# THE USE OF PIECEWISE GROWTH MODELS TO ESTIMATE LEARNING TRAJECTORIES AND RTI INSTRUCTIONAL EFFECTS IN A COMPARATIVE INTERRUPTED TIME-SERIES DESIGN

Keith Zvoch
UNIVERSITY OF OREGON

**ABSTRACT**
Piecewise growth models (PGMs) were used to estimate and model changes in the preliteracy skill development of kindergartners in a moderately sized school district in the Pacific Northwest. PGMs were applied to interrupted time-series (ITS) data that arose within the context of a response-to-intervention (RtI) instructional framework. During the kindergarten year, multiple literacy assessments were conducted and supplemental instruction was initiated with struggling readers to promote the attainment of literacy benchmark goals. The use of PGMs provided analytic flexibility to specify a discontinuous learning function and test a number of within- and between-group model parameters related to the evaluation of program efficacy. Results revealed a statistically significant linear increment in student learning following the onset of the intervention. However, absolute and relative gains in literacy performance were more modest during the second half of the academic year. Key issues in applying PGMs within the RtI framework, including the coding of time functions and the interpretation of model results, are addressed.

THE ELEMENTARY SCHOOL JOURNAL
Volume 116, Number 4. Published online May 11, 2016
 0013-5984/2016/11604-0008$10.00

PASSAGE of the Individuals with Disabilities Education Improvement Act (IDEA, 2004) has resulted in the widespread implementation of "response-to-intervention" (RtI) instructional models (Bradley et al., 2011; GlobalScholar/Spectrum K12 Solutions, 2011) that necessitate the collection and use of multiple waves of assessment data on individual students during each academic year. The frequent collection of formative assessment data is designed to facilitate the early screening of student academic performance, enable continuous monitoring of student learning progress, and allow targeted intervention in cases where learning difficulties are observed (Fletcher & Vaughn, 2009; Fuchs & Fuchs, 2006; Fuchs, Vaughn, & Fuchs, 2008). The advent of an integrated system of assessment and instruction offers school personnel a data-driven means for identifying and proactively intervening with students at risk of poor academic outcomes. However, many empirical questions regarding the nature of underlying growth functions, the "effects" of treatment, and variability in student outcomes are often left unaddressed in current applications. In the following, an analytic approach that closely aligns with the dynamic and fluid nature of contemporary RtI-based instructional practice is presented. The model provides the analytic flexibility requisite to estimate student learning trajectories over the course of one or more academic years and statistically identify the short- and longer-term effects associated with application of supplemental instructional regimes that are designed to meet the needs of struggling learners.

## Response to Intervention (RtI)

The use of early screening and progress-monitoring data to inform instructional decisions regarding at-risk students is generally viewed as more effective and equitable than the "wait to fail" IQ/achievement discrepancy model that has historically been utilized to identify and serve students with learning disabilities (Fuchs & Fuchs, 2006; Fuchs et al., 2008). Yet, the nonrandom application of tiered intervention regimes that are specifically designed to vary in response to student assessment outcomes also makes identification of RtI program effects difficult. With some notable exceptions (e.g., more tiers or problem-solving approaches; see Burns, Appleton, & Stehouwer, 2005; Fuchs, Mock, Morgan, & Young, 2003; Mellard, Stern, & Woods, 2011), RtI typically involves tri-annual universal screening and high-quality, evidenced-based, direct core instruction delivered to all students in regular education classrooms (Tier I). For students struggling with knowledge and skill acquisition, more frequent assessment is conducted and intensive Tier II supplemental small-group instruction is implemented. If a student "responds" to the instructional supplement, Tier II services may be continued for a period of time or a return to normative Tier I instruction with continued monitoring may be prescribed. However, if a student remains nonresponsive, alternative, more intensive Tier III instructional supplements (e.g., individual intervention) can be enacted. As a result, individual students may be systematically entered and exited from a program of supplemental instruction at different points in time, have their instructional dosage incrementally strengthened by lengthening the duration or increasing the intensity of the supplement (e.g., individual tutoring vs.

small-group instruction), and/or be transferred to one or more alternative instructional programs over the course of an academic year. The assessment schedules of individual students can be distinct as well.

## Methodological Challenges and Opportunities

From a methodological perspective, the design and analytic complexities that arise in conjunction with implementation of an RtI model pose a significant challenge to identifying whether and to what extent a particular instructional intervention is effective for one or more struggling learners. Of particular concern is the pedagogically appropriate and administratively cost-effective strategy of delivering targeted supplemental instruction only to those students most at risk of poor academic outcomes. The common use of nonrandom, need-based assignment practices results in either no comparison group or nonequivalent group designs that hinder the separation of instructional effects from the background and motivational characteristics of students. In addition, to promote widespread implementation, school personnel are trained in the use of rules for visual inspection of data that involve examination of trend lines applied to performance benchmarks derived from local or national norms (e.g., Busch & Reschly, 2007; Deno et al., 2009). The graphical representation of student assessment scores is designed to enable teachers and instructional coordinators to quickly determine whether a student is "on track" to becoming skill or content-area proficient or whether a more intensive or alternative instructional treatment should be considered. The simplified set of decision rules offers a straightforward, practitioner-friendly means for monitoring and evaluating individual learning progress, but the visual inspection of data does not allow for identification of chance fluctuations in performance or statistical evaluation of key aspects of the intervention context. As a result, current evaluative practice does not fully utilize the wealth of available data to systematically investigate the short- and longer-term pattern of student performance, the responsiveness of students to similar or distinct intervention regimes, and whether variations in student background or instructional practice associate with learning outcomes.

In many respects, the underutilization of rich sets of treatment-related, time-series data is not unique to the educational setting (see Duncan & Duncan, 2004; Moore, Osgood, Larzelere, & Chamberlin, 1994). School-based RtI instructional regimes are similar to the individualized treatment approaches that are common in the health and social service fields whereby an individual presents with symptoms, is screened and diagnosed, receives a treatment, and progress is monitored over time. If the individual is "nonresponsive," the treatment is modified and continued tracking occurs. Over time, the nature, intensity, and sequence of the implementation of additional treatment regimes varies in response to recipient outcomes and is designed to closely align with client needs. Yet, as with RtI-based instruction, the manner and extent to which data are used to inform decisions regarding the continuation or modification of a physical or mental health treatment is often limited to simple descriptive comparisons with a "performance" benchmark. However, when different drugs or dosages are prescribed, physical and mental health therapies are changed, and clinic reward structures are modified in an attempt to optimally align program services with individual needs, the opportunity for systematic evaluation of treatment effects and individual response patterns is presented. More specifically, under conditions where the type and timing of treatment or programmatic changes are recorded and key outcomes are contiguously measured, researchers and practitioners can draw on the power of the associated interrupted time series (ITS) to increase the validity of inferences regarding the average treatment effect of an intervention as well as ascertain whether responsiveness to a course of supplemental treatment varies as a function of demographic, instructional, and/or institutional characteristics.

## Interrupted Time-Series Designs and Piecewise Growth Models

When an RtI model is applied, an ITS design develops by default. The default implementation of a strong quasi-experimental design is a coincidental but fortuitous opportunity for researchers. In applied treatment contexts, ITS designs are often recommended as an alternative to the randomized experiment to provide strong control over threats to internal validity (Bloom, 2003; Campbell, 1969; Shadish, Cook, & Campbell, 2002). A time series is realized when an outcome of interest (e.g., reading performance) is repeatedly measured over an extended period. An interruption to the time series occurs when an intervention (e.g., supplemental Tier II instruction) is enacted or ended at a known point. Treatment effects are demonstrated when the level or slope of the intervention time series statistically deviates from the preintervention time series. ITS designs are particularly useful in applied field settings as they are minimally intrusive and are consistent with the need-based provisioning of limited resources often found in education and social and health service contexts (Gordon & Heinrich, 2004; Osgood & Smith, 1995). ITS-based designs may also yield inferences with strength comparable to those associated with randomized experiments as knowledge of the point in time in which the intervention was initiated enables rigorous comparison of adjacent developmental trends. Causal inference is particularly robust as the number of observations increases, as comparison or control groups are added, and as an abrupt change in the outcome can be demonstrated at the point at which the intervention was initiated or ended (Shadish et al., 2002; St. Clair, Cook, & Hallberg, 2014).

In ITS design contexts, conventional growth-modeling techniques that yield a single trajectory for the growth function over all time points are not appropriate (Osgood & Smith, 1995; Singer & Willett, 2003). Instead, an alternative model specification that accounts for the discontinuous effects associated with an episodically administered treatment intervention schedule is needed. A recent solution to capturing the richness and complexity of the RtI decision-making framework is offered by the procedural and analytic flexibility of piecewise growth models. The piecewise growth model (PGM) is useful in the ITS design context as individual changes in performance level and growth coordinated with the exact time at which the intervention is delivered can be estimated. By breaking the time function into pieces that correspond to each period of interest, separate parameters that explicitly represent distinct segments of the overall growth trajectory are specified and tests for discontinuities between model parameters are conducted (Hindman, Cromley, Skibbe, & Miller, 2011; Singer & Willett, 2003). Of particular importance for the current investigation is the flexibility that the PGM affords for estimating and modeling changes in the learning trajectories that occur in conjunction with the initiation and removal of supplemental instruction. In addition, unlike traditional time-series models, observed variability in the short- and longer-term responses of recipients offers the opportunity to investigate whether characteristics of the individual or treatment context are associated with period-specific changes in outcome level and growth (Gordon & Heinrich, 2004; Hindman et al., 2011; Singer & Willett, 2003).

## Study Purpose

As the collection and use of multiple waves of assessment data to screen and proactively intervene with students at risk of poor academic outcomes has become increasingly widespread, the opportunity for more sophisticated evaluation of intervention effects and student learning progress has emerged. The purpose of this article is to present a statistical approach that more coherently aligns with the organizational realities of the RtI-based instructional regimes now commonly delivered in elementary and middle school contexts. The PGM presented herein is designed to specifically map the structure of contemporary educational practice by tracking the development of a preliteracy skill over the course of one academic year. The demonstration draws on a unique data set that contains a time series of upward of 16 test scores that are aligned with key dates in the academic calendar, including the date at which a supplemental literacy intervention was initiated, withdrawn, and subsequently reinitiated at points contiguous with the onset and end of the winter break. The availability of a second time series on a sample of students who did not receive supplemental Tier II instruction extended the demonstration to a multiple-group ITS design context. The following research questions were investigated: (*a*) What was the shape of the growth function for students assigned to supplemental Tier II instruction? (*b*) On average, did students' literacy performance and rate of literacy skill growth increase relative to baseline after supplemental reading support was initiated? and (*c*) Were there distinct differences between the growth functions and performance outcomes of treatment and "control" students?

## Method

### Data Source

Data records obtained from a moderately-sized school district in the Pacific Northwest served as the basis for the current investigation. The sample comprised 356 students from three schools who began kindergarten either in 2008–2009 ($n$ = 114, 32%), 2009–2010 ($n$ = 135, 38%), or 2010–2011 ($n$ = 107, 30%) and whose preliteracy skill development was assessed multiple times throughout the academic year.[1] During the study period, two of the study schools did not meet annual yearly progress (AYP) requirements (No Child Left Behind Act, 2002) in one or more years due to the performance of students in two disaggregated subgroups—students with disabilities and economically disadvantaged. These schools tended to have greater percentages of ethnic minority students and students eligible for a lunch subsidy than their peer school that achieved AYP benchmarks each year. Overall, 33% ($n = 117$) of the sample were identified as early struggling readers and received RtI-based supplemental literacy instruction. Eighty-one percent ($n = 288$) of students were identified as non-Latino White. In addition, 56% ($n = 200$) of students received a free or reduced-price lunch, and 5% ($n = 19$) were identified as English learners. Girls constituted 49% ($n = 176$) of the sample.

## RtI Model

Each academic year the school district employs a two-tier RtI model to identify and proactively intervene with kindergarten students struggling to acquire basic early literacy skills.[2] Tier I involves the systematic delivery of high-quality, evidenced-based, direct literacy instruction (National Reading Panel, 2000) and tri-annual universal screening of student literacy performance. Each instructional day, all students receive 1 hour of differentiated literacy instruction in alphabetic understanding (letter names and sounds, decoding, phonic analysis) and phonemic awareness (oral blending and segmentation). Daily literacy instruction is provided in a 30-minute whole-group and in a 30-minute small-group setting. Lessons begin in a whole-group setting, where literacy skills are modeled and practiced. Students then cycle through small-group and individual reading activities delivered at mixed-use literacy stations. The literacy stations allow for individual reading activities and collaborative group learning and also provide opportunity for need-based differentiated instruction. Lessons close with a whole-group review period. In each whole- and small-group arrangement, teachers provide direct modeling of literacy skills and offer multiple opportunities for practice with corrective feedback.

During the first week of the academic year, all students are administered select subtests from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002). For students who show initial performance difficulties, follow-up screeners are then administered on a biweekly basis. After 6 weeks of rigorous instruction and three formative literacy assessments, students who score in the lowest quartile of district performance on the DIBELS Letter Naming Fluency (LNF) subtest (~10 or fewer letter names) are considered to be at heightened risk for future reading difficulty, a benchmark consistent with a number of federal and state guidelines (Piasta, Petscher, & Justice, 2012). These students and those whose teacher-based referral is approved by the district literacy coordinator are enrolled in a Tier II instructional supplement during the seventh week of the academic year. Tier II instruction reinforces and complements regular classroom literacy instruction through the provision of 30 additional minutes of daily small-group literacy instruction (see below). After the supplement is begun, students' responsiveness to the intervention is monitored through administration and evaluation of scores obtained from biweekly LNF assessments.

## Supplemental Literacy Intervention: Extended-Day Kindergarten

Extended-day kindergarten (EK) was used as the mechanism for delivering Tier II instruction to students struggling to acquire basic literacy skills. Students received the 30-minute instructional supplement directly after the close of classes each day, resulting in an additional 150 minutes of instruction in a typical 5-day school week. EK was designed to help struggling learners gain the foundational skills requisite to becoming successful readers and otherwise shrink the preliteracy skills gap with their peers. EK was delivered entirely in a small-group setting (~5:1 student-teacher ratio). Students were grouped based on skill level to allow teacher-directed instruction to be maximally supportive and aligned with individual student needs. As with Tier I small-group instruction, teachers modeled critical preliteracy skill components (e.g., the production of letter names and sounds, the segmentation and blending of multisyllable words) and provided each child with frequent practice opportunities and individual feedback to correct errors and otherwise positively reinforce student responses.

## Sample Characteristics

Table 1 presents the characteristics of the analytic sample classified by treatment status (i.e., participant/nonparticipant in EK, the Tier II instructional supplement). Examination of Table 1 reveals relatively higher percentages of free-lunch recipients, English language learners, and ethnic minority students in the treatment group. EK students were also younger, had more LNF assessment occasions throughout the academic year, and by virtue of the district's need-based assignment practice had lower initial LNF assessment scores in comparison to students in the control condition. The exit LNF status of EK students was also markedly lower than their peers.[3] A series of chi-square and independent *t*-tests revealed

Table 1. Student Characteristics by Treatment Status

| | Treatment ($n$ = 117) | | Control ($n$ = 239) | |
|---|---|---|---|---|
| Student Characteristics | *M* | *SD* | *M* | *SD* |
| Female | .43 | – | .53 | – |
| Ethnic minority | .23 | – | .17 | – |
| English language learner | .09 | – | .04 | – |
| Free lunch recipient | .61 | – | .54 | – |
| Age in months at entry[a] | 64.58 | 3.48 | 65.44 | 3.52 |
| Initial LNF status[a] | 2.75 | 4.99 | 16.41 | 12.51 |
| Exit LNF status[a] | 29.92 | 14.64 | 44.81 | 15.70 |
| No. of observations[a] | 11.21 | 3.24 | 8.67 | 3.28 |

Note.—The mean of each dichotomous variable represents the proportion of the sample with the identified characteristic (e.g., 61% of the treatment group were free lunch recipients).

[a] Group mean difference statistically significant ($p < .05$).

that with the exception of the age difference, observation number, and the initial and exit performance contrast, none of the other demographic differences were statistically significant.

## Placement and Outcome Measure

The number of letters named in a one-minute assessment frame was used to determine a student's eligibility for EK and to measure short- and longer-term changes in early literacy skill development. Letter naming was used as the focal outcome for conceptual and practical reasons. Conceptually, letter naming has been established as a critical early predictor of future reading success (Ritchey & Speece, 2006; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004) and serves as an important risk factor in the development of reading disabilities (Catts, Fey, Zhang, & Tomblin, 2001; Puolakanaho et al., 2007). The ability to identify individual letters is a fundamental component of alphabetic knowledge (i.e., that printed letters symbolize and represent the sounds in spoken words) and letter recognition is the first step in the association of letter sounds and the beginning of phonemic awareness (Ehri, 1998). Although letter naming has repeatedly been shown to predict later reading ability, there is no current consensus on the nature of the causal relationship between early literacy skill components or on the manner and extent to which letter names should be explicitly taught (National Early Literacy Panel, 2008; Piasta & Wagner, 2010). Nonetheless, in their recent meta-analytic review, Piasta and Wagner (2010) demonstrate that evidence-based, explicit instruction (National Reading Panel, 2000) was associated with a range of alphabetic knowledge outcomes. Effect-size estimates were larger for letter knowledge than for letter-fluency outcomes, and varied across a range of instructional and contextual conditions.

Letter naming was assessed on the basis of scores obtained from administration of the DIBELS LNF subtest. LNF is a standardized, individually administered measure of lowercase- and capital-letter-naming fluency. Since the test is timed, fluency refers to students' letter-naming speed. The LNF subtest is typically administered throughout kindergarten and at the beginning of first grade to permit early identification of students at risk of developing future reading difficulties. In the study schools, letter-naming fluency is the only preliteracy skill that is assessed across the kindergarten year, offering the practical advantage of a common assessment on which to track and evaluate student performance. At each administration, students are presented with a page of 110 upper- and lowercase letters randomly arranged in 11 rows with 10 letters on each row. Students receive one point for each correct letter name provided during a one-minute assessment frame, with the total number of points serving as the LNF score. Scores can range from a lower bound of 0 to an upper bound of 110. In the current sample, the maximum score on the first assessment in September was 62, with 98% of initial scores <40. The maximum across all assessment occasions was 90, obtained by a student on one of the final assessments in May.

In investigations of the psychometric properties of the DIBELS LNF, an alternative-form reliability estimate of .93 (Good, Kaminski, Simmons, & Kame'enui, 2001) and a 2-week test-retest reliability estimate of .90 have been reported (Elliott, Lee, & Tollefson, 2001). The test developer also reported a concurrent relationship of .70 between the LNF and the Woodcock-Johnson Psycho-Educational Battery readiness score (Good et al., 2001). Other investigators have reported concurrent correlations of .62 with scores on the Developmental Reading Assessment (DRA), .52 with scores on the Test of Early Reading Ability (TERA)–reading, and .59 with TERA–alphabet (Rouse & Fantuzzo, 2006). A similar range of correlations (.52–.59) have been reported between LNF scores and subtest scores on the Comprehensive Test of Phonological Processing (Hintze, Ryan, & Stoner, 2003). Larger correlations between LNF scores and scores on the Woodcock-Johnson Revised (WJ-R) Skills cluster (.75) and WJ-R letter-word identification (.71) have also been obtained (Elliott et al., 2001). The concurrent validity evidence suggests that the DIBELS LNF is a related but somewhat conceptually distinct measure of early literacy knowledge and readiness. In addition, the diagnostic value that LNF offers researchers and practitioners has been further established in a study that examined the measure's predictive validity. Predictive relationships between kindergarten LNF scores and first-grade DRA Instructional Reading, TerraNova vocabulary, TerraNova language, and TerraNova reading scores have been reported as .67, .63, .57, and .48, respectively (Rouse & Fantuzzo, 2006).

## Analytic Procedures

Two-level PGMs were applied to the data obtained over the course of the study. PGMs were utilized in order to systematically represent the ITS design that arose in conjunction with the RtI-based delivery of EK. The adoption of a piecewise growth-modeling approach allowed estimation of the preliteracy skill performance and growth of students during periods of time adjacent to and contiguous with implementation of EK and with respect to a naturally occurring break in the academic calendar. The splicing of a growth trajectory into discrete parts is particularly useful when a nonlinear function is hypothesized or observed and representation and comparison of definable segments is sought (Hindman et al., 2011; McCoach, O'Connell, Reis, & Levitt, 2006; Singer & Willett, 2003). Estimating the performance level and growth of struggling readers during specific time frames first enabled a within-group examination of the rate of literacy skill acquisition before, during, and after a course of supplemental instruction was stopped and then reinitiated. A second PGM that included time-series data on a group of students not assigned to treatment was then estimated to facilitate between-group contrasts of key growth-trajectory components. To accommodate the structure of the RtI instructional model, assessment schedules were matched with treatment start and end dates. Assessment and treatment-related events were represented in an elapsed calendar week metric.

An abbreviated version of the coding scheme used to estimate the growth portion of the piecewise model is presented in Table 2. The coding associated with the PGM was relatively complex due to the need to specify multiple segments to represent the ITS design. Additional guidance and detailed examples for coding discontinuous growth models similar to the present application can be found in several sources (e.g., Fitzmaurice, Laird, & Ware, 2011; Raudenbush & Bryk, 2002; Singer & Willett, 2003). In the table, the 40-week academic year is coded in multi-

Table 2. Incremental Coding Scheme for Academic Year Changes in Letter-Naming Fluency

| | Weeks of Instruction | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $X_{1i}$ | 0 | 3 | 6 | 9 | 12 | 15 | 18 | 21 | 24 | 27 | 30 | 33 | 36 | 39 | Baseline (pretreatment slope) |
| $X_{2i}$ | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | Tx-level |
| $X_{3i}$ | 0 | 0 | 0 | 3 | 6 | 9 | 12 | 15 | 18 | 21 | 24 | 27 | 30 | 33 | Tx-slope $\Delta$ |
| $X_{4i}$ | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | Winter break |
| $X_{5i}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 6 | 9 | 12 | 15 | 18 | 21 | Postbreak slope $\Delta$ |
| $X_{6i}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 36 | 81 | 144 | 225 | 324 | 441 | Postbreak slope$^2$ |

Note.—The intervention was initiated in week 7 and winter break covered the period of time between weeks 16 and 18. For brevity, time codes are presented in multiples of 3. In the estimated models, the exact number of weeks for each time point was specified. Postbreak slope$^2$ refers to the squared linear slope values that specify the post-winter-break quadratic term.

ples of three. The first row shows the time codes used to estimate a linear growth trajectory over the measurement occasions representing the baseline (i.e., pre-intervention period). The second and fourth rows show the codes used to estimate the change in the level of literacy performance at the point at which treatment was initiated (i.e., week 7) and over the course of the winter break (i.e., weeks 16–18). In the third and fifth rows, codes representing the linear change in slope from baseline (row 3) and from the initial intervention period (row 5) are presented. The codes in row 3 facilitate the capture of the relative change in slope (from baseline) during the initial intervention period, while the codes in row 5 capture the relative change in slope (from the prewinter-break intervention period) during the period of time following the winter break. Weekly unit increments in the squared linear slope values (row 6: Postbreak slope$^2$) were also used to specify a quadratic term to test for the potential of a curvilinear relationship between time and literacy performance during the post-winter-break treatment period.[4] The applied coding scheme defines the status parameter as expected LNF performance upon entry to kindergarten.

Two-level PGMs (observations within students) were estimated using the hierarchical linear modeling (HLM) program, version 7.0 (Raudenbush et al., 2011), but it should be noted that PGMs can be fit in almost any multilevel or structural equation modeling software package. The level 1 model used to estimate the parameters of the growth function is specified in the equation below. In the equation, the trajectory of change is conceived as a function of three linear growth terms, two terms that represent level changes, a term for the quadratic growth component, and an initial status parameter ($\pi_{0i}$). The parameter $\pi_{1i}$ represents the linear rate of LNF growth prior to the onset of the EK treatment, $\pi_{2i}$ represents the change in LNF performance level at the onset of EK, $\pi_{3i}$ represents the change in the linear rate of LNF growth during the initial intervention period, $\pi_{4i}$ represents the change in LNF associated with the winter break, $\pi_{5i}$ represents the change in the linear rate of LNF growth during the post-winter-break intervention period, and $\pi_{6i}$ represents the quadratic growth component associated with the post-winter-break intervention period. All level 1 growth-function parameters were allowed to vary across students. Level 1 (measurement occasions): $Y_{ti} = \pi_{0i} + \pi_{1i}$ (pretreatment slope) $+ \pi_{2i}$ (treatment level) $+ \pi_{3i}$ (treatment slope $\Delta$) $+ \pi_{4i}$ (winter break) $+ \pi_{5i}$ (post-winter-break slope $\Delta$) $+ \pi_{6i}$ (post-winter-break slope$^2$) $+ e_{ti}$. Level 2 (students): $\pi_{0i} = \beta_{p0} + r_{0i}$, and $\pi_{pi} = \beta_{p1} + r_{1i}$, for each slope parameter.

In the two-group model, a dummy-coded vector representing treatment status (i.e., 0 = treatment, 1 = control) was added as a student-level predictor. The treatment group (i.e., EK students) was used as the referent in order to align estimates with the one-group model. The inclusion of the treatment-status variable enabled estimation and comparison of the direction and size of group differences in each growth-trajectory component. Variability in each of the level 1 growth function parameters was also conditionally modeled by student demographic variables, including age at entry to kindergarten. Dummy codes were used to identify girls and ethnic minority students as well as free-lunch recipients and English language learners. Finally, it should be noted that while the incremental coding scheme outlined above provides estimates and statistical tests of the change in level and slope from one period to another, the PGM was also reparameterized to obtain actual slope values and alternative tests of interest. To test whether slopes were statistically different from zero, a differential-rate coding scheme was applied (e.g., Gordon & Heinrich, 2004; Raudenbush & Bryk, 2002) and models were reestimated. The alternative parameter estimates and statistical tests are presented and discussed in the Results section.

## Results

### One-Group ITS Model

Table 3 presents the analytic results associated with the one- and two-group ITS models. In the one-group model, growth-parameter estimates revealed that upon entry to kindergarten, students who would later be assigned to EK identified approximately three letter names on average during the initial one-minute assessment period ($\pi_{0i} = 2.73$), and in the 6 weeks prior to the start of the intervention produced an additional two-thirds of a letter name per week ($\pi_{1i} = 0.66$). Imme-

Table 3. One- and Two-Group ITS Model Results

| | One-Group ITS | | Two-Group ITS | |
|---|---|---|---|---|
| | Coefficient (*SE*) | Variance | Coefficient (*SE*) | Variance |
| Intercept, $\pi_{0i}$ | 2.73 (.44)* | 11.94 | 2.69 (.44)* | 88.06 |
| Control, $\beta_{01}$ | | | 13.86 (.95)* | |
| Baseline, $\pi_{1i}$ | .66 (.10)* | .34 | .66 (.10)* | .31 |
| Control, $\beta_{11}$ | | | 1.01 (.19)* | |
| Tx-level, $\pi_{2i}$ | .63 (.49) | 2.00 | .68 (.49) | 1.49 |
| Control, $\beta_{21}$ | | | −1.15 (.97) | |
| Tx-slope Δ, $\pi_{3i}$ | .73 (.15)* | .79 | .69 (.16)* | 1.33 |
| Control, $\beta_{31}$ | | | −1.28 (.28)* | |
| Winter break, $\pi_{4i}$ | −5.38 (.99)* | 11.77 | −5.09 (.98)* | 14.77 |
| Control, $\beta_{41}$ | | | 1.03 (1.43) | |
| Postbreak slope Δ, $\pi_{5i}$ | .45 (.20)* | 1.80 | .42 (.20)* | 2.10 |
| Control, $\beta_{51}$ | | | −.20 (.27) | |
| Postbreak slope$^2$, $\pi_{6i}$ | −.06 (.01)* | .01 | −.06 (.01)* | .01 |
| Control, $\beta_{61}$ | | | .02 (.01) | |

Note.—The treatment group (EK participants) served as the referent in the two-group model; *SE* = standard error.
* $p < .05$.

diately after the onset of the intervention during week 7, LNF performance remained statistically constant (i.e., no change in level; $\pi_{2i} = 0.63$, $p > .05$), but the linear rate of growth doubled relative to the pretreatment period ($\pi_{3i} = 0.73$; $0.66 + 0.73 = 1.39$). In other words, the rate of LNF growth was twice as fast during the 9-week period that bounded the start and end of the intervention (at the winter break) than during the 6-week pretreatment period. However, over the course of the 3-week winter break, letter-naming performance declined by over one letter name ($\pi_{4i} = -5.38/3 = -1.79$; $1.39 - 1.79 = -0.40$; $-0.40 * 3 = -1.20$). After the winter break, the linear growth increment was again positive ($\pi_{5i} = 0.45$) and statistically different from the growth rate observed during the initial intervention period.

In the post-winter-break period, a decelerating positive growth function was observed, instantaneous rate of growth ($\pi_{5i} = 0.45$; $0.66 + 0.73 + 0.45 = 1.84$), quadratic change ($\pi_{6i} = -0.06$). The negative relationship between the linear and quadratic growth components (–.79) indicated that students with a faster instantaneous linear growth rate had a steeper rate of deceleration during the 22-week instructional period following the winter break. Computation of the transition point revealed that at the end of April, approximately 15 weeks after instruction resumed, the quadratic trajectory reversed direction (i.e., $-1.84/(2 * -.06) = 15.3$) and LNF performance began to decline.

To directly estimate the linear slope value for each intervention period and to test whether the slopes were statistically different than zero, the one-group ITS model was reparameterized using differential rate codes. The linear rate of LNF growth was statistically greater than zero directly after the onset of EK ($\pi_{3i} = 1.39$, $p < .05$) as well as during the post-winter-break period ($\pi_{5i} = 1.84$, $p < .05$). The deviance statistics associated with the increment and differential-rate models revealed that a better fit to the data was obtained when each growth-trajectory component was specified to vary randomly (deviance = 8706.76, parameters estimated = 29) than when growth components were constrained to be equivalent (deviance = 9393.86, parameters estimated = 2). Multiparameter variance component tests indicated that the improvement in model fit was statistically significant, $\chi^2 \Delta (27) = 687.10$, $p < .05$. To illustrate the variability in LNF outcomes, Figure 1 presents a sample of estimated growth trajectories.

## Two-Group ITS Model

Estimation of the two-group model revealed that students who were not assigned to EK (i.e., the comparison group) entered kindergarten with the ability to name close to 14 more letters than their peers ($\beta_{01} = 13.86$, $p < .05$). Comparison-group students also produced an average of one additional letter name per week more than their peers during the preintervention period ($\beta_{11} = 1.01$, $p < .05$). Immediately after the onset of the intervention period, comparison students' relative LNF performance remained statistically constant (i.e., no level change) while their rate of LNF growth slowed with respect to the baseline period. The slope decrement was statistically different from the positive growth increment experienced by treatment students. The relative change in LNF trajectories during the

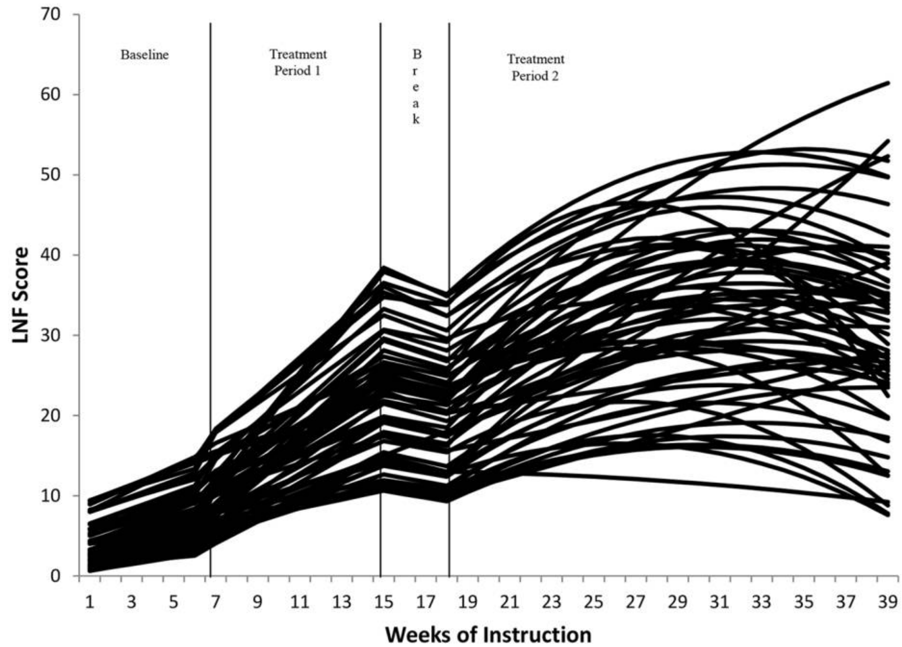


Figure 1. Letter-naming fluency as a function of instructional schedule and time.

initial intervention period (i.e., weeks 7 to 15) enabled treatment students' rate of letter-naming fluency (i.e., 0.66 + 0.69 = 1.35) to exceed that associated with the group of nonparticipants (i.e., 0.66 + 1.01 + 0.69 – 1.28 = 1.08). The deflection in growth rates prevented the initial and growing LNF performance gap from further widening and facilitated a reduction in some of the additional disparity that accumulated during the preintervention period (i.e., a relative gain of approximately 2.5 letter names for the treatment group, 1.35 – 1.08 = 0.27; 0.27 * 9 = 2.43).

Estimated between-group performance differences over the winter break and during the post-winter-break instructional period indicated that the group of EK students had a relatively greater LNF performance loss during the break, stronger postbreak instantaneous growth, and a more severe deceleration in LNF performance during the post-winter-break period. However, none of the observed group differences were statistically significant. Figure 2 presents the group-specific learning patterns across the study period. In the first panel of Figure 2, the initial and growing preintervention literacy gap between EK students and their peers can clearly be seen. Also of note is the positive deflection in LNF performance for EK participants that emerged in conjunction with the start of the intervention (panel 2) as well as the drop in performance for both student groups during the winter break (panel 3). Of additional relevance are the nonlinear functions that represent LNF performance during the latter part of the academic year (panel 4). Despite the persistent LNF performance level difference, comparison-group students (0.66 + 1.01 + 0.69 – 1.28 + 0.42 – 0.20 = 1.30; –1.30/(2 * –.04) = 16.25) transitioned to a negative trajectory at a nearly identical point (early May) as EK students during the

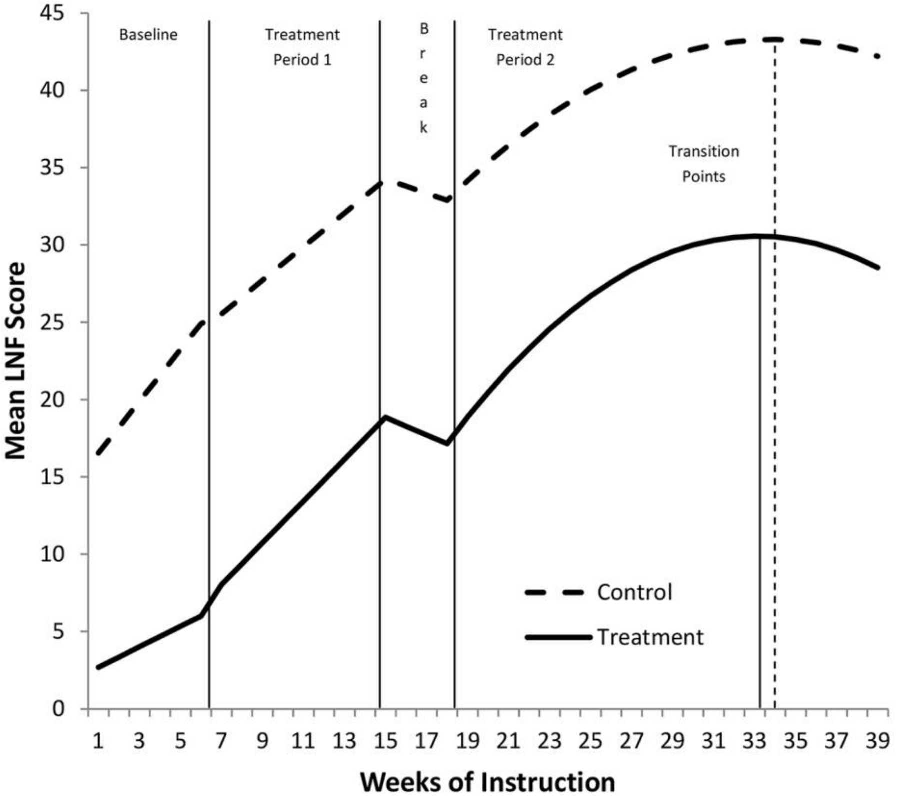


Figure 2. Mean letter-naming fluency as a function of instructional schedule, treatment status, and time.

post-winter-break period. Finally, it should be noted that with the exception of a positive relationship between age of entry and initial LNF status (i.e., initial status increased .42 letter names for each additional month in age at kindergarten entry) and the larger drop in LNF performance over the winter break for boys (i.e., an additional one-half of a letter name decrement for boys), demographic characteristics were not associated with any of the other LNF outcomes.[5]

## Discussion

Interest in preventing initial reading difficulties from progressing into long-term reading failure (Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Snow, Burns, & Griffin, 1998; Torgesen, 2002) has promoted the development and widespread implementation of the RtI instructional model (e.g., Bradley et al., 2011). Yet, questions regarding how to best identify students in need of supplemental instruction, structure and deliver instructional support, and determine the adequacy of response to strategic intervention continue to be investigated and debated (Compton, Fuchs, Fuchs, & Bryant, 2006; Compton et al., 2010; Denton, 2012; Fuchs & Vaughn, 2012; Gersten et al., 2009). The purpose of the current study was to provide an additional dimension to the discussion by highlighting some of the methodological opportunities and challenges that follow from adoption and implementation of RtI-based treatment models. By recognizing the richness of typical RtI data-collection practice, key issues surrounding the shape of measure-specific growth functions, the efficacy of particular instructional interventions, and variation in individual students' response to supplemental treatment can be directly addressed. When combined with the descriptive approaches currently used by practitioners to track and evaluate student performance, the strategic use of statistical techniques appropriate for modeling data obtained from complex longitudinal research designs may aid in facilitating a more robust implementation of the RtI instructional framework.

## Key Results and Inferences

In the current study, PGM were applied to the time-series data obtained in conjunction with implementation of a two-tier RtI model during the kindergarten year. Results revealed a relatively complex, nonlinear growth pattern with changes in performance observed contiguous with key events and time periods. With the exception of the end-of-year leveling and modest reversal in performance, letter-naming fluency increased during periods of instruction but declined when students were out of school during the winter break. The performance changes that were observed contiguous with increments and decrements in instructional dosage provide some evidential support regarding the benefit of schooling and the efficacy of EK, the Tier II instructional supplement. However, it should be noted that without a comparison or control group to reference, the growth increment associated with the onset of supplemental instruction in the initial one-group "A-B" ITS design could also be attributed to the background or maturational characteristics of the student sample rather than the positive impact of EK. The potential value of strategically targeting supplemental instruction to those most in need was further bolstered by results associated with the two-group ITS model. The two-group model facilitated a comparison of students who received Tier I instruction with those who participated in Tier I and the daily 30-minute Tier II EK supplement starting in week 7 of the academic year. Although the student groups were nonequivalent by design, the between-group comparison served to strengthen the inference derived from the one-group ITS model by making a maturational or seasonality explanation for the LNF growth increment somewhat less plausible (Shadish et al., 2002; St. Clair et al., 2014).

In addition to strengthening the inference regarding the impact of Tier II instruction, the two-group model also served to contextualize the relative magnitude of the increment to learning experienced by EK students. Notably, while the rate of letter-naming fluency doubled from baseline during the initial intervention period (weeks 7–15) and the relative gap in letter-naming fluency was reduced, EK students remained over a standard deviation behind their peers at the end of the fall term. Moreover, the statistically equivalent group trajectories that followed in the post-winter-break period served to maintain the performance disparity. These results suggest that the provision of 30 minutes of daily supplemental instruction may boost the rate at which struggling learners acquire preliteracy skills, but delivery of a Tier II supplement may not be sufficient to ameliorate a large initial performance disparity. The consequence of the sustained performance differential is not entirely clear, however, as the strong and struggling learners transitioned to a negative growth function at virtually the same point toward the end of the academic year. If it can be assumed that the leveling and modest decline in letter-naming fluency reliably marks a transition to the increased acquisition of a more advanced early literacy skill (e.g., phoneme segmentation), efforts to reduce a preliteracy skill gap may be of secondary importance to ensuring that struggling learners instead attain the threshold level of performance requisite for continued developmental advancement.

## Implications

The within- and between-group ITS-based comparisons that arise within the context of RtI provide system stakeholders with a means for generating tangible evidence regarding the impact associated with the provision of supplemental instruction. The presence of individual differences during key time frames and instructional exposures also offers an opportunity to systematically examine whether background and contextual characteristics are predictive of response variability. In the current example, the modeling of LNF outcomes as a function of demographic predictor variables yielded only a limited number of statistical relationships. The relative lack of relationships tentatively suggests that background characteristics were somewhat immaterial to student performance.[6] Yet, with a more focused data-collection effort, it would be possible to ascertain whether other student attributes or dispositions (e.g., student engagement) or aspects of the instructional environment (e.g., instructional fidelity) associate with interindividual differences in learning outcomes. The identification of variables that reliably covary with individual or group responsiveness to Tier II instruction would enable school personnel to closely monitor students most at risk of not responding to supplemental instructional practice and more rapidly intervene in cases where increasingly intensive and targeted supports are needed.

In addition to the potential for systematic investigation of interindividual differences, the statistical analysis of RtI progress-monitoring data may also enable researchers and stakeholders to gain a more thorough understanding of student learning patterns. With the collection of weekly, biweekly, or monthly assessment scores, linear developmental trends no longer have to be assumed, as more complex functions can be fit to the time-series data. In the current application, nonlinear growth was observed in conjunction with the winter break and during the post-winter-break period. Over the 3-week winter recess, letter-naming fluency declined before resuming an upward (nonlinear) trend with reinstatement of Tier I and II instruction in the post-winter-break period. The zigzag pattern of performance is generally consistent with findings in the summer-learning literature regarding the in- and out-of-school performance of struggling readers (Alexander, Entwisle, & Olson, 2001; Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996; Heyns, 1987; Zvoch & Stevens, 2015), but the nonlinear growth pattern also suggests that benchmark goals derived from a linear change projection may need to be revised and student growth expectations may need to be raised or lowered at different points in the academic year (Christ, Silberglitt, Yeo, & Cormier, 2010; Nese et al., 2011).

The discovery that literacy or math skill development follows a nonlinear trajectory may therefore enable educators to recognize the need for more intensive instructional support during periods when rapid growth is expected or to shift the instructional focus when a point of diminishing returns is reached.

## Qualifications and Limitations

The application of piecewise growth models to the ITS data that arises with the delivery of RtI-based instruction provides a more sophisticated means for examining and evaluating student response to intervention. However, it should be reiterated that the strength of inference regarding treatment effects not only depends on the strength of the design, but also on the absence of any spurious relationship coincidental with the intervention point and statistical conclusion validity in modeling the string of discontinuous growth segments. In practice, the availability of multiple data points adjacent to and contiguous with the treatment intervention, access to a comparison time series, and correct modeling of functional form will often be effective in reducing, controlling, or eliminating most major ITS validity threats. Nonetheless, the ability to adequately model and evaluate period-specific growth trajectories and identify associated treatment effects or population subgroup differences will also be limited by the amount of statistical power available to detect such effects.

In a growth-modeling context, statistical power is impacted by the size of the analytic sample, the frequency of observation during specific time periods of interest, the order of the polynomial(s) under consideration, the amount of within- and between-person variability in growth segments, and the magnitude of growth-parameter or group difference to be detected. All else equal, statistical power is higher when group mean or parameter differences are greater and samples are larger. The power to detect statistical effects will also be enhanced as the reliability of growth-parameter estimates is increased. Greater observation frequency, longer segment duration, and less within- and between-person variability (within treatments/groups) boost parameter reliability and thereby, statistical power (see Muthén & Curran, 1997; Raudenbush, 2008; Raudenbush & Lui, 2001; Zhang & Wang, 2009). Careful consideration of the design context associated with a particular RtI application is thus recommended to ensure that the measurement schedule and sample characteristics are adequate for estimating period-specific growth segments and testing hypotheses of interest.

In the present example, the timing of the supplemental intervention was linked with the schedule of student observations and the structure of the academic calendar to inform the piecewise model specification. With the rich set of assessment scores, the relatively large analytic sample, and the expectation that the provision of supplemental instruction would have both short- and longer-term impacts, the individual growth model contained seven terms that represented multiple time segments and allowed testing of various hypotheses. Model estimates generally supported expectations regarding the positive impact of schooling and the value added by supplemental instruction on student learning trajectories. However, an immediate change in LNF level after the onset of Tier II instruction was not observed and the reversal of LNF performance levels toward the end of the academic year were not anticipated. The lack of an immediate detectable change in LNF level may be attributed in part to the initial novelty and attendant disruption in student schedules and the organizational realities that accompany the implementation of a supplemental intervention. Conversely, the deceleration and reversal in LNF performance may reflect a developmental process that until multiple time-structured measures became available was assumed to progress in positive linear increments. The pattern of results could also be particular to the specific fluency skill that served as the basis of the demonstration, including possible practice and form effects (e.g., Francis et al., 2008) and/or the unique characteristics of the sample of kindergartners that were studied.

Another study aspect that merits consideration is the absence of data on the fidelity with which instruction was delivered. By design, students were to receive explicit, direct instruction by trained and experienced teachers. The use of a scripted instructional framework by veteran instructional staff was intended to minimize deviation from program protocol and ensure a consistent, high-quality instructional experience for students. Previous research on the delivery of direct literacy instruction to early struggling readers attending summer school in the same school district provides some indirect evidence on the fidelity with which direct literacy instruction is typically delivered (Zvoch, 2012). In the previous study, a high-quality, uniform implementation of instructional protocol was observed. However, without data that capture the extent to which classroom instruction was implemented as designed in the current investigation, inferences regarding the size and variability of RtI-based impacts are necessarily limited. As a result, additional study with different samples and measures and robust collection of fidelity of implementation data is recommended to shed further light on the extent to which the patterns observed herein generalize to similar and/or diverse RtI contexts.

Lastly, it should be noted that although RtI typically involves three instructional tiers and a fluid provision of services, the current demonstration was based on the application of a two-tier RtI model with a uniform start date and a standard instructional protocol (e.g., all struggling learners received EK starting in week 7 of the academic year). In the more common application, supplemental tiered services can be initiated with different students for different durations with differential intensity between different points in time during the academic year. When the RtI model is implemented in this fashion, a staggered ITS design develops. Staggered ITS designs are more complex than a standard ITS design as the point of intervention and instructional dosage may differ by student. Nonetheless, the structure of the staggered design can be accommodated by the piecewise modeling framework by matching intervention dates with the academic year calendar and calculating the number of elapsed instructional days or weeks prior to the start and for the duration of supplemental instruction. To apply, coding schemes individualized to reflect the treatment regime experienced by specific students need to be specified. The staggered ITS/PGM model is then applicable to a variety of RtI models and in any number of educational, social, and health contexts (e.g., cardiovascular fitness monitoring, substance abuse and recovery) where a developmental outcome is followed over time, the number of measurement occasions is sufficient to establish pre- and postintervention trends, and the timing, nature, and duration of an intervention for a group of treatment recipients are recorded.

## Conclusion

When an RtI model is implemented, individual response to intervention may differ initially and over time. Variability in outcomes may be due to characteristics of the individual and/or characteristics of the treatment context. Beyond descriptively tracking outcomes, understanding whether and to what extent supplemental instruction impacts treatment group outcomes and why different students respond in different ways should be a goal of researchers and practitioners. The piecewise growth models presented herein provide a systematic means to represent and evaluate the dynamic and fluid nature of contemporary instructional practice. Specification of parameters for each segment of the time series enables estimation of distinct growth-trajectory components, facilitates tests of level and slope discontinuities, and permits examination of hypotheses regarding the predictors of change between discrete time points. Nonetheless, the statistical modeling of RtI progress-monitoring data represents a clear departure from the descriptive approaches currently used by practitioners to track and evaluate student performance. At the same time, the explicit modeling of RtI data opens the potential for new collaborative opportunities that ultimately may strengthen the learning outcomes for students struggling with the acquisition of academic skills. To further the evaluation and enhance understanding of RtI instructional practice, researchers and analysts are encouraged to map the episodic structure of the instructional regimes delivered to struggling learners using contextually diverse data sets that include a variety of outcome measures and indicators describing the attributes of students and the components of instruction and their delivery.

## Notes

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A090369. The opinions expressed are the author's and do not represent the views of the Institute of Education Sciences or the U.S. Department of Education. Keith Zvoch is an associate professor in the Department of Educational Methodology, Policy, and Leadership at the University of Oregon. Address all correspondence to Keith Zvoch, University of Oregon, Department of Educational Methodology, Policy and Leadership, College of Education, 5267 University of Oregon, Eugene, OR 97403–5267; e-mail: kzvoch@uoregon.edu.

1. Three district elementary schools repeatedly assessed all kindergarten students regardless of initial or subsequent performance and treatment assignment. The universal progress monitoring assessments facilitated estimation of the two-group ITS model described below.

2. The district uses a two-tiered RtI model during kindergarten. A more traditional three-tiered model is instituted starting in first grade.

3. Exit LNF status was evaluated in mid-May, approximately 4 weeks prior to the end of the academic year. The mid-May assessment serves as the third universal LNF screener, so the majority of students (93%) had a score at this reference point.

4. In preliminary analyses, a quadratic term was also specified to test for the presence of a nonlinear relationship during the initial intervention period. The nonlinear term was not statistically related to LNF performance in either the one or two group ITS model and was not included in subsequent models.

5. For brevity, with the large number of statistically nonsignificant findings, coefficients for student predictor variables are not presented. Complete model results are available from the author upon request.

6. Some may wonder if the lack of subgroup differences could be due to interrelationships among the set of level 2 predictors. In the current sample, the largest bivariate correlation

(.30) was between English learner and ethnic minority status. All other correlations were < |.16|.

## References

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis*, **23**(2), 171–191.

Bloom, H. S. (2003). Using "short" interrupted time series analysis to measure the impacts of whole-school reforms: With applications to a study of accelerated schools. *Evaluation Review*, **27**, 3–49.

Bradley, M. C., Daley, T., Levin, M., O'Reilly, R., Parsad, A., Robertson, A., & Werner, A. (2011). *IDEA national assessment implementation study* (NCEE 2011-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment*, **23**, 381–394.

Busch, T. W., & Reschly, A. L. (2007). Progress monitoring in reading: Using curriculum-based measurement in a response-to-intervention model. *Assessment for Effective Intervention*, **32** (4), 223–230.

Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, **24**, 409–429.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools*, **32**, 38–50.

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-based measurement of oral reading: An evaluation of growth rates and seasonal effects among students served in general and special education. *School Psychology Review*, **39**, 447–462.

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., Cho, E., & Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology*, **102**(2), 327–340.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology*, **98**(2), 394–409.

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, **66**(3), 227–268.

Deno, S. L., Reschly, A. L., Lembke, E. S., Magnusson, D., Callender, S. A., Windram, H., & Stachel, N. (2009). Developing a school-wide progress-monitoring system. *Psychology in the Schools*, **46**(1), 44–55.

Denton, C. A. (2012). Response to intervention for reading difficulties in the primary grades: Some answers and lingering questions. *Journal of Learning Disabilities*, **45**(3), 232–243.

Duncan, T. E., & Duncan, S. C. (2004). A latent growth curve modeling approach to pooled interrupted time series analyses. *Journal of Psychopathology and Behavioral Assessment*, **26**(4), 271–278.

Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential to learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3–40). Mahwah, NJ: Erlbaum.

Elliott, J., Lee, S. W., & Tollefson, N. (2001). A reliability and validity study of the Dynamic Indicators for Basic Early Literacy Skills—Modified. *School Psychology Review*, **30**(1), 33–49.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis* (2nd ed.). Hoboken, NJ: Wiley.

Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives*, **3**(1), 30–37.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at risk children. *Journal of Educational Psychology*, **90**(1), 37–55.

Francis, D. J., Santi, K. L., Barr, C., Fletcher, J. M., Varisco, A., & Foorman, B. R. (2008). Form effects on the estimation of students' oral reading fluency using DIBELS. *Journal of School Psychology*, **46**, 315–342.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, **41**(1), 93–99.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities*, **45**(3), 195–203.

Fuchs, D., Vaughn, S., & Fuchs, L. S. (2008). *Response to intervention*. Newark, DE: International Reading Association.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice*, **18**, 157–171.

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 for reading in the primary grades; a practice guide* (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

GlobalScholar/Spectrum K12, AASA, CASE, NASDSE, & RTO Action Network/NCLD. (2011). *Response to Intervention (RTI) adoption survey 2011*. Seattle: Author.

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

Good, R. H., Kaminski, R. A., Simmons, D., & Kame'enui, E. J. (2001). *Using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model: Steps to reading outcomes*. Eugene: University of Oregon.

Gordon, R. A., & Heinrich, C. J. (2004). Modeling trajectories in social program outcomes for performance accountability. *American Journal of Evaluation*, **25**(2), 161–189.

Heyns, B. (1987). Schooling and cognitive development: Is there a season for learning? *Child Development*, **58**, 1151–1160

Hindman, A. H., Cromley, J. G., Skibbe, L. E., & Miller, A. L. (2011). Conventional and piecewise growth modeling techniques: Applications and implications for investigating Head Start children's early literacy learning. *Evaluation Review*, **35**(3), 204–239.

Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Processing. *School Psychology Review*, **32**(4), 541–556.

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

McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology*, **98**(1), 14–28.

Mellard, D. F., Stern, A., & Woods, K. (2011). RTI school-based practices and evidence-based models. *Focus on Exceptional Children*, **43**(6), 1–15.

Moore, K. J., Osgood, W., Larzelere, R. E., & Chamberlin, P. (1994). Use of pooled time series in the study of naturally occurring clinical events and problem behavior in a foster care setting. *Journal of Consulting and Clinical Psychology*, **62**, 718–728.

Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, **2**, 371–402.

National Early Literacy Panel. (2008). *Developing early literacy*. Washington, DC: National Institute for Literacy.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Pub. No. 00-4754). Rockville, MD: National Institutes of Health.

Nese, J. F. T., Biancarosa, G., Anderson, D., Lai, C. F., Alonzo, J., & Tindal, G. (2011). Within-year oral reading fluency with CBM: A comparison of models. *Reading and Writing: An Interdisciplinary Journal*, **25**(4), 887–915.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Osgood, D. W., & Smith, G. L. (1995). Applying hierarchical linear modeling to extended longitudinal evaluations: The Boys Town follow-up study. *Evaluation Review*, **19**(1), 3–38.

Piasta, S. B., Petscher, Y., & Justice, L. M. (2012). How many letters should preschoolers in public programs know? The diagnostic efficiency of various preschool letter-naming benchmarks for predicting first-grade literacy achievement. *Journal of Educational Psychology*, **104**(4), 945–958.

Piasta, S. B., & Wagner, R. K. (2010). Developing emergent literacy skills: A meta-analysis of alphabet learning and instruction. *Reading Research Quarterly*, **45**, 8–38.

Puolakanaho, A., Ahonen, T., Aro, M., Eklund, K., Leppnen, P. H. T., Poikkeus, A. M., Tolvanen, A., Torppa, M., & Lyytinen, H. (2007). Very early phonological and language skills: Estimating individual risk of reading disability. *Journal of Child Psychology and Psychiatry*, **48**(9), 923–931.

Raudenbush, S. W. (2008). Many small groups. In J. de Leeuw & E. Meijer (Eds.), *Handbook of multilevel analysis* (pp. 207–236). New York: Springer.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, A. S., Fai, Y. F., Congdon, R. T., & du Toit, M. (2011). *HLM: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Raudenbush, S. W., & Lui, X. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, **6**(4), 387–401.

Ritchey, K. D., & Speece, D. L. (2006). From letter names to word reading: The nascent role of sublexical fluency. *Contemporary Educational Psychology*, **31**, 301–327.

Rouse, H. L., & Fantuzzo, J. W. (2006). Validity of the Dynamic Indicators for Basic Early Literacy Skills as an indicator of early literacy for urban kindergarten children. *School Psychology Review*, **35**(3), 341–355.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, **96**(2), 265–282.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. Oxford: Oxford University Press.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

St. Clair, T., Cook, T. D., & Hallberg, K. (2014). Examining the internal validity and statistical precision of the comparative interrupted time series design by comparison with a randomized experiment. *American Journal of Evaluation*, **35**(3), 311–327.

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology*, **40**, 7–26.

Zhang, Z., & Wang, L. (2009). Statistical power analysis for growth curve models using SAS. *Behavior Research Methods*, **41**(4), 1083–1094.

Zvoch, K. (2012). How does fidelity of implementation matter? Using multilevel models to detect relationships between participant outcomes and the delivery and receipt of treatment. *American Journal of Evaluation*, **33**(4), 541–559.

Zvoch, K., & Stevens, J. J. (2015). Identification of summer school effects by comparing the in- and out-of-school growth rates of struggling early readers. *Elementary School Journal*, **115**(3), 433–456.