



Published in final edited form as:

*Elem Sch J*. 2011 June ; 111(4): 535–560. doi:10.1086/659031.

## Assessment Data-Informed Guidance to Individualize Kindergarten Reading Instruction: Findings from a Cluster-Randomized Control Field Trial

Stephanie Al Otaiba, Carol M Connor, Jessica Sidler Folsom, Luana Greulich, Jane Meadows, and Zhi Li

Florida Center for Reading Research at Florida State University

### Abstract

The purpose of this cluster-randomized control field trial was to examine the extent to which kindergarten teachers could learn a promising instructional strategy, wherein kindergarten reading instruction was differentiated based upon students' ongoing assessments of language and literacy skills and documented child characteristic by instruction (CXI) interactions; and to test the efficacy of this differentiated reading instruction on the reading outcomes of students from culturally diverse backgrounds. The study involved 14 schools and included 23 treatment ( $n = 305$  students) and 21 contrast teacher ( $n = 251$  students). Teachers in the contrast condition received only a baseline professional development that included a researcher-delivered summer day-long workshop on individualized instruction. Data sources included parent surveys, individually administered child assessments of language, cognitive, and reading skills and videotapes of classroom instruction. Using Hierarchical Multivariate Linear Modeling (HMLM), we found students in treatment classrooms outperformed students in the contrast classrooms on a latent measure of reading skills, comprised of letter-word reading, decoding, alphabetic knowledge, and phonological awareness ( $ES = .52$ ). Teachers in both conditions provided small group instruction, but teachers in the treatment condition provided significantly more individualized instruction. Our findings extend research on the efficacy of teachers using Individualized Student Instruction to individualize instruction based upon students' language and literacy skills in first through third grade. Findings are discussed regarding the value of professional development related to differentiating core reading instruction and the challenges of using Response to Intervention approaches to address students' needs in the areas of reading in general education contexts.

### Keywords

Response to intervention; kindergarten; reading instruction; child by instruction interactions

---

In 2005, the National Institute of Child Health and Human Development funded several Learning Disability Center Grants and challenged these projects to extend the knowledge base about how to prevent, as well as to identify, reading disabilities using *Response to Intervention* approaches. Such knowledge is vital because far too many children struggle to

learn to read primarily because they do not receive adequate reading instruction in the primary grades (Vellutino et al., 1996), and subsequently do not succeed in school. Over a third of all American fourth graders, and even more disturbingly, over half of students from minority backgrounds, performed below basic on the reading comprehension portion of the National Assessment of Educational Progress (NAEP; National Center for Educational Statistics, 2007). It is alarming that so many students face difficulties comprehending grade level texts because converging evidence suggests it is challenging to remediate the reading problems of older students and subsequently, their motivation to read and their self efficacy is negatively affected (e.g., Klingner, Vaughn, Hughes, Schumm & Elbaum, 1998; Morgan & Fuchs, 2007).

Fortunately, several seminal research reports have shown that early intervention efforts using evidence-based reading programs can be highly successful in preventing reading difficulties for most students (National Reading Panel, 2000; National Early Literacy Panel, 2008). In light of this research base, general and special education policy in the United States has shifted; for example, the Individuals with Disabilities Educational Improvement Act (IDEA; 2004) allows school districts to use up to 15% of special education funds for prevention and early intervention. In addition, to encourage schools to provide intervention to struggling readers before they fall far enough behind to qualify for special education, IDEA supports a shift to documenting students' response to evidence-based instruction and intervention in order to identify which students have a reading disability (Pub. L. No. 108–446 § 614 [b][6][A]; § 614 [b] [2 & 3]). This new identification process is known as Response to Intervention, or RTI. There are different RTI models, but the common foundation for success in any model is effective beginning reading instruction, that is provided by classroom teachers (Gersten et al., 2008). The purpose of the present study was to learn the extent to which kindergarten teachers could learn a promising instructional strategy, wherein kindergarten reading instruction was differentiated based upon students' ongoing assessments of language and literacy skills and to test the efficacy of differentiated reading instruction on the reading outcomes of students from culturally diverse backgrounds.

## What Is RTI?

Current research on RTI demonstrates that multi-tiered models of early literacy intervention can reduce referrals to special education and improve the accuracy in identifying individuals with reading disabilities (e.g., Mathes, Denton, Fletcher, Anthony, Francis, & Schatschneider, 2005; Speece & Case, 2001; Torgesen et al., 1999; Vaughn & Fuchs, 2003; Vellutino et al., 1996; Wanzek & Vaughn, 2007). Torgesen (2002) argued that a well-implemented multi-tiered RTI approach could reduce the prevalence of severe word-level reading disabilities to about six percent within most elementary schools. All multi-tiered RTI models begin with effective initial reading instruction, or Tier 1. Components of effective Tier 1 include: (1) a core reading program grounded in scientifically-based reading research, (2) assessment of all students to determine instructional needs and identify students who are at risk, and (3) ongoing, targeted professional guidance to provide teachers with the necessary tools to differentiate or individualize instruction to ensure most students learn to read (Gersten et al., 2008). Well-implemented Tier 1 instruction should help at least 80% of

students meet grade-level reading proficiency (Batsche, Curtis, Dorman, Castillo, & Porter, 2007), although this percentage may vary somewhat across schools depending on student need and also upon the definition for grade-level proficiency or for responsiveness. To date, researchers have used several response criteria (see Fuchs & Deschler, 2007 for a thorough discussion), but one of the mostly commonly used, that is also applied in the present study, is normalization - reading scores above an 85 on standardized, nationally-normed test of reading (i.e., within one standard deviation of the normative sample).

Thus, even with well-implemented Tier 1, some students are likely to need more assistance and so would receive Tier 2 interventions. Tier 2 interventions, provided as a supplement to Tier 1, involve classroom teachers, specialists, or even paraprofessionals providing small-group reading intervention that are individualized for the needs of students, that offer frequent opportunities for feedback and practice. Tier 2 is typically delivered three to five days per week and students' progress is assessed more frequently than in Tier 1. Students who do not respond to Tier 2 then receive more intensive supplemental Tier 3 interventions (up to 60 mins a day in groups of 1–3 students). The number of Tiers (and length of time students spend in tiers) varies across models, but the underlying theory is that students who struggle with reading require more time in explicit and systematic instruction in their assessed areas of weakness. Further, because all students receive Tier 1, whereas only a few students also receive additional intervention, a well-implemented Tier 1 is the foundation of all RTI. Then, if these specialized interventions, supplementing Tier 1, are still not powerful enough, identification of a reading disability with participation in special education would likely follow.

Many schools that have been implementing RTI-like approaches for more than two decades have identified some challenges related to Tier 1 that impact the efficiency of the entire RTI system. Chief among these are the need to ensure that teachers provide instruction and interventions that are sufficiently intense and implemented with fidelity (e.g., Ikeda, Rahn-Blakeslee, Niebling, Gustafson, Allison, & Stumme, 2007), and the need to connect assessment data to interventions (Marston, Pickart, Reschly, Heistad, Muyskens, & Tindal, 2007). Historically, researchers in the fields of special education and school psychology have demonstrated that it is challenging to help classroom teachers to use assessment data to guide small group instruction or to individualize interventions based on students' strengths and weaknesses (e.g., Fuchs, Fuchs, Hamlett, Phillips, & Bentz, 1994). Even with support, researchers found that most general educators tended to provide the same activities to all children, rather than using data to provide different types of activities to students with weaker skills. More recently, in the context of RTI, researchers have rekindled their efforts to assist teachers in using assessment data to individualize beginning reading instruction and interventions (Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007; Gersten et al., 2008; Mathes et al., 2005; Scanlon, Gelzheiser, Vellutino, Schatschneider, & Sweeney, 2008; Vellutino et al., 1996).

Providing individualized amounts and types of instruction based upon students' assessed skills and needs is vital in light of accumulating knowledge that the effect of specific amounts and types of instruction depend on students' vocabulary and reading skills. That is, students' response to any instruction will likely depend in part on how their initial language

and literacy skills interact with the instruction they receive. This idea of child characteristic X instruction interactions has been confirmed in a number of studies in preschool (Connor, Morrison, & Slominski, 2006), first grade (Connor, Morrison, & Katch, 2004; Foorman, Francis, Fletcher, Schatschneider & Mehta, 1998; Juel & Minden-Cupp, 2000), second grade (Connor, Morrison & Underwood, 2007) and most recently in kindergarten (Al Otaiba et al., 2008). Many of these studies use a multi-dimensional framework to describe classroom instruction. This is with the understanding that effective teachers are frequently observed to use multiple strategies even if the school core curriculum emphasizes a particular type of instruction (Connor et al., 2009). Subsequently, a small but growing number of cluster randomized control studies in schools conducted by Connor and colleagues have shown that helping teachers provide the amount and type of instruction tailored to students' skills, as informed by assessment data, results in stronger reading growth (Connor et al., 2007; Connor et al., 2009).

## **Rationale for the Present Study: Dimensions of Instruction and Individualizing Student Instruction**

Well-implemented Tier 1 instruction is the foundation to successful RTI because students who do not make adequate growth through Tier 1 will need to be provided more expensive specialized interventions. Thus, because students enter school with vastly different background knowledge, having experienced different language and literacy experiences at home, the first line of defense against reading disabilities is their initial classroom literacy instruction. As a result, kindergarten classrooms are becoming increasingly academic in focus. Our rationale for this cluster-randomized control field trial was to examine whether kindergarten teachers could improve the beginning literacy trajectories of their students by implementing a hybrid of Tier 1 instruction and Tier 2 small group individualized intervention in their classrooms that was assessment data-informed.

Our conceptualization of classroom reading instruction is multi-dimensional and draws on converging evidence that the effect of a particular type of instruction depends upon students' language and literacy skills (Al Otaiba et al 2008; Connor et al., 2007; 2009; Foorman et al., 2003; Cronbach & Snow, 1977; Juel & Minden-Cupp, 2000). Grouping is one dimension of instruction: specifically, researchers have found that small group instruction is relatively more powerful than whole class (Connor et al., 2006). During whole class instruction, by necessity, most teachers teach to the "middle;" in contrast, small group instruction lends itself to differentiation, which is associated with stronger reading growth, particularly for the more struggling and also better achieving students. The second dimension, the content of instruction, considers two types of instruction that are essential for reading: code-focused or meaning-focused. Code-focused instruction addresses phonological awareness, print knowledge, and beginning decoding; whereas meaning-focused instruction addresses vocabulary development, listening and reading comprehension. Research has shown that the amount of each type of content needed to be a successful reader depends upon students' language and literacy skills (Al Otaiba et al., 2008; Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007; Connor, 2009). The final dimension is management: instruction is considered child-managed when students are working independently or with a

peer and teacher-child managed when the teacher is actively instructing. These three dimensions function such that instruction can be whole class or small group, teacher-child or child-managed, and code- or meaning-focused.

There is a growing body of evidence that teachers who use these dimensions to individualize instruction based upon students' language and literacy skills in first through third grade achieve stronger student reading performance than teachers who do not (Connor et al., 2004; Connor et al, 2009; Connor Piasta et al., 2009). Specifically, the Individualized Student Instruction intervention was designed by Connor and colleagues to support teachers' ability to use assessment data to inform instructional amounts, types, and groupings. Child assessment data and data from classroom observations were used to develop and fine tune algorithms that use a predetermined end-of-year target outcome. The students' assessed language and reading scores were entered into the Assessment to Instruction (A2i) software, which then computes the recommended amounts of instruction in a multidimensional framework of teacher- or child-managed instruction that is either code- or meaning-focused. Thus, the ISI intervention includes three components: A2i software, ongoing teacher professional development, and in-class support.

Although ISI as a means of PD has been carefully studied in first through third grades, the present study extends this research in several innovative ways that are very relevant for RTI. First, consistent with the notion of early intervening, this is the initial field test of ISI at kindergarten, when formal reading instruction begins for most children. Second, our treatment aimed to help teachers differentiate their core reading instruction, essentially creating within their classrooms a hybrid of Tier 1 (classroom instruction) and Tier 2 (targeted and differentiated small group interventions). Third, teachers in both the ISI-K treatment and the professional development contrast conditions received a common baseline of professional development that included (1) a researcher-delivered summer day-long workshop on RTI and individualized instruction, (2) materials and games for center activities, and (3) data on students' reading performance provided through Florida's Progress Monitoring Reporting Network (PMRN), the states' web-accessed data base. At the time of the study, these data included students' scores and risk status on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002). Furthermore, schools all had reading coaches who worked with the teachers. Thus, our test of the ISI-K treatment represents a rigorous test of all three components of the ISI regimen used in prior investigations of ISI efficacy A2i software, ongoing professional development, and bi-weekly classroom-based support. Given the newness of ISI, it is also important to learn if and how it works to change teacher behavior, and through that change, to improve students' responsiveness to initial reading instruction. Fall and winter classroom observations of both treatment and contrast classrooms were used to investigate potential changes in practice.

## Research Questions

The purpose of this cluster-randomized control field trial was to examine the effects of ongoing professional development, guided by ISI-K techniques, the A2i software recommendations and planning tools, and in-classroom support, on the kindergarteners' reading outcomes. Specifically, we addressed two research questions:

1. Is there variability in the implementation of literacy instruction and individualization of instruction within ISI-K treatment and contrast professional development classrooms? We hypothesized that teachers in the treatment condition would provide more small-group individualized literacy instruction relative to teachers in the contrast condition.
2. Would students in the treatment classrooms demonstrate stronger reading outcomes than students in the contrast classrooms? We hypothesized that the kindergarten students in the ISI-K treatment classrooms would demonstrate stronger reading outcomes. Relatedly, we hypothesized that a higher proportion of students would be responsive to individualized instruction, which we defined as reaching grade-level word-reading skills. If so, fewer students in ISI-K classrooms would be expected to need additional interventions in first grade.

## Method

### Participants

This cluster-randomized control field trial took place during the second year of a large-scale Learning Disabilities Center project investigating the efficacy of Tier 1 core kindergarten reading instruction. For the present study, the school district in a mid-size city in northern Florida nominated 14 schools to be recruited and the principals all agreed to participate and to be randomly assigned to the ISI-K treatment (hereafter *treatment*) or to a wait-list contrast professional development (hereafter *contrast*) condition. These schools served an economically and ethnically diverse range of students; six schools received Title I funding and four received Reading First funding. Although the schools served a culturally diverse population, the percentage of the schools' students who were identified as Limited English Proficient (LEP) was not typical for the state, and ranged from less than 1% to 4.5%.

Prior to assigning schools to condition, we matched them on several criteria: proportion of students who received free or reduced price lunch (as a proxy for socioeconomic status), Title I and Reading First participation (which entitled schools to extra resources), and the schools' reading grades based upon the proportion of students passing the Florida high stakes reading test at third grade (as a proxy for overall response to reading instruction K-3). One of each pair was then randomly assigned to the ISI-K treatment or to the wait-list contrast condition. The present study describes pre- and post-test data from the 2007–2008 school year involving 7 treatment schools ( $n = 23$  teachers) and 7 contrast schools ( $n = 21$  teachers). In the following year, all teachers in the contrast schools were provided with the full ISI-K treatment and teachers in the initial 7 treatment schools received a second year of ISI-K, but presentation of these data is beyond the scope of this present study.

Across these 14 schools, kindergarten was provided for the full-day and, as increasingly typical in North American schools, there was a strong focus on reading and language arts instruction. Per district policy, reading and language arts instruction was provided for an uninterrupted block of a minimum of 90 minutes and the instruction was guided by a core reading program that was explicit and systematic. Specifically, all but one school utilized *Open Court* as the core reading program (Bereiter et al., 2002), this remaining school used

*Reading Mastery Plus* published by SRA (Engelmann & Bruner, 2002). Further, during the present study all schools had reading coaches and all collected student reading assessment data three times per year to enter into the state's reading assessment data base. However, RTI was not yet being implemented within the district.

**Teachers and students**—A total of 44 credentialed teachers, ranging from two to five teachers per school, agreed to participate. A majority (31 teachers, 70.5%) were Caucasian, 10 (22.7%) were African American, and 3 were Hispanic. Ten teachers held graduate degrees (22.7%) and the majority held Bachelor's degrees (77.3%). On average, teachers had taught for 10.55 years ( $SD = 8.66$ ). There was only one first-year teacher, although 18 teachers reported having 0–5 years of teaching experience. Six teachers reported having between 6–10 years, 11 had 11–15 years, and 9 had more than 15 years of teaching experience. A chi-square analysis revealed no significant difference across conditions.

With the teachers' assistance, we recruited all students in their classrooms (including students who qualified for special education) and subsequently, we received consent from 605 parents. During the course of the study, 49 children (from different schools and classrooms and evenly distributed across conditions) moved to schools not participating in the study. Table 1 describes the 556 remaining students' demographics, including age, gender, ethnicity, free and reduced lunch status, retention, number of student absences during the year of the study, special education classification, IQ and vocabulary. After applying the Benjamini-Hochberg (BH) correction (Benjamini & Hochberg; 1995) for multiple comparisons, the only significant difference between groups was seen for the rate of student absences. On average, students in the treatment group were absent for about 7 days and students in the contrast group were absent about three days during the school year.

Along with the consent form, we sent parents a survey to complete regarding their home literacy practices, parent education, and whether their child had attended preschool. Table 2 shows the means and standard deviations by condition; a chi-square test revealed no significant difference across conditions. In summary, the cluster randomization process yielded samples of students in each condition who were similar in demographics, pre-treatment reading scores, and home and preschool experiences.

## Procedures

**Common baseline professional development provided to treatment and contrast conditions**—As previously stated, all teachers in both conditions received a common baseline of professional development that included a researcher-delivered summer day-long workshop on RTI and individualized instruction. We conducted separate one-day workshops for teachers in each condition to reduce potential treatment diffusion. The content and materials were identical; with one exception, namely that contrast teachers received no ISI-K or A2i training.

The first author led both workshops, initially providing an overview about RTI research, with a focus on Tier 1 and individualized instruction. Research staff assisted with presenting information about the NRP (2000) five components of reading and how these related to teachers' core reading program. Staff modeled dialogic reading (Whitehurst & Lonigan,

1998) and vocabulary instructional strategies (Beck, McKeown, & Kucan, 2002) and trained them to work in teams to select vocabulary words and to create dialogic questions using narrative and expository texts.

Next, we used Gough's Simple View of Reading (Gough & Tunmer, 1986) to cluster phonological, phonics, spelling, and fluency instruction as "code-focused" and vocabulary, comprehension, and writing as "meaning-focused" instruction. We emphasized that no one size fits all students, and described the research behind providing small group individualized instruction and the need to manage centers efficiently. Further, we reminded teachers how data, such as the DIBELS data provided through the PMRN, could be used to group children with similar instructional needs. Using the PMRN was familiar to teachers, as they had been provided DIBELS (Good & Kaminski, 2002) data (scores and risk status) for several years regarding students' initial letter naming and phonological awareness and, beginning in winter decoding skill development.

Teachers then worked in small groups using their core reading program to distinguish activities within the dimensions of code vs. meaning instruction and teacher-child managed vs. child managed instruction. We also provided colored and laminated activities developed from downloadable templates created by the FCRR for use in teacher-child managed and child managed centers (FCRR, 2007).

In summary, during the initial summer workshop, teachers in both conditions were trained about: RTI, how to individualize instruction based upon student assessment data, how to manage centers, and how to use FCRR center activities. Like all other teachers in the school district, teachers in the present study used an explicit and systematic core reading program, they had access to DIBELS data through the PMRN, and they had a reading coach at their school.

**Treatment: professional development and classroom support for A2i and ISI-K**—Professional development to conduct the ISI-K intervention began with this initial baseline professional development workshop, but also involved ongoing monthly coaching and bi-weekly classroom based support which was not provided to teachers in the contrast condition. This coaching model is supported by prior research (e.g., Gersten et al., 2008; Showers, Joyce, & Bennett, 1987) and specifically by prior ISI research (e.g., Connor et al., 2007). We also emphasized to teachers that our research goal was to learn whether ISI-K and A2i had a value-added in providing data-based guidance individualization.

We showed the teachers screenshots (see Appendix A) depicting an overview of how A2i would use a mathematical formula using their students' scores from picture vocabulary and word reading assessments to suggest homogeneous groups, and to predict recommended amounts (in minutes) in each dimension of instruction for each child. We explained that the formulas incorporated growth and were based upon data from previous correlational studies. Teachers saw screenshots (also in Appendix A) depicting what data would look like for each student- with a graph indicating a desired target outcome of grade-level language and reading performance (or for a child with strong initial skills a target of 9 months growth) and



using simulated data for fall and winter to create a trend line to show adequate and inadequate progress toward the target.

Then once a month for the next 8 months, research staff met at each school site with grade-level teams, or if needed, met with individual teachers, to train them how to use the technology. We entered all data for teachers (this included WJ-III Letter Word Identification, Picture Vocabulary scores). The first school-based meeting focused on the recommended minutes of instruction based on the student data and on scheduling minutes of instruction. This was very challenging for teachers, particularly when their only extra adult help was the biweekly research partners, so we addressed this challenge by sharing articles as well as describing and modeling “what works”. We learned that it was difficult for some teachers to move away from whole class undifferentiated instruction toward small group individualized instruction and to design child-managed center activities that were meaningful and engaging. The FCRR center activities were very helpful and widely-used. In winter, we did a mid-test using the WJ-III Picture Vocabulary and Letter Word Identification subtests and again entered data, which generated updated groupings for the teachers. Thus, each month we reinforced the project goal to individualize instruction in our meetings. During their biweekly visits, research partners reinforced the professional development, assisted if needed with technology, modeled small group strategies, and often led a center or read-aloud to students.

**Research staff and training**—A total of 19 research staff assisted with various research tasks within this large-scale project. Staff included mostly graduate students from special education and school psychology; one additional staff member was a speech and language pathologist, two were former elementary education teachers, and the remaining member held a BA in foreign language. All research staff performed one or more of the primary tasks: assessing students, videotaping teachers, and entering and analyzing data. The most highly experienced staff worked as research partners and as classroom observation video coders (coding training descriptions follow).

Research partners played at least three coaching roles during their bi-weekly classroom visits. First, they provided in-classroom support for teachers’ use of the A2i software. Second, they modeled small group individualized instruction. Third, they assessed all students’ Letter Sound Fluency once a month. Research partners were assigned to specific schools (although given the size of the project and that most schools had similar times for reading instruction, schools had more than one research partner). There were no systematic differences among schools or research partners. On average, across the study year, research partners provided a total of (2 hours a month for 8 months) of in-classroom support.

Thus, due to the size and complexity of the project, all staff members were aware of assignment to condition. Under more ideal circumstances, assessors and coders would be blind to condition. Because of this potential problem, we took steps throughout each phase of the project, and particularly prior to each wave of assessment, to remind staff that the purpose of the study was to learn which condition was more effective and to caution them that experimenter bias could undermine an otherwise very carefully planned study (e.g., Rosenthal & Rosnow, 1984). Furthermore, we instructed staff that examiner unfamiliarity

could negatively impact the reading assessment performance of kindergartners, particularly for students from minority backgrounds (Fuchs & Fuchs, 1986) and students with disabilities (Fuchs, Fuchs, Power, & Dailey, 1985). To reduce the extent to which examiner familiarity was problematic, research partners also visited contrast classrooms once a month to drop off FCRR activities, but this presence was less than in the treatment classrooms where they worked bi-weekly for an hour.

The first author trained all research staff to individually administer assessments prior to each wave of testing. During these sessions, staff learned about developing rapport to reliably assess young students as well as how to administer and score each test following the directions from the assessment examiner manuals. Prior to testing staff had to reach 98% accuracy on this checklist that evaluated the accuracy with which the assessor followed the directions in administering and scoring the assessments (adapted from Sattler, 1982).

**Assessment and scoring procedures**—In the present study, data were collected from five sources: parents, teachers, individual staff-administered child assessments, individual district-administered child assessments, and classroom observational videotapes. All students in all conditions were individually assessed on most measures by research staff in quiet areas near their classrooms. Because children were young, testing was divided into 30 minute sessions; resulting in three sessions in fall, one in winter, and two in spring. Each test protocol was scored by the member of the research staff who administered the assessment. Then all protocols were checked to ensure accurate basal and ceiling rules were followed and that addition was accurate prior to data entry. Then, for relevant subtests, compuscoring from the commercial test producer was used to calculate standard and W scores, which are Rasch ability scores that provide equal-interval measurement characteristics. W scores, which are centered at 500, represent the typical achievement for a 10-year-old. Data were entered into SPSS builder windows, and to reduce errors, this entry process was double entered, meaning that two independent entries occurred by two different data entry personnel.

Another set of measures, DIBELS, (Good & Kaminski, 2002), were administered by well-trained district personnel who also entered data into the PMRN. Teachers in both conditions could access their students' test score via a web-based interface to the Florida Progress Monitoring and Reporting Network (PMRN). The districts' testing protocol incorporated three assessment periods: The first occurred within the first 20 30 days of school, the second occurred in winter, and the third occurred in the late spring, ending the week prior to our posttesting. We accessed the DIBELS data through a query process following IRB approval, and for the purposes of this study, analyzed Phoneme Segmenting Fluency and Nonsense Word Fluency

**Videotaping training**—Research staff who became videographers attended a two hour group training session prior to each round of videotaping. The training session discussed the purpose of videotaping and provided examples of detailed notes and guidance about taking detailed field notes about classroom instruction. Staff members were taught how to operate the videotaping equipment, to record accurate target student descriptions, and to take detailed notes about these targets. This was important for two reasons: (1) to ensure that we

did not film children without parental consent; and (2) to ensure that we did capture each target student's instruction. It was crucial that all student instructional activities were described as well as management (e.g., teacher/child-managed, child-managed) and grouping (e.g., whole group, small group, peer, individual). After video taping, videos were uploaded onto a dedicated computer equipped with the Noldus Observer program.

**Coding training**—All coders had or were pursuing a graduate level degree in education or speech and language. Coders participated in a careful training process conducted in small groups and individually. We benefited from guidance at the start of this process from Connor's coding team, who had extensive experience using this coding scheme in first grade (Connor et al., 2009) and from attending a 2 day training session presented by a Noldus Observer consultant to better understand the software features.

First, coders were trained on the content of the manual through review. Second, the coders sat with an experienced coder to observe the coding system. Third, the coder was assigned a tape to code independently. Next, the reliability was then obtained using Cohen's kappa. The reliability for each coder was checked against a master coder and then the other coders. Coders could not code independently until a kappa of .75 was acquired. The reliability of the coders ranged from .77–.83 with a mean of .80. Coding meetings were held weekly to discuss any coding issue or question about a specific activity. During the coding meetings disagreements were resolved by the master coder.

In addition, these same coders were trained to complete a low-inference instrument, after watching each tape, that evaluates overall effectiveness of implementation of literacy instruction captured on the videotapes. The scale ranged from 0 – 3 with (0) for content that was not observed, (1) for “not effective,” (2) for “moderately,” and (3) for “highly effective.” The inter-rater reliability ranged from .92 to 100%, with a mean of .98 and kappas ranged from a low of .64 to 1.00.

## Teacher and Student Measures

**Teacher fidelity of individualized instruction, fidelity of A2i usage, and instructional effectiveness**—Research staff videotaped reading instruction in all 44 classrooms in November and in February of 2007. These video recordings ranged in length from 60 120 minutes and averaged 90 minutes. We carefully scheduled these observations in advance to meet teachers' convenience and teachers were given the opportunity to reschedule as needed. The videos focused on a stratified sample of 10 children; to select students for observation we rank ordered students on their fall Letter Word Identification scores (Woodcock Johnson-III Tests of Achievement; WJ-III; Woodcock, McGrew, & Mather, 2001) and randomly selected low-, average-, and high-performing target students from the class. During video taping, staff used two digital video cameras with wide-angle lenses to best capture classroom instruction. In addition, detailed field notes were kept on students, particularly those who were “off camera.” For example “Child A was reading a book in the book corner” or “Child H went to the restroom.” If a target child was absent, a substitute child with the same ranking was used as a replacement for the day.

In addition to video-coding, to address the overall effectiveness of implementation of literacy instruction captured on the videotapes, we created a low-inference observational instrument with a scale of 0 – 3 with (0) for content that was not observed, (1) for “not effective,” (2) for “moderately,” and (3) for “highly effective.” Specifically, coders first used the scale to rate individualization of instruction. Second, coders used the same scale to rate teachers’ warmth and sensitivity, classroom organization, and the degree to which teachers were effective at keeping students on-task during instruction. Third, coders reported the instructional effectiveness of code focused (letter-sound, phonological awareness, word identification, fluency, spelling) and meaning focused (vocabulary, comprehension, and writing) instruction.

Additionally, as part of the professional development protocol, to assess ongoing fidelity of implementation of A2i and ISI-K, we asked research partners to rate their ISI teachers only on use of A2i components and software using a low inference Likert scale checklist used in prior ISI research that ranged from 1 (consistently weak) to 6 (exemplary). Ratings occurred in January, March, and April. These ratings were based on the research partner’s collective experiences with the teacher, thus it would not be feasible for more than one research partner to conduct an inter-rater reliability analysis, since the rating was not based on one single observation time period.

**Student measures**—We administered a battery of assessments related to students’ language and conventional literacy skills using both criterion and norm-referenced tests. Students’ letter sound correspondence was assessed using the AIMSWEB Letter Sound Fluency (Shinn & Shinn, 2004). In this task, children are presented an array of 10 rows of 10 lower case letters per line and are asked to name as many of the sounds that the letters make as quickly as they can within 1 minute. Testing is discontinued if the child can not produce any correct sounds in the first 10. Raw scores are reported and alternate form reliability is .90. Scores from the fall and spring administration are used in the present study.

To assess students’ expressive vocabulary growth, we selected the Picture Vocabulary (PV) subtest of the Woodcock Johnson-III Tests of Achievement (Woodcock et al., 2001). In this subtest, students name pictured objects which increase in difficulty. Testing is discontinued after 6 consecutive incorrect items. According to the WJ-III test authors, reliability of this subtest is .77. The test (and all other WJ-III subtests used in the present study) yields a standard score with a mean of 100 and a standard deviation of 15 and W-scores, which are a Rasch ability scale score, that provide an equal-interval form of measurement. These W-scores are centered at 500 and this score represents average achievement for a 10-year-old student.

Students’ word reading skills were assessed using WJ-III Letter Word Identification (Woodcock, et al., 2001) subtest. Examiners asked students to identify letters in large type, and then to read words in arrays of about 8 per page. The subtest consists of 76 increasingly difficult words. Testing is discontinued after 6 consecutive incorrect items. Inter-rater reliability is high for this age group ( $r = .91$ ); concurrent intercorrelations with the WJ-III Word Attack and Passage Comprehension subtests are .80 and .79 respectively.

We assessed students' ability to decode pseudowords using the Word Attack subtest of the WJ-III (Woodcock et al., 2001). The initial items require students to identify the sounds of a few single letters; remaining items require the decoding of increasingly complex letter combinations that follow regular patterns in English orthography, but are nonwords. Testing continues until the examinee makes six errors on a page. Word Attack has a high test-retest reliability ( $r = .94$ ) for this age group.

We also accessed spring scores from the district-administered DIBELS Nonsense Word Fluency (NWF) and Phoneme Segmenting Fluency (PSF) tasks. The NWF task measures students' ability to read letter sounds and to blend letters into words. The examiner presents the student with a sheet of paper with two and three letter nonsense words (e.g., *ov*, *rav*). Two practice items are given with feedback. Students get a point for each letter sound produced correctly, whether pronounced individually (e.g., /o/ /v/) or blended together (e.g., "ov"); raw scores are reported. Students' scores on the DIBELS phoneme segmentation fluency (PSF) were used as a measure of their ability to segment words into phonemes. The PSF task measures students' ability to say the individual phonemes in three and four-phoneme words. The examiner says a word, then asks the student to say the sounds in the word. As with the NWF administration, two practice items are given with feedback before testing begins. For example, the examiner says "sat", and the student says /s/ /a/ /t/. Students get a point for each segment pronounced correctly in 1 min. and raw scores are reported.

### Data Analytic Procedures

To address our first research question regarding the variability in the overall implementation of literacy instruction and individualization of instruction within treatment and comparison classrooms, teacher data were analyzed using Analyses of Variance (ANOVAs) using the SPSS version 17.0 statistical program. Given the nested nature of our data, to address our second question regarding the student outcomes, we conducted a hierarchical multivariate linear model (HMLM; Raudenbush & Bryck, 2002) using the HLM version 6.02 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004).

We chose to use HMLM because it allows us to evaluate multiple outcome measures of one latent construct, reading, thus providing stronger construct validity (Hox, 2002). Rather than conducting a series of univariate tests on each measure where type I error may become inflated, by conducting an HMLM there is more control for type I error and stronger power with a set of multiple measures where the joint effect of those measures may produce a significant effect (Hox, 2002). Furthermore, when conducting research in schools where students are nested in classrooms, and classrooms in schools, HMLM accounts for the nesting nature of the data and account for shared classroom variance so as to avoid mis-estimation of standard errors (Raudenbush & Bryk, 2002) and the covariances can be decomposed over each level.

In HMLM, three levels of data exist. The first level is the measurement level, which includes the latent construct of reading as measured by Letter Sound Fluency, Letter Word Identification, Word Attack, PSF, and NWF. The second level is the child level and includes the covariate of fall letter-sound fluency, which was the only pre-test variable where significant differences existed, in this case, favoring the contrast group. Using this covariate

increases power to find effects. The third level is the classroom level, which included the dummy coded variable for treatment (1 = treatment; 0 = contrast). Because all indicators were not measured in the same metric, the raw score or W score where available, was transformed into a z-score. Because we are evaluating a single latent construct, rather than adding individual interactions for each indicator, by conducting a HMLM, the explanatory variables may be considered the loading or regression weight for each indicator of the latent construct, constraining all indicators to have equal loadings, and a common effect size (Hox, 2002). Limitations of the software precluded including a 4<sup>th</sup> school-level equation. Three level HLM models (students nested in classrooms, nested in schools) revealed no significant school level variance ( $p > .05$ ). Where appropriate, we applied a Benjamini-Hochberg (BH) correction for multiple comparisons at the  $p < .05$  level (Benjamini & Hochberg; 1995).

## Results

### Variability in Teacher Fidelity of Individualization, Fidelity of A2i Usage, and Instructional Effectiveness within ISI-K Treatment and Contrast Classrooms

**Fidelity of individualized instruction and fidelity of A2i usage**—We examined two types of fidelity: the overall individualized instruction, which was targeted in the summer workshop provided to teachers in both conditions and the specific fidelity of A2i, which was only for our treatment condition. Fidelity of individualized instruction ranged from 0 to 3, where 0 indicates no individualization of instruction, 1 indicates small group instruction but all children doing the same activity, 2 indicates small group and individualized activities, and 3 indicates not only small group and individualized activities, but that the overall content of literacy instruction is differentiated and regrouping was observed. Findings suggest that although teachers in the contrast condition attempted to individualize instruction, they were less effective than were the teachers in the ISI-K treatment condition, as seen in Figure 1.

As seen in Figure 1, an ANOVA revealed that teachers in the ISI-K treatment condition provided significantly more individualized instruction than the PD contrast group in both the fall ( $p < .005$ ) and the winter ( $p < .01$ ). The mean fall fidelity of individualized instruction for the treatment group was 1.78 ( $SD = 1.00$ ) and the mean fidelity for the contrast condition was 1.00 ( $SD = .71$ ). ( $M = 1.96$ ,  $SD = 1.11$ ); the mean winter fidelity of individualized instruction for the treatment condition was 1.96 ( $SD = 1.11$ ) and for teachers in the contrast condition the mean was  $M = 1.19$  ( $SD = .75$ ).

With regard to fidelity of the A2i usage, scores ranged from 0–6, with a 0 meaning no usage (thus all contrast teachers scored a 0 because they did not use A2i), and 0–6 being given to treatment teachers depending on their actual use of the software (fidelity rubric available from the first author). The overall mean fidelity of A2i usage among the ISI-K treatment teachers ranged from 2.74 to 3.00 ( $SD = 1.054$ – $1.168$ ), indicating moderate use of A2i and that on average, A2i use increased from January ( $M = 2.87$ ;  $SD = 1.18$ ) to March ( $M = 3.00$ ;  $SD = 1.17$ ) and then dipped slightly in April ( $M = 2.74$ ;  $SD = 1.05$ ).

The correlations among fidelity measures were strong and were also significant at the  $p < .01$  level. Fall and Winter observations of fidelity of individualized instruction were

correlated at  $r = .659$ . Correlations between January, March, and April on fidelity of A2i usage ranged from  $r = .875$  to  $r = .957$ . Intercorrelations between individualization and A2i usage ranged from  $r$ s of  $.471$  to  $.576$ .

**Instructional effectiveness**—Table 3 displays the descriptive statistics and results of a series of one-way ANOVAs on instructional effectiveness; the p-values for all ANOVAs were adjusted for 26 comparisons using the BH procedure at the  $p < .05$  level. Teachers in both conditions provided similar levels of organization of instruction, warmth and sensitivity, and ensured similar levels of on task behaviors among their students in fall and winter. Mean instructional effectiveness ratings ranged from 1.95 (SD = 0.74) to 2.10 (SD = 0.70) suggesting that teachers quality of instruction was rated as relatively effective.

### Kindergarteners' Reading Outcomes and Response to Instruction

Our second research aim was to learn whether students in the treatment classrooms would demonstrate stronger reading outcomes; in addition we wanted to examine the percentage of students who would likely need additional intervention in first grade. Descriptive statistics for each of the outcomes by fall (pre) and spring (post), and by treatment and contrast conditions are provided in Table 4. Students in both conditions improved relative to national norms on word identification and word attack. Students also increased their letter-sound fluency raw scores between the fall and spring in both groups.

To investigate the effect of ISI-K on reading outcomes, we conducted an HMLM analysis. To put the outcomes in the same metric, all spring outcomes (W or raw scores) were z-scored. Table 5 displays the descriptive statistics of the all the variables included in the HMLM analysis in both the W or raw score and z-score metrics. With the exception of the fall Letter Sound Fluency score, which was higher in the contrast group, students did not generally differ by condition on any of the fall measures. Therefore in our final models we controlled only for fall Letter Sound Fluency scores to preserve parsimony in these complex models. Zero-order correlations among outcome and variables and the fall covariate, revealed that our outcome variables ranged from weak to high, with Pearson correlation coefficients ranging from  $.15$  to  $.59$ .

First we ran the unconditional model with homogeneous variance as the best fit. In this model, Tau ( $\pi$ ) was  $.11$  (SE =  $.03$ ) and sigma squared ( $R_0$ ) was  $.44$  (SE =  $.01$ ). This yielded an intraclass correlation (ICC) of  $.20$ . We then added the fall Letter Sound Fluency score at the child level (level 2) and Condition at the classroom level (level 3) where classrooms in the ISI-K condition were coded 1 and the contrast classrooms = 0. In Figure 2 we display the unrestricted and homogeneous hierarchical multivariate linear models that were tested. The model fit statistics revealed that the unrestricted model fit the data significantly better than did the model assuming homogeneous variance hence we report the unrestricted model results with random intercepts and fixed slopes in Table 6. Table 7 displays the variance-covariance matrix of error variances between indicators.

Modeled results revealed a significant effect of the ISI-K treatment; students in ISI-K intervention classrooms achieved significantly greater spring word reading outcomes compared to students in contrast classrooms. Because we used z-scores, a standard deviation

of 1 was used to compute effect size  $d$ , which was .52. This is a moderate effect size (Rosenthal & Rosnow, 1984). Tau decreased from .11 to .07 (i.e., the final model explained 36% of the classroom level variance). Fall Letter Sound Fluency score significantly predicted to students' spring outcomes. Figure 3 shows a graphical representation of the model results.

**Description of classwide responsiveness to Tier 1 instruction**—We also considered the success of ISI-K based upon the percentages of children who would likely need extra intervention in first grade based upon their end of kindergarten skills. Based on prior research (Batsche et al., 2007), up to 20% of students in many classrooms would be expected to need more assistance. Additionally, only 6.9% of students in the treatment condition and 7.6% of students in the contrast condition ended kindergarten with Letter Word Identification standard scores below 85, indicating that their word reading was not normalized.

## Discussion

This cluster-randomized control field trial examined the effect of two forms of professional development on kindergarten teachers' ability to individualize, or differentiate, Tier 1 classroom reading instruction and on reading outcomes of students from economically and culturally diverse backgrounds. Both professional development conditions addressed the need for kindergarten teachers to differentiate core instruction using instructionally meaningful assessment and progress monitoring. Teachers in both the ISI-K treatment and in the professional development contrast conditions received a common baseline of professional development that included (1) a researcher-delivered summer day-long workshop on RTI and individualized instruction, (2) materials and games for center activities, and (3) data on students' reading performance provided through Florida's Progress Monitoring Reporting Network (PMRN), the states' web-accessed data base.

In addition, the treatment teachers received ISI-K training, which included 16 hours of ongoing professional development, guided by ISI-K techniques, A2i software recommendations and planning tools, and bi-weekly in-classroom support from research partners who helped kindergarten teachers differentiate Tier 1 classroom reading instruction based upon students' language and literacy skills. Thus, our test of the ISI-K treatment represents a rigorous test of all three components of the ISI regimen used in prior investigations of ISI efficacy A2i software, ongoing professional development, and bi-weekly classroom-based support.

We found that teachers in the ISI treatment condition provided similarly effective classroom and reading instruction compared to teachers in the contrast condition, but they delivered significantly more individualized small group instruction. Encouragingly, although students in both conditions improved their reading skills relative to national norms, students in treatment classrooms outperformed students in the contrast classrooms on a latent measure of conventional literacy skills, comprised of letter word reading, decoding, alphabetic knowledge, and phonological awareness ( $ES = .52$ ). Our findings add to the growing body of evidence that teachers who use ISI to individualize instruction based upon students'



language and literacy skills in first through third grade achieve stronger student reading performance than teachers who do not (Connor et al., 2004; Connor, Morrison, et al., 2009; Connor, Piasta et al., 2009). ISI is still fairly new so it was important to demonstrate that, consistent with the notion of early intervening, ISI was as effective in kindergarten, when formal reading instruction begins.

There are some potentially important lessons to be gleaned from the present study about changing teacher behavior, and through that change, improving reading instruction and subsequent student outcomes. First, it was challenging, but possible, for teachers to differentiate their core reading instruction when they received ISI-K. Teachers in both conditions provided small group instruction, but teachers in the contrast condition did not differentiate activities or materials provided to groups, despite having received: the workshop on RTI and individualized instruction, games for center activities, and data on their students' reading performance provided through the state database. Thus the contrast teachers' tendency to provide the same activities to all children, regardless of their skill level is consistent with prior research (Fuchs et al., 1994). By contrast, treatment teachers, using ISI-K, provided significantly more effective individualized instruction than contrast teachers, and generally provided each group different activities tailored to their ability levels. Thus, an important implication of the present study is that the combination of A2i, ongoing professional development and bi-weekly support that comprised the ISI-K treatment were important and active ingredients that supported kindergarten teachers in differentiating instruction based upon students' ongoing assessments of language and literacy skills, which is consistent with prior investigations of ISI in first through third grades (Connor et al., 2004; Connor, Morrison, et al., 2009; Connor, Piasta et al., 2009).

Second, not all of our treatment teachers made full use of the A2i web-based resources. The overall mean fidelity of A2i usage indicates that on average, the treatment teachers were using the software only to a moderate level, attending to the recommended groupings, but they had not yet mastered the ability to implement grouping that consistently followed recommended amounts and types of instruction for all individual students. This finding confirms prior concerns related to ensuring teachers provide interventions that are sufficiently implemented with fidelity (e.g., Ikeda et al., 2007). Connor and colleagues have reported in prior studies of ISI that the amount of time teachers viewed A2i resources (assessment graphs, recommended groupings) and the more precisely they followed the recommendations for amount and types of literacy instruction (e.g., more phonetic decoding for students with weaker skills, especially in the beginning of the school year and more comprehension instruction for students with stronger skills with increasingly greater amounts of child-managed instruction), the stronger the reading growth of all students (cite).

Third, we confirmed that individualizing led to stronger student conventional literacy outcomes at the end of kindergarten within a diverse group of students. No children were excluded from the study and so the sample included many students with disabilities. More than half of students were from minority backgrounds and received free and reduced price lunch. Confidence in our findings is strengthened by our research design that involved successful randomization (groups were similar on IQ, vocabulary, ethnicity, gender,

eligibility for free and reduced lunch, parental education, preschool experiences, home literacy exposure, and all but one measure of conventional literacy at the start of the school year). Further, we used a sophisticated data analysis to account for students being nested within classrooms, which accounted for students' weaker initial letter sound scores within the treatment classrooms, and we analyzed multiple child measures of literacy, careful assessments of classroom instruction, and fidelity of treatment within both conditions. Differences in instructional implementation, rather than individualization could have led to differences in student outcomes. However, we learned, as shown in Table 3, that both groups of teachers were similarly effective in their organization of instruction, warmth and sensitivity, and ability to ensure most students were on-task. Additionally, teachers in both conditions provided similarly well-implemented reading instruction. This finding may be partly explained by the common explicit and systematic core reading program used in most classrooms. There was, however, considerable variability in quality ratings for some reading instruction components (most standard deviations were half as large as the means; some were even greater). The correlations between teachers' fall and winter ratings ranged from strong (on-task  $r = .62$ ) to weak (decoding  $r = .05$ ), indicating generally higher correlations among the overall instructional quality than among the reading instructional effectiveness ratings. These findings provide support that differences in student outcomes were not due to overall instructional differences between the two treatment conditions, but to individualization and that also suggest that when teachers are providing effective instruction, some individualization was better than none.

Finally, we learned that the overall percentage of students who did not reach grade level reading expectations was notably small (about seven percent, which is far less than the 20 percent rate considered acceptable for Tier 1 by Batsche et al., 2007 and more in line with the six percent failure rate deemed acceptable for RTI systems by Torgesen, 2002). This success rate may be attributed to professional development that aimed to help teachers differentiate their core reading instruction, essentially creating within their classrooms a hybrid of Tier 1 (classroom instruction) and Tier 2 (targeted and differentiated small group interventions). This finding is very encouraging, given that our sample was considerably at risk for academic underachievement - more than half were minority and received free and reduced price lunch and nearly 20 percent had a speech or language delay or other disability. Furthermore, there did not appear to be a trend of weaker outcomes that related to the percentage of children within a school receiving free and reduced price lunch. By the end of kindergarten, the mean sample reading scores were 107, which is more than a standard deviation higher than their IQ or Vocabulary standard scores. Findings may not generalize to other curriculum programs that have less explicit reading instruction, which is important in light of at least one prior larger-scale investigation that examined the efficacy of reading kindergarten curricula (Foorman et al., 2003). Foorman and colleagues conducted a multi-year, multi-cohort study with over 4,800 kindergartners who attended struggling schools. Teachers were provided even more extensive professional development (that included graduate coursework) than we did in the present study. Foorman and colleagues examined the extent to which kindergarten reading curricula supported the teaching of connections between phonemic awareness and letter sound knowledge explicitly and systematically. Similar to our study findings, Foorman et al. (2003) reported that curricula that explicitly

linked phonemic awareness and the alphabetic principle in kindergarten led to reading performance that was at the national average, despite considerable variability in teachers' instructional quality and fidelity of implementation.

**Limitations and directions for future research**—As is true with any research, and particularly with school-based research, there are several important limitations to the present study, but these too are lessons that will inform future research on RTI implementation. First, due to the complexity of this large project and scheduling challenges related to administering such a large number of assessments, our RAs who tested children and who videotaped observations were not blind to condition. A stronger design would include a greater number of schools but resources precluded this. Only 14 schools left the study somewhat underpowered although we were able to improve power by adding covariates and using HMLM analyses. Additionally a stronger design would incorporate a separate team of assessors. Ideally as we learn which assessments are most important, the number of measures can be trimmed and perhaps if teachers conduct some of the assessments, it would help them be even more aware of student progress. Furthermore, research partners were in treatment classrooms regularly to provide in-class support and so we cautioned them against examiner bias and about familiarizing themselves with children prior to testing.

Second, teachers were told in advance that we would be videotaping classroom instruction. It is possible that had we visited more often or unannounced, we would have observed different behavior. That said, we did not observe perfect implementation in any case. Furthermore, the small group center activities and materials that we observed were clearly familiar to children and so it is likely that what we videotaped was consistent with typical practice.

Third, in the future, we need to unpack ISI-K to learn how much of which of the three active ingredients (ongoing professional development, in-class support, and A2i) is enough. In other words, how much support is enough and for who? Can we individualize teacher support to improve individualization and fidelity to ISI-K and particularly to the A2i software? We are also in the process of analyzing data from the second year of the study that will allow us to learn whether a second year of treatment will have an additional impact on instructional practices. Instructional practice data can also be analyzed from a within-teacher perspective. Will teachers improve their individualization and implementation and if so, will subsequent cohorts of students achieve greater growth?

Fourth, we are conducting longitudinal follow up for our students to learn more about the longer-term impact of effectively individualized instruction and about identification of students with persistent reading difficulties as reading disabled. We can not assume that all students who achieve a standard score of 85 on a measure of word reading will learn to reading fluently and with comprehension by the end of the primary grades. Even though the majority of parents reported reading to their child daily, and most reported having done so for several years prior to kindergarten, especially with our high-risk students, there will likely be some summer recidivism. Thus we hope to learn more about how many children will need extra intervention in first grade and beyond; this will be possible because the state has just mandated RTI. From this longitudinal work we also will explore whether an 85

standard score is a good response criteria and will compare this to other benchmarks on progress monitoring measures that could be more teacher-friendly.

Finally, future research is needed to replicate the efficacy of ISI-K as a means of ensuring instruction is matched to students' developing skills and to test its efficacy as an RTI school-wide system. We believe that ISI could be expanded beyond the classroom for use by principals and reading coaches as a dynamic forecasting intervention school-wide model. It is clear that schools need assistance to help teachers use assessment data to individualize beginning reading instruction and also to plan which children need additional interventions (Connor, Morrison, et al., 2007; Gersten et al., 2008; Mathes et al., 2005; Scanlon et al., 2008; Vellutino et al., 1996). Some of the students with disabilities in the present study were recommended to receive over 90 minutes per day of small group individualized teacher managed code-focused instruction. This intensity is likely more heroic than can be provided within general education. Data from the present study and from previous investigations in kindergarten (Al Otaiba et al., 2008) and in later grades has demonstrated that students' response to instruction depends in part on how their initial language and literacy skills interact with the instruction they receive (Connor et al., 2004; Connor, Morrison, et al., 2007; Connor, Piasta et al., 2009). This accumulating data can be used to derive equations that compute recommended amounts and types of instruction based on a set target outcome (grade level comprehension or a year's worth of growth) and individual students' assessed skills. Because RTI would be data-informed, teachers will provide the instruction, students will be reassessed and the recommended instruction modified so that children continue to show progress in key component skills across the primary grades. In conclusion, lessons learned from this empirical study uniquely contribute to the professional dialogue about the value of professional development related to differentiating core reading instruction and the challenges of using RTI approaches to address students' needs in the areas of reading in general education contexts.

## Acknowledgments

This work was supported by (a) a Multidisciplinary Learning Disabilities Center Grant P50HD052120 from the National Institute of Child Health and Human Development, (b) a Predoctoral Interdisciplinary Research Training Grant R305B04074 from the Institute for Education Science, (c) grants R305H04013 and R305B070074, "Child by Instruction Interactions: Effects of Individualizing Instruction" from the US. Department of Education, Institute for Education Sciences and (d) by grant R01HD48539 from the National Institute for Child Health and Human Development. We appreciate the teachers and students who have participated in this project and our numerous project staff for their help on the study. We are grateful for the helpful feedback from our anonymous reviewers.

## References

- Al Otaiba S, Connor C, Lane H, Kosanovich M, Schatschneider C, Dyrland AK, Miller MS, Wright TL. Reading First kindergarten classroom instruction and students' early literacy growth. *Journal of School Psychology*. 2008; 46:281–314. [PubMed: 19083361]
- Batsche, GM.; Curtis, MJ.; Dorman, C.; Castillo, JM.; Porter, LJ. The Florida problem-solving/response to intervention model: Implementing a statewide initiative. In: Jimerson, SR.; Burns, MK.; VanDerHeyden, AM., editors. *Handbook of response to intervention: The science and practice of assessment and intervention*. New York, NY: Springer Science + Business Media; 2007. p. 378-395.
- Beck, IL.; McKeown, MG.; Kucan, L. *Bringing words to life: Robust vocabulary instruction*. New York: The Guilford Press; 2002.

- Benjamini Y, Hochberg Y. Controlling the false discovery rate - a new and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*. 1995; 57:289–300.
- Bereiter, C.; Brown, A.; Campione, J.; Carruthers, I.; Case, R.; Hirshberg, J.; Treadway, GH. *Open court reading*. Columbus, OH: SRA McGraw-Hill; 2002.
- Connor CM, Morrison FJ, Katch LE. Beyond the reading wars: Exploring the effect of child-instruction interactions on growth in early reading. *Scientific Studies of Reading*. 2004; 8:305–336.
- Connor CM, Morrison FJ, Slominski L. Preschool instruction and children’s emergent literacy growth. *Journal of Educational Psychology*. 2006; 98:665–689.10.1037/0022-0663.98.4.665
- Connor CM, Morrison FJ, Underwood PS. A second chance in second grade: The independent and cumulative impact of first and second-grade reading instruction and students’ letter-word reading skill growth. *Scientific Studies of Reading*. 2007; 11:199–233.
- Connor CM, Morrison FJ, Fishman BJ, Ponitz CC, Glasney S, Underwood P, et al. The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher*. 2009; 38:85–99.
- Connor CM, Morrison FJ, Fishman BJ, Schatschneider C, Underwood P. The Early Years: Algorithm-guided individualized reading instruction. *Science*. 2007; 315:464–465. [PubMed: 17255498]
- Connor CM, Piasta SB, Fishman B, Glasney S, Schatschneider C, Crowe E, Underwood P, Morrison FJ. Individualizing student instruction precisely: Effects of child × instruction interactions on first graders’ literacy development. *Child Development*. 2009; 80:77–100.10.1111/j.1467-8624.2008.01247.x [PubMed: 19236394]
- Cronbach, L.J.; Snow, RE. *Aptitudes and instructional methods: A handbook for Research on Interactions*. New York: Irvington; 1977.
- Engelmann, S.; Bruner, EC. *Reading mastery plus*. Columbus, OH: SRA McGraw-Hill; 2002.
- Florida Center for Reading Research. 2007 Student Center Activities: Grades K-1. [Instructional routines]. Retrieved from <http://www.fcrr.org/curriculum/SCAindex.shtm>
- Folsom, JS.; Al Otaiba, S.; Greulich, L.; Meadows, J. Examining the relations between kindergarten vocabulary, instructional practices and quality: A longitudinal study. Poster presented at the Sixteenth Annual Meeting of the Society for the Scientific Studies of Reading Conference; Boston, MA. 2009 Jul.
- Foorman BR, Chen D, Carlson C, Moats L, Francis DJ, Fletcher JM. The necessity of the alphabetic principle to phonemic awareness instruction. *Reading and Writing: An Interdisciplinary Journal*. 2003; 16:289–324.
- Foorman BR, Francis DJ, Fletcher JM, Schatschneider C, Mehta P. The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology*. 1998; 90:235–235.10.1037/0022-0663.90.2.235
- Fuchs D, Fuchs LS. Test procedure bias: A meta-analysis of examiner familiarity effects. *Review of Educational Research*. 1986; 56:243–262.
- Fuchs D, Fuchs LS, Power MH, Dailey AM. Bias in the assessment of handicapped children. *American Educational Research Journal*. 1985; 22:185–198.10.2307/1162839
- Fuchs LS, Fuchs D, Hamlett CL, Phillips NB, Bentz J. Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children*. 1994; 60:516–537.
- Fuchs D, Deschler D. What we need to know about responsiveness to intervention (and shouldn’t be afraid to ask). *Learning Disabilities Research and Practice*. 2007; 22:129–136.
- Gersten, R.; Compton, D.; Connor, CM.; Dimino, J.; Santoro, L.; Linan-Thompson, S.; Tilly, WD. *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; 2008. Retrieved from <http://ies.ed.gov/ncee/wwc/publications/practiceguides/>
- Good, RH.; Kaminski, RA. *Dynamic Indicators of Basic Early Literacy Skills*. 6. Eugene, OR: Institute for Development of Educational Achievement; 2002.
- Gough PB, Tunmer WE. Decoding, reading, and reading disability. *RASE: Remedial & Special Education*. 1986; 7:6–10.10.1177/074193258600700104

- Hox, J. *Multilevel Analysis*. Mahwah, NJ: Lawrence Erlbaum; 2002.
- Ikeda, MJ.; Rahn-Blakeslee, A.; Niebling, BC.; Gustafson, JK.; Allison, R.; Stumme, J. The Heartland Area Education Agency 11 problem-solving approach: An overview and lessons learned. In: Jimerson, SR.; Burns, MK.; VanDerHeyden, AM., editors. *Handbook of response to intervention: The science and practice of assessment and intervention*. New York, NY: Springer Science + Business Media; 2007. p. 255-268.
- Juel C, Minden-Cupp C. Learning to read words: Linguistic units and instructional strategies. *Reading Research Quarterly*. 2000; 35:458–492.10.1598/RRQ.35.4.2
- Kaufman, A.; Kaufman, N. *Kaufman Brief Intelligence Test*. 2. Circle Pines, MN: American Guidance Service Publishing; 2004.
- Klingner JK, Vaughn S, Hughes MT, Schumm JS, Elbaum B. Outcomes for students with and without learning disabilities in inclusive classrooms. *Learning Disabilities Research & Practice*. 1998; 13:153–161.
- Marston D, Pickart M, Reschly A, Heistad D, Muyskens P, Tindal G. Early literacy measures for improving student reading achievement: Translating research into practice. *Exceptionality*. 2007; 15:97–117.10.1080/09362830701294177
- Mathes PG, Denton C, Fletcher J, Anthony J, Francis D, Schatschneider C. The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly*. 2005; 40:148–182.
- Morgan PL, Fuchs D. Is there a bidirectional relationship between children’s reading skills and reading motivation? *Exceptional Children*. 2007; 73:165–183.
- National Assessment of Educational Progress, National Center for Education Statistics. *The nation’s report card: Reading 2007*. Washington D.C: National Center for Education Statistics; 2007.
- National Early Literacy Panel. *Developing early literacy: Report of the National Early Literacy Panel*. Washington, DC: National Center for Family Literacy; 2009.
- National Reading Panel. *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (National Institute of Health Report No. 00-4769)*. Washington, DC: National Institute of Child Health and Human Development; 2000.
- Raudenbush, SW.; Bryk, AS. *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage; 2002.
- Raudenbush, SW.; Bryk, AS.; Cheong, YF.; Congdon, R.; du Toit, M. *HLM6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International; 2004.
- Rosenthal, R.; Rosnow, RL. *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill; 1984.
- Sattler, JM. *Assessment of Children’s Intelligence and Special Abilities*. Boston, MA: Allyn & Bacon; 1982.
- Scanlon DM, Gelzheiser LM, Vellutino FR, Schatschneider C, Sweeney JM. Reducing the incidence of early reading difficulties: Professional Development for classroom teachers versus direct interventions for children. *Learning and Individual Differences*. 2008; 18:346–359. [PubMed: 19325710]
- Shinn, MM.; Shinn, MR. *AIMSweb*. Eden Prairie, MN: Edformation; 2004.
- Showers B, Joyce B, Bennett B. Synthesis of research on staff development: A framework for future study and a state-of-the-art analysis. *Educational Leadership*. 1987; 45:77–87.
- Speece D, Case L. Classification in context: An alternative approach to identifying early reading disability. *Journal of Educational Psychology*. 2001; 93:735–749.
- Torgesen JK. The prevention of reading difficulties. *Journal of School Psychology*. 2002; 40:7–26.
- Torgesen JK, Wagner RK, Rashotte CA, Rose E, Lindamood P, Conway T, Garvan C. Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of*. 1999
- Vaughn S, Fuchs LS. Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice*. 2003; 18:137–146.10.1111/1540-5826.00070

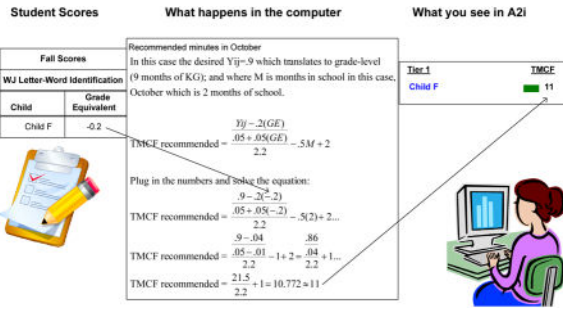
- Vellutino FR, Scanlon DM, Sipay ER, Small SG, Pratt A, Chen R, Denckla MB. Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*. 1996; 88:601–638.
- Wanzek J, Vaughn S. Research-bases implications from extensive early reading interventions. *School Psychology Review*. 2007; 36:541–561.
- Whitehurst GJ, Lonigan CJ. Child development and emergent literacy. *Child Development*. 1998; 69:848–872.10.2307/1132208 [PubMed: 9680688]
- Woodcock, RW.; McGrew, KS.; Mather, N. Woodcock-Johnson III tests of achievement. Itasca, IL: Riverside Publishing; 2001.

## Appendix A

### How does the computer know what my students need?

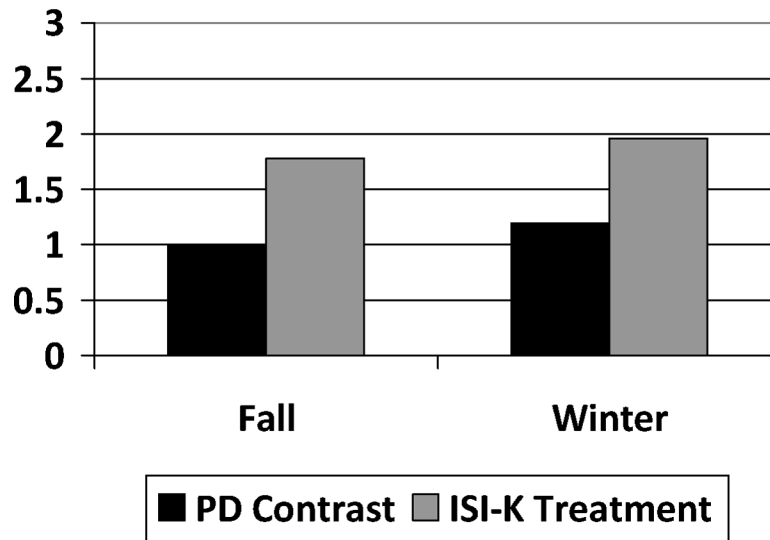


Based on previous research we found that we could predict where students ended the school year by his/her reading scores, the month of school in which they were in, and the types and amounts of reading instruction they received. A basic algebraic equation was created based on these results and programmed into A2i which now uses the the students test scores, month in school, and desired end of year outcome, to calculate the amounts and types of instruction needed. It's pretty simple! Let's walk through an example of a student that starts the year 2 months behind a normal kindergarten student:





## Extent of Individualized Instruction



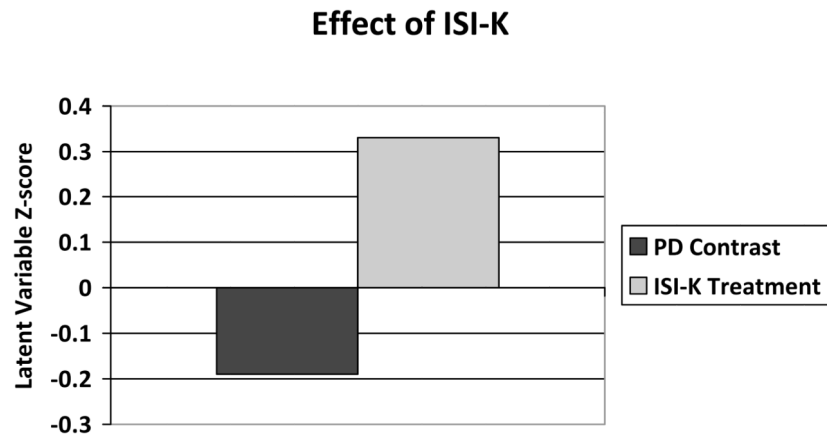
**Figure 1.**

Extent to which teachers individualized instruction. Teachers in the ISI-K treatment group provided significantly more individualized instruction than the PD contrast group in both the fall ( $p < .005$ ) and in the winter ( $p < .01$ ). In this figure 1 represents the use of small groups however students are receiving the same instruction and 2 represents clear evidence of differentiation where small groups and content vary by student and skill level.

UNRESTRICTED MODEL  
 LEVEL 1 MODEL  
 $READING = (LSF\_S)*READING' + (LW)*READING' + (NWF)*READING' + (PSF)*READING' + (WA)*READING'$   
 $READING' = \pi_0 + \varepsilon$   
 LEVEL 2 MODEL  
 $\pi_0 = \beta_{00} + \beta_{01}(LSF\_F)$   
 Combined level-1 and level-2 Model  
 $READING = \beta_{00} + \beta_{01}(LSF\_F) + \varepsilon$   
 $Var(\varepsilon) = \Delta$   
 LEVEL 3 MODEL  
 $\beta_{00} = \gamma_{000} + \gamma_{001}(TREATMENT) + u_{00}$   
 $\beta_{01} = \gamma_{010}$

HOMOGENOUS MODEL  
 LEVEL 1 MODEL  
 $READING = (LSF\_S)*READING' + (LW)*READING' + (NWF)*READING' + (PSF)*READING' + (WA)*READING'$   
 $READING' = \pi_0 + \varepsilon$   
 LEVEL 2 MODEL  
 $\pi_0 = \beta_{00} + \beta_{01}(LSF\_F) + \tau_0$   
 Combined level-1 and level-2 Model  
 $READING = \beta_{00} + \beta_{01}(LSF\_F) + \varepsilon$   
 $\varepsilon = \tau_0 + e$   
 $Var(\varepsilon) = Var(Ar+e) = \Delta = A\tau_e A' + \sigma^2 I$   
 LEVEL 3 MODEL  
 $\beta_{00} = \gamma_{000} + \gamma_{001}(TREATMENT) + u_{00}$   
 $\beta_{01} = \gamma_{010}$

**Figure 2.**  
 The unrestricted and homogenous hierarchical multivariate linear models.



**Figure 3.** The effect of ISI-K treatment compared to PD contrast after controlling for fall letter-sound fluency.

**Table 1**

Student demographic Data

Variables	Treatment Group (n=305)			Contrast Group (n=251)		
	M	SD	P	M	SD	P
Age at fall testing	5.21	0.37	.534 <sup>†</sup>	5.19	0.35	.534 <sup>†</sup>
Nonverbal IQ <i>K</i>	91.80	12.44	.087 <sup>†</sup>	93.50	10.47	.087 <sup>†</sup>
Verbal IQ <i>K</i>	92.14	15.16	.341 <sup>†</sup>	93.33	14.00	.341 <sup>†</sup>
Picture Vocabulary <i>WJ</i>	99.65	10.80	.817 <sup>†</sup>	99.86	10.94	.817 <sup>†</sup>
Absence*	7.04	9.41	.000 <sup>†**</sup>	3.29	5.34	.000 <sup>†**</sup>
	<i>n</i>	%		<i>n</i>	%	<i>P</i>
Male	164	53.77	.351 <sup>†</sup>	125	49.80	.351 <sup>†</sup>
Ethnicity			.924 <sup>†</sup>			.924 <sup>†</sup>
Caucasian	105	34.43		83	33.07	
Black	175	57.38		151	60.16	
Hispanic	14	4.59		10	3.98	
Other	11	3.61		7	2.79	
Eligible for Free or Reduced Lunch	182	59.67	.248 <sup>†</sup>	137	54.58	.248 <sup>†</sup>
Retained	35	11.48	.543 <sup>†</sup>	25	9.96	.543 <sup>†</sup>
Special Education Classification						
Speech/Language	55	18.03	.202 <sup>†</sup>	21	8.37	.202 <sup>†</sup>
EMH, DD, ASD	4	1.31	.394 <sup>†</sup>	8	3.19	.394 <sup>†</sup>
Sensory Impaired	4	1.31	.890 <sup>†</sup>	3	1.20	.890 <sup>†</sup>
Orthopedically or OHI	8	2.62	.037 <sup>†</sup>	1	0.40	.037 <sup>†</sup>
Learning Disabled	1	0.33	.356 <sup>†</sup>	2	0.80	.356 <sup>†</sup>
Gifted	1	0.33	.289 <sup>†</sup>	3	1.20	.289 <sup>†</sup>

Note.

*K* = Kaufman Brief Intelligence Test (Kaufman & Kaufman, 2004);

*WJ* = Woodcock Johnson III (Woodcock, McGrew, & Mather, 2001)

EMH = Educably Mentally Handicapped; DD = Developmentally Delayed; ASD = Autism Spectrum Disorders; OHI = Other Health Impaired

\* n=284 for treatment; n=242 for contrast.

<sup>†</sup> p-value is computed using ANOVA.

<sup>‡</sup> p-value is computed using Chi-square test.

\* = significant at the  $p = .05$  level after the Benjamini Hochman correction was applied.

Table 2

Parent Survey Data

Variables	Treatment (n=305)		Contrast (n=251)		p
	n	%	n	%	
Parent Education					.185 <sup>‡</sup>
Some high school	28	9.18	21	8.37	
High school diploma	44	14.43	30	11.95	
Some college/vocational training	90	29.51	86	34.26	
College degree	90	29.51	58	23.11	
Graduate degree	24	7.87	30	11.95	
No response	29	9.51	26	10.36	
How many years of reading to child					.816 <sup>‡</sup>
< 1 year	4	1.31	3	1.20	
1 – 2 years	71	23.28	66	26.29	
2 – 3 years	79	25.90	67	26.69	
3 – 4 years	42	13.77	27	10.76	
4 – 5 years	25	8.20	16	6.37	
> 5 years	13	4.26	12	4.78	
Not Reported	71	23.28	60	23.90	
Time spent on reading per day					.328 <sup>‡</sup>
None	15	4.92	10	3.98	
About 10 minutes	98	32.13	72	28.69	
Between 15–30 minutes	147	48.20	133	52.99	
Over 30 minutes	22	7.21	9	3.59	
Not Reported	23	7.54	27	10.76	
Attended Preschool	241	79.02	201	80.08	.126 <sup>‡</sup>

Note. P-values are not adjusted for Type-I error because all values are already non-significant.

<sup>‡</sup> p-value is computed using Chi-square test.

Table 3

Instructional Effectiveness	Contrast ( <i>n</i> = 21)		Treatment ( <i>n</i> = 23)		P
	M	SD	M	SD	
<b>Fall Quality of Instruction</b>					
Organization of instruction	2.10	0.70	1.91	0.95	.476
Warmth & Sensitivity	1.95	0.74	1.83	0.83	.599
Level of classroom on-task behaviors	2.10	0.70	1.96	0.56	.471
<b>Fall Instructional Effectiveness</b>					
Phonological Awareness	2.10	0.89	2.09	1.08	.978
Alphabets	1.76	0.83	2.09	0.73	.175
Decoding	1.62	0.80	2.09	0.85	.068
Vocabulary	1.33	1.06	1.57	1.16	.495
Comprehension	1.62	1.07	1.83	1.15	.542
Fluency	0.19	0.60	0.26	0.75	.735
Spelling	0.76	1.09	1.48	1.12	.038
Writing	1.10	1.00	2.22	0.90	.000*
Child Managed Activities	1.76	1.18	2.30	1.02	.109
Other Adults	0.95	0.92	1.26	1.21	.351
<b>Winter Quality of Instruction</b>					
Organization of instruction	2.29	0.72	2.04	0.88	.325
Warmth & Sensitivity	2.19	0.81	2.26	0.75	.767
Level of classroom on-task behaviors	2.33	0.66	2.13	0.63	.301
<b>Winter Instructional Effectiveness</b>					
Phonological Awareness	2.62	0.74	2.35	0.83	.261
Alphabets	1.90	0.94	1.74	1.21	.618
Decoding	2.00	0.95	2.17	0.83	.521
Vocabulary	1.29	1.15	1.48	1.20	.590
Comprehension	2.19	1.17	1.87	1.10	.353
Fluency	0.52	0.81	0.43	0.84	.724
Spelling	0.19	0.51	1.00	1.17	.005*

	Contrast ( $n = 21$ )		Treatment ( $n = 23$ )		P
	M	SD	M	SD	
Writing	0.86	0.57	1.70	0.88	.001*
Child Managed Activities	2.48	0.93	2.26	1.01	.467
Other Adults	1.38	1.24	1.96	1.02	.100

Note. All items were on a 0–3 scale, where 0 indicated not observed, 1 indicated not effective, 2 indicated effective, and 3 indicated highly effective.

\* = significant at the  $p = .05$  level after the Benjamini Hochman correction was applied.



Table 4

Pre/Post Measures by Treatment and Contrast

	Treatment						Contrast					
	Pre			Post			Pre			Post		
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Letter Word <i>S WJ</i>	95.53	12.23	103.81	13.88	97.27	13.52	104.84	14.77				
Word Attack <i>S WJ</i>	96.37	22.14	107.76	13.82	98.86	21.94	107.85	13.95				
Letter Sound Fluency <sup>A</sup>	8.15	9.61	37.16	17.42	9.98	10.26	34.02	14.26				
Phoneme Segmenting Fluency <sup>D</sup>			42.2	22.97			30.07	15.61				
Nonsense Word Fluency <sup>D</sup>			39.25	24.66			36.74	23.04				

Note.

*S* = Standard Score,*WJ* = Woodcock-Johnson III (Woodcock, McGrew, Mather, 2001),*A* = AIMSWeb (Shinn & Shinn, 2004),*D* = DIBELS (Good & Kaminski, 2002), *SS* = standard score, *W* = *W* score

Table 5

Variables in HMLM analysis

	Treatment			Contrast		
	n	M	SD	n	M	SD
<b>Covariate</b>						
Letter Sound Fluency <sup>A,F</sup>	305	8.15	9.61	249	9.98	10.63
Letter Sound Fluency <sup>A,S</sup>	305	37.16	17.42	251	34.02	14.26
Letter Sound Fluency <sup>A</sup> z-score	305	0.09	1.08	251	-0.11	0.88
Letter Word <sup>WJ</sup> W	305	396.90	28.40	251	397.69	29.78
Letter Word <sup>WJ</sup> z-score	305	-0.01	0.98	251	0.01	1.03
Word Attack <sup>WJ</sup> W	305	447.04	26.86	251	445.65	27.28
Word Attack <sup>WJ</sup> z-score	305	0.02	0.99	251	-0.03	1.01
Phoneme Segmenting Fluency <sup>D</sup>	303	42.20	22.97	245	30.07	15.61
Phoneme Segmenting Fluency <sup>D</sup> z-score	303	0.26	1.10	245	-0.32	0.75
Nonsense Word Fluency <sup>D</sup>	303	39.25	24.66	245	36.74	23.04
Nonsense Word Fluency <sup>D</sup> z-score	303	0.05	1.03	245	-0.06	0.96

Note. Z-scores of all measures were used in the HMLM analysis, for Letter Word and Word Attack, z-scores were calculated from the w scores. The fall LSF raw score was used as a contrast variable. ANOVAs were conducted only on the data included in the actual model, thus p-values are based on the z-scores of the measures in the latent variable and on the raw score of the covariate.

<sup>A</sup> = AIMSWeb (Shinn & Shinn, 2004),

<sup>F</sup> = Fall assessment data,

<sup>S</sup> = Spring assessment data,

<sup>WJ</sup> = Woodcock Johnson III (Woodcock, McGrew, Mather, 2001),

<sup>D</sup> = DIBELS (Good & Kaminski, 2002), SS = standard score, W = W score

\* = significant at the  $p = .05$  level after the Benjamini Hochman correction was applied.

**Table 6**

Summary of Hierarchical Multilevel Linear Model Results

Variable	Coefficient	SE	T-ratio	df	p
<b>Fixed Effects</b>					
Intercept	G <sub>000</sub>	-0.19	0.07	-2.68	42 0.011
Fall LSF-A	G <sub>010</sub>	0.04	0.00	15.64	544 <.001
Condition	G <sub>001</sub>	0.33	0.10	3.33	42 0.002
<b>Random Effects</b>					
School	Tau <sub>B</sub>	0.07	0.02		

Note. LSF = Letter Sound Fluency (AIMSweb). This model was estimated with random intercepts and fixed slopes model. Fall LSF was used as a covariate to control for initial status. Condition was coded 0 and 1 where 1 indicated that the student was in a treatment classroom.

**Table 7**

Variance-Covariance Matrix of Error Variance Between Indicators

	1.	2.	3.	4.	5.
1. Letter Sound Fluency <sup>A</sup>	0.71	0.29	0.32	0.18	0.36
2. Letter Word <sup>WJ</sup>	0.29	0.64	0.34	0.05	0.42
3. Nonsense Word Fluency <sup>D</sup>	0.32	0.34	0.71	0.29	0.29
4. Phoneme Segmenting Fluency <sup>D</sup>	0.18	0.05	0.29	0.84	0.18
5. Word Attack <sup>WJ</sup>	0.36	0.42	0.29	0.18	0.67

Variance-Covariance Matrix of Standard Errors Variance Between Indicators					
1. Letter Sound Fluency <sup>A</sup>	0.04	0.03	0.03	0.03	0.04
2. Letter Word <sup>WJ</sup>	0.03	0.03	0.03	0.03	0.03
3. Nonsense Word Fluency <sup>D</sup>	0.03	0.03	0.04	0.03	0.05
4. Phoneme Segmenting Fluency <sup>D</sup>	0.03	0.03	0.03	0.04	0.03
5. Word Attack <sup>WJ</sup>	0.04	0.03	0.05	0.03	0.04

<sup>A</sup> = AIMSWeb (Shinn & Shinn, 2004),

<sup>WJ</sup> = Woodcock Johnson III (Woodcock, McGrew, Mather, 2001),

<sup>D</sup> = DIBELS (Good & Kaminski, 2002)