schools. *Exceptional Children*, *76*, 394–416.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–Third Edition (PPVT-III)*. Bloomington, MN: Pearson.

Enders, C. K., & Tofighi, D. (2008). The impact of misspecifying class-specific residual variances in growth mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*, 75–95. doi:10.1080/10705510701758281

Engelmann, S. (1996). Theory of mastery and acceleration. In J. Lloyd, E. Kame'enui & D. Chard (Eds.), *Issues in educating students with disabilities* (pp. 177–195). Mahwah, NJ: Lawrence Erlbaum.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities*, *45*, 195–203. doi:10.1177/0022219412442150

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2009). *Assisting students struggling with reading: Response to intervention and multi-tier intervention in the primary grades. A practice guide* (NCEE 2009-4045). Washington, DC: U.S. Department of Education, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Good, R. H., III, & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Guskey, T. R. (2009). Mastery learning. In T. L. Good (Ed.), *21st century education: A reference handbook* (Vol. 1, pp. 194–202). Thousand Oaks, CA: Sage.

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*, 341–370. doi:10.3102/1076998606298043

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.

Individuals with Disabilities Education Improvement Act, Pub.L.No. 108-446, 118 Stat. 2647 (2004).

Kwok, O., Luo, W., & West, S. G. (2010). Using modification indexes to detect turning points in longitudinal data: A Monte

Carlo study. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*, 216–240. doi:10.1080/10705511003659359

Larsen, S. C., Hammill, D. D., & Moats, L. C. (2005). *Test of Written Spelling* (4th ed.). Austin, TX: PRO-ED.

Lembke, E. S., McMaster, K. L., & Stecker, P. M. (2010). The prevention science of reading research within a response-to-intervention model. *Psychology in the Schools*, *47*, 22–35. doi:10.1002/pits.20449

Little, M. E., Rawlinson, D., Simmons, D. C., Kim, M., Kwok, O., Hagan-Burke, S., & Coyne, M. D. (2012). A comparison of responsive interventions on kindergarteners' early reading achievement. *Learning Disabilities Research & Practice*, *27*, 189–202. doi:10.1111/j.1540-5826.2012.00366.x

Makar, B. W. (1995). *Primary phonics*. Cambridge, MA: Educators Publishing Service.

Mellard, D., McKnight, M., & Jordan, J. (2010). RTI tier structures and instructional intensity. *Learning Disabilities Research & Practice*, *25*, 217–225. doi:10.1111/j.1540-5826.2010.00319.x

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107–122. doi:10.1007/BF02294746

Ming, K., & Rosenbaum, P. R. (2001). A note on optimal matching with variable controls using the assignment algorithm. *Journal of Computational and Graphical Statistics*, *10*(3), 455–463. doi:10.1198/106186001317114938

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345–368). Thousand Oaks, CA: Sage.

Muthén, L. K., & Muthén, B. O. (1998–2011). *Mplus user's guide* (6th ed.). Los Angeles, CA: Author.

National Center on Response to Intervention. (2010, March). *Essential components of RTI—A closer look at response to intervention*. Washington, DC: U.S. Department of Education, Office of Special Education Programs, National Center on Response to Intervention. Retrieved from http://www.rti4success.org/resourcetype/essential-components-rti-closer-look-response-intervention

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*, 535–569. doi:10.1080/10705510701575396

Pearson/Scott Foresman. (2004). *Scott Foresman sidewalks: Early reading intervention*. Glenview, IL: Author.

Ram, N., & Grimm, K. J. (2009). Methods and measures: Growth-mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development*, *33*, 565–576. doi:10.1177/0165025409343765

Reis, S. M., McCoach, D. B., Little, C. A., Muller, L. M., & Kaniskan, R. B. (2011). The effects of differentiated instruction and enrichment pedagogy on reading achievement in five elementary schools. *American Educational Research Journal*,*48*, 462–501. doi:10.3102/0002831210382891

Robins, J. M., Rotnitzky, A., & Scharfstein, D. (1999). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. *Statistical Methods in Epidemiology, the Environment and Clinical Trials*, *92*(1), 1–94. doi:10.1007/978-1-4612-1284-3_1

Rosenbaum, P. R., & Rubin, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society Series B (Methodological)*, *45*(2), 212–218. doi:10.1017/CBO9780511810725.017

Simmons, D. C., Coyne, M. D., Hagan-Burke, S., Kwok, O., Simmons, L., Johnson, C., & Crevecoeur, Y. C. (2011). Effects of supplemental reading interventions in authentic contexts: A comparison of kindergarteners' response. *Exceptional Children*, *77*, 207–228.

Simmons, D. C., Kame'enui, E. J., Harn, B., Coyne, M. D., Stoolmiller, M., Santoro, L. E., & Kaufman, N. K. (2007). Attributes of effective and efficient kindergarten reading intervention: An examination of instructional time and design specificity. *Journal of Learning Disabilities*, *40*, 331–347. doi:10.1177/00222194070400040401

Stecker, P. S., Fuchs, D., & Fuchs, L. S. (2008). Progress monitoring as essential practice within response to intervention. *Rural Special Education Quarterly*, *27*(4), 10–17.

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trial with paraeducator implementers. *Journal of Educational Psychology*, *98*, 508–528. doi:10.1037/0022-0663.98.3.508

Vadasy, P. F., Sanders, E. A., Peyton, J. A., & Jenkins, J. R. (2002). Timing and intensity of tutoring: A closer look at the conditions for effective early literacy tutoring. *Learning Disabilities Research & Practice*,*17*, 227–241. doi:10.1111/1540-5826.00048

Valentine, J. C., & Cooper, H. (2003). *Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes*. Washington, DC: What Works Clearinghouse.

Vaughn, S., Wexler, J., Roberts, G., Barth, A., Cirino, P., Romain, M., & Denton, C. (2011). Effects of individualized and standardized interventions on middle school students with reading disabilities. *Exceptional Children*, *77*, 391–407.

Vellutino, F. R., Scanlon, D. M., Small, S., & Fanuele, D. P. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities: Evidence for the role of kindergarten and first-grade interventions. *Journal of Learning Disabilities*, *39*, 157–169. doi:10.1177/00222194060390020401

Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1998). The case for early intervention in diagnosing specific reading disability. *Journal of School Psychology*, *36*, 367–397. doi:10.1016/S0022-4405(98)00019-3

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing*. Austin, TX: PRO-ED.

What Works Clearinghouse. (2008). *Procedures and standards handbook, version 2.1*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf

Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests–Revised/Normative Update*. Bloomington, MN: Pearson. (Original work published 1987)