HHS Public Access

Author manuscript

Read Writ Q. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

Read Writ Q. 2016 ; 32(5): 454–476. doi:10.1080/10573569.2015.1021060.

Professional development to differentiate kindergarten Tier 1 instruction: Can already effective teachers improve student outcomes by differentiating Tier 1 instruction?

Stephanie Al Otaiba,
Southern Methodist University, Dallas, Texas, USA

Jessica S. Folsom,
Florida State University, Tallahassee, Florida, USA

Jeannie Wanzek,
Florida State University, Tallahassee, Florida, USA

Luana Greulich,
Andrews University, Berrien Springs, Michigan, USA

Jessica Wasche,
University of Central Florida, Orlando, Florida, USA

Christopher Schatschneider, and
Florida State University, Tallahassee, Florida, USA

Carol Connor
Arizona State University, Phoenix, Arizona, USA

Abstract

Two primary purposes guided this quasi-experimental within-teacher study: (1) to examine changes from baseline through two years of professional development (Individualizing Student Instruction) in kindergarten teachers' differentiation of Tier 1 literacy instruction; (2) to examine changes in reading and vocabulary of three cohorts of the teachers' students (n = 416). Teachers' instruction was observed and students were assessed on standardized measures of vocabulary and word reading. Results suggested that teachers significantly increased their differentiation and students showed significantly greater word reading outcomes relative to baseline. No change was observed for vocabulary. Results have implications for supporting teacher effectiveness through technology-supported professional development.

More students in the U.S. are reaching basic and proficient levels of reading achievement in the elementary and middle grades than at any other time in history (National Center for Education Statistics, 2011). However, there are still too many students who fail to achieve basic reading levels by fourth grade, underscoring the importance of ensuring effective early

Address correspondence to Stephanie Al Otaiba, Department of Teaching and Learning, Annette Caldwell Simmons School of Education and Human Development, Southern Methodist University, PO Box 750455, Dallas, TX 75275-0455, USA. salotaiba@smu.edu.

reading instruction. As a result, across all 50 states, schools and districts have begun to organize elementary reading instruction around multi-tiered response to intervention (RTI) models for guiding instruction, intervention, and early identification of students with reading difficulties (O'Connor, Fulmer, Harty, & Bell, 2005; Vaughn, Woodruff, Murray, Wanzek, Scammacca et al., 2008; Wanzek & Vaughn, 2007; Zirkel & Thomas, 2010). Essential to the implementation of RTI is an effective core reading program (Tier I) based on scientific reading research to ensure all students receive quality instruction (Crowe, Connor, Petscher, 2009; Fletcher, Denton, Fuchs, & Vaughn, 2005; Vaughn, Wanzek, Woodruff, & Linan-Thompson, 2007; Gersten et al., 2008).

Tier I is the classroom reading instruction that all students receive and typically consists of (a) screening assessments administered several times per year to monitor student reading achievement and identify students at-risk for reading failure, (b) research-based differentiated instruction that focuses on the essential reading components for the grade level, both code-focused and meaning-focused elements, and (c) ongoing professional development for teachers. This later emphasis on professional development to ensure that teachers know how to differentiate instruction is focal to the present study because well-implemented Tier 1 is foundational to an RTI model, yet researchers have noted that many classroom teachers have limited knowledge about how to differentiate instruction for their most-at risk students (e.g., Al Otaiba, Clancy-Menchetti, & Schatschneider, 2006; Fuchs & Vaughn, 2006; Harn, Chard, Biancarosa, & Kame'enui, 2011; Scanlon, Gelzheizer, Schatschneider, & Sweeney, 2008). The benefits of differentiating using small group instruction when used to target student needs have also been noted in several research studies (Lou et al., 1996; McCoach, O'Connell, & Levitt 2006; Vaughn et al., 2003). Specifically in kindergarten, the use of homogenous small group instruction has been associated with improved reading growth when used in classrooms that devote more than one hour to reading instruction (Hong & Hong, 2009). In fact, increased time spent in reading instruction was only associated with increased student reading achievement when teachers used homogeneous grouping practices to differentiate instruction for part of the instructional time devoted to reading (Hong & Hong). Small group instruction can increase opportunities for students to respond and stay engaged (e.g., La Paro et al., 2009). Fortunately, there are also converging findings that teachers who receive professional development to use data formatively to differentiate instructional decisions demonstrate higher student outcomes than teachers who do not implement these practices (e.g., Conte & Hintze, 2000; Denton, Swanson, & Mathes, 2007; Wanzek & Vaughn, 2011; Fuchs, Fuchs, Hamlett, & Allinder, 1991). In the following section we review the evidence on professional development on early literacy instruction.

Is Beginning Reading Professional Development Effective and Does the Effect Accumulate over Time?

Some evidence from quasi-experimental longitudinal studies suggests that, on average, the impact of professional development may accumulate over time. For example, Vaughn et al. (2008) examined the effects of enhancing Tier I on the early literacy outcomes across grades, but added an additional cohort of students in order to examine the cumulative effects

of professional development for the teachers. Three cohorts of kindergarten students participated in the study and were followed through first grade. Only students identified as at-risk for reading difficulties in the winter of kindergarten were followed. The first cohort entered kindergarten the year prior to the enactment of professional development and were followed through first grade. The second cohort represented students in the same teachers' classes the first year they received professional development. Finally, the third cohort consisted of the students in the teachers' classes during the second year they received professional development. Professional development included seven in-service sessions throughout the school year, monthly grade level meetings, and biweekly classroom support from a research team member aimed at addressing the differentiated needs of students at-risk for reading difficulties.

At the end of first grade, student outcomes in Cohort 3 were significantly higher than student outcomes in Cohort 1 on measures of word attack and passage comprehension. Effect sizes were in the moderate to large range. These results suggested the two years of professional development may have resulted in the highest outcomes for students at-risk for reading difficulties. However, the authors noted a significant variance around the slope on all outcome measures suggesting unspecified factors could be contributing to group differences.

Accumulating effects for professional development were also observed by Biancarosa, Bryk, and Dexter (2010), who examined several years of implementation of the Literacy Collaborative coaching model of professional development in grades K-2. Specifically, they found increases in the value of the intervention on student reading outcomes in each of the three years of the implementation, again suggesting the value of cumulative professional development for improving student outcomes. In summary, research findings indicate that professional development to help teachers learn to differentiate instruction based on student needs is effective on average and in some studies, these effects accumulated over time.

Individualized Student Instruction (ISI) as model of PD

Connor and colleagues have designed the Individualized Student Instruction (ISI) intervention to support teachers' ability to use assessment data to inform instructional amounts, types, and groupings. The ISI intervention includes three components: Assessment to Instruction (A2i) software, ongoing teacher professional development, and in-class support. A2i, incorporates both child assessment data and classroom observations in software to provide teachers with recommended amounts of instruction in a tri-dimensional framework of teacher- or child-managed instruction that is either code- or meaning-focused and that includes guidance for placing students in homogeneous instructional groupings. The theoretical frame for the particular three dimensions draws on converging evidence that the effect of a particular type of instruction depends upon students' language and literacy skills (Connor, Morrison, Fishman, Schatschneider, & Underwood, 2007, Connor, Piasta, Fischman, Schatschneider, et al., 2009; Cronbach & Snow, 1977; Juel & Minden-Cupp, 2000).

The first dimension of A2i is how instruction is managed; it is either teacher- managed or child-managed if students are working independently or with a peer. The second dimension

of A2i is reading instructional content, which is framed by the Simple View of Reading (Gough & Tunmer, 1986). Code-focused instruction addresses phonological awareness, print knowledge, and word study. Meaning-focused instruction addresses vocabulary development, listening, and reading comprehension. The third dimension of A2i is grouping, given research that small-group instruction is relatively more powerful than whole class (Connor et al., 2007) and that differentiation is easier during small-group instruction.

Al Otaiba and colleagues (Al Otaiba et al., 2011) conducted a year-long cluster randomized study to examine the effects of training kindergarten teachers to use ISI for kindergarten (ISI-K). The comparison group received a one-day summer workshop and materials that could be used to differentiate instruction. Interestingly, there were no significant differences between conditions in the amount or types of instruction teachers provided, the overall quality of instruction (warmth, classroom control, organization, and the degree to which teachers were effective at keeping students on-task during instruction), or the instructional quality of literacy (phonological awareness, letter-sound, word identification, fluency, spelling, vocabulary, and comprehension). However, teachers in the ISI-K condition did provide statistically significantly more differentiated instruction and student outcomes on a latent variable of literacy skills were over a half a standard deviation higher for students than in control classrooms (Al Otaiba, et al.). One explanation for why the only significant difference in teaching across conditions was that training teachers to differentiate was a potent ingredient of ISI-K.

Furthermore, in a three year longitudinal experimental study of ISI (first through third grade), Connor and colleagues (Connor et al., 2013) demonstrated that the effects for students of being in classrooms when teachers participated in ISI accumulated. Students whose teachers participated in first through third grade showed stronger reading performance by the end of third grade compared to students whose teachers participated in fewer years. The effects were strongest in first and second grade, but this study did not include kindergarten or address teacher change over time. In summary, there is a growing body of evidence that teachers who use these dimensions to individualize instruction based upon students' language and literacy skills in grades 1–3 achieve stronger student reading performance than teachers who do not (Connor, Piasta, Fishman, Schatscheider et al., 2009).

The previous research on professional development provides evidence for the value of Tier I instruction in a RTI model and acknowledges that many teachers may not know how to differentiate small group instruction, particularly for students at risk (Harn et al., 2011). Although ISI as a means of professional development (PD) has been carefully studied in grades K–3 using experimental methods and has been found to increase teachers' differentiation and increase student outcomes, the questions remain, is it necessary for all teachers? Do already effective teachers also benefit from ISI? Would the impact of ISI accumulate if teachers receive a second year of treatment? In this study, we sought to replicate and extend prior research related to training teachers with a goal of preventing reading difficulties. First, we use a quasi-experimental design (Shadish, Cook & Campbell, 2001) involving a subset of kindergarten teachers who participated in the initial baseline observation year for ISI-K and then received two years of ISI-K training. It is important to understand that these teachers had been nominated by their principals as being effective

beginning reading teachers and we verified their effectiveness through direct observation. Our rationale for focusing on these effective teachers was to learn whether there was a value added of professional development to differentiate instruction or whether their typical effective practice was good enough. Second, we considered their teaching experience as a moderator of professional development given that Vaughn and colleagues (2008) noted considerable variance in teacher performance and given that teaching experience can influence student outcomes (e.g., Croninger, Rice, Rathbun, & Nishio, 2007). Third, we examined whether the effects of professional development accumulated for these effective teachers. We asked two research questions. First, would providing professional development for individualizing instruction (ISI-K) to relatively effective kindergarten teachers be associated with changes in instruction? Relatedly, would teachers' experience moderate this relation? Specifically, we were interested in documenting changes in instruction from baseline across two years of professional development in terms of (1) differentiation, (2) overall quality of classroom instruction (classroom control, organization, warmth, students on-task), (3) quality of instruction directly related to the components of literacy (phonological awareness, letter-sound correspondence, decoding, vocabulary, comprehension, fluency, spelling, and writing), and the proportion of time allotted to various dimensions of instruction (i.e., teacher- or child-managed, code- or meaning-focused instruction). Second, would providing professional development for individualizing instruction (ISI-K) to relatively effective kindergarten teachers improve their students' reading and vocabulary performance?

Methods

Background and Context of the Larger Study

Data for this study come from the Florida Learning Disabilities Center Grant funded by the National Institute of Child Health and Human Development, which was a large multi-year study. In the initial baseline year of this larger study, seven schools were recruited from a mid-size city in the Southeast to over-represent students from lower socioeconomic status; four of the 7 schools were receiving Title I funding, two were designated as Reading First schools. Principals nominated effective kindergarten teachers within their schools to participate in a baseline observation of effective kindergarten reading instruction, and subsequently additional schools were added to form a cluster randomized field trial wherein schools were assigned to either ISI-K or to a workshop contrast condition that functioned as a waitlist control. The subsample for the current study consists of those effective teachers who participated in all three years of the study and were therefore eligible for two years of ISI-K. Thus, the research design is a quasi-experimental repeated measures design that examines change in teacher and student performance with a baseline (historical control) year, and two treatment years. The baseline year may help mitigate the threat that other variables rather than ISI-K were responsible for these changes, as may the efficacy findings from the cluster randomized control trial indicating that on average, kindergarten teachers in the ISI-K treatment condition did learn to differentiate instruction and that their students outperformed students in the classrooms where teachers did not receive ISI-K (Al Otaiba et al., 2011).

Setting and Participants

Schools and context—In this current analysis, four schools and 10 teachers were included. The percentage of students receiving free or reduced price lunch in these schools ranged from 34.9 to 89.2. Although the schools served a diverse population, the percentage of the schools' students who were identified as limited English proficient (LEP) was notably small, but typical for the district, and ranged from 0.0% to 2.6%. In each school, kindergarten programs were full-day and children were provided a minimum uninterrupted block of 90 minutes of instructional time for reading and language arts per district policy. Across the study, *Open Court* (Bereiter, et al., 2002) materials were implemented as the core reading program.

Teachers—The 10 teachers were female; eight were Caucasian, one was African American, and one was Hispanic. One teacher held a graduate degree, the rest held Bachelor's degrees. In the first year of the study, teachers had an average of 8.9 ($SD = 7.7$) years of teaching experience. There were no first-year teachers; five teachers reported having between 1–5 years of teaching experience. One teacher reported having between 6–10 years, three had 11–15 years, and one had 27 years of teaching experience. At the conclusion of the study, the first author interviewed teachers who confirmed that during their three years of participation in the study, ISI-K was their primary form of professional development and their only training to differentiate instruction.

Students—With the teachers' assistance, we recruited all students in their classrooms (including students who qualified for special education) for the study, which resulted in 12–15 out of a total average of 20 students per classroom for whom we received consent and who did not move during the study year. Across the three years, a total of 416 students participated (123 from the baseline year [Cohort1], 143 from the first treatment year [Cohort 2], and 150 from the second treatment year [Cohort 3]). Table 1 describes the participating student demographics including gender, age, race/ethnicity, free and reduced lunch status, and special education status. In general, there were slightly more males than females, minorities outnumbered white students, more than half of students received free or reduced price lunch, and a notably high proportion, about 15%, participated in special education.

Individualized Student Instruction Professional Development

In both treatment years, teachers received professional development using ISI-K. Training included three components. The first was a summer day long workshop that provided an overview and update of RTI research, with a focus on evidence-based Tier 1 instruction and the need for differentiated instruction (e.g., National Reading Panel, 2000). We also provided teachers with a conceptual framework of the Simple View of Reading (Gough & Tunmer, 1986) to help them to think about clustering phonological, phonics, spelling, and fluency instruction as code-focused instruction and vocabulary and comprehension as meaning-focused instruction. We showed teachers how student data would be used to group children with similar instructional needs in order to differentiate their small group activities. Teachers were also given materials and games for centers that included colored and laminated activities developed from downloadable templates created by the Florida Center for Reading Research for use in teacher/child-managed and child-managed centers (FCRR,

2007). These included both code- and meaning-focused activities. For a detailed description, see Al Otaiba et al., (2011).

The second component involved monthly in service training about differentiating reading instruction and about using A2i software (Connor et al., 2007). Reading instruction topics ranged from using dialogic small group reading strategies in homogenous groups, to planning centers for child-managed activities, to using the specific FCRR activities. After all students' language and reading scores were entered into the software, A2i provided recommendations to teachers about which children had similar skills and should be in the same small group; moreover, the software provided optimal amounts (min/day) for each of the four types of instruction for each child. These recommendations involved algorithms derived from data collected during the kindergarten baseline year and were also informed by Connor and colleagues' previous ISI studies in preschool and first grade (Connor et al., 2007). The software recommended different amounts for code- and meaning-focused activities as well. In other words, children with the weakest decoding skills, including many students receiving special education, would receive recommendations for relatively more teacher/child-managed small group code-focused instruction than peers with stronger skills.

The third component was biweekly classroom-based coaching to use ISI-K provided by staff that we referred to as "research partners". During these 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. This coaching model is supported by prior research (e.g., Gersten et al., 2008; Showers et al., 1987) and specifically by prior ISI research (e.g., Connor et al., 2007). Research partners included certified teachers, a speech language pathologist, and advanced graduate students from special education. The research partners were not always the same across years due to graduation of students. Training for research partners was provided each year by senior research staff, led by the first author. In addition, partners met regularly with senior staff to problem solve and track progress.

Measures

Teachers' literacy instruction observations—Research staff videotaped reading instruction in all 10 classrooms in November and February during all three study years. These video recordings, scheduled with teachers' advanced consent, ranged in length from 48.8 to 146.6 minutes and averaged 93.13 minutes. Videotapes focused on a stratified random sample of 10 children; we rank ordered students on their fall Letter Word Identification scores and randomly selected low-, average-, and high-performing target students from the class. During videotaping, staff used two digital video cameras with wide-angle lenses to best capture classroom instruction and field notes were kept on students who might be off camera. Thus, the videos focused on students and included whole group, small group, and individual instruction during the language arts block.

Videos were uploaded to the Noldus Observer Pro software and were coded according to the ISI Classroom Observations Coding Manual (see Connor, Morrison, Fishman, Ponitz, Underwood et al., 2009). Briefly, all actions that lasted at least 15 seconds were coded at the individual child level across three dimensions: grouping, management, and content. As the

A2i software provided teachers with recommended amounts of types of instruction, for this analysis we extracted the proportions of time allotted to the following dimensions of interest (Teacher/child managed code-focused, teacher/child managed meaning-focused, child managed code-focused, child managed meaning focused). Examples of teacher/child-managed activities that were code-focused include small group phonological blending or segmenting activities, or the teacher working with a pair of students to provide extra practice with letter sound correspondence. Examples of teacher/child-managed meaning-focused activities include dialogic or shared book reading, word webs (Tell me all the words you can think of to describe this zebra), or the teacher leading a guided writing activity. By contrast child-managed activities often occurred as centers; pairs of students might be sorting pictures by initial sounds or playing a game of Go Fish "Do you have a card that starts with /d/ like my dog card?" (one of the Florida Center for Reading Research games that is a sound matching game) that the teacher had previously taught as a small group lesson. Similarly, examples of child-managed meaning-focused activities include book browses or re-telling a story with pictures or acting it out with prompts. A goal of this analysis was to learn if the proportions of these types of instruction changed across the three years. A minimum reliability of .80 Cohen's Kappa was established among all coders.

Additionally, to address the overall quality of implementation of literacy instruction captured on the videotapes, after watching the entire videotape, coders used an observational instrument originally developed by Haager et al., (2003). We adapted the checklist by adding items related to differentiation and reading instruction and we provided target examples for each point on the scale to make this even lower inference and to increase our ability to have a high inter-rater reliability. The scale ranged from 0–3 and was rated based upon the overall video of the lesson: with a 0 for content that was not observed, 1 for "not effective," 2 for "moderately," and 3 for "highly effective."

First coders used the scale to rate differentiation of instruction and anchors for the scale were provided. For example, in the domain of differentiation, a 0 indicated no differentiation of instruction was observed and that students were mainly seen in whole group instruction, 1 indicates small-group instruction with all children doing the same activity, 2 indicates small-group and differentiated activities, and 3 indicates not only small-group and differentiated activities but that the overall content of literacy instruction is differentiated and regrouping was observed.

Second, coders used the same scale with similar anchors to rate teachers' overall quality of classroom instruction (warmth, classroom control, organization, and the degree to which teachers were effective at keeping students on-task during instruction). Third, coders rated the specific instructional quality on the same scale with analogous anchors for each literacy component (phonological awareness, letter-sound, word identification, fluency, spelling, vocabulary, comprehension, and writing). The percent of inter-rater agreement across the differentiation and quality codings and across years ranged from .92 to 1.00 with a mean of . 98; Cohen's Kappas ranged from.64 to 1.00.

Student measures—Students' word reading skills were assessed using the Woodcock Johnson-III Tests of Achievement, [WJ-III] Letter Word Identification (Woodcock, McGrew

& Mather, 2001) subtest. In this subtest, examiners asked students to identify letters in large type and then to read words that are presented 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 inter-correlations with the WJ-III Word Attack and Passage Comprehension subtests are .80 and .79 respectively.

Students' expressive vocabulary growth was assessed using the Picture Vocabulary (PV) subtest of the WJ-III (Woodcock et al., 2001). In this subtest, examiners ask students to name pictured objects that increase in difficulty. Testing is discontinued after six consecutive incorrect items. According to the WJ-III test authors, the split-half reliability of this subtest is .77.

Procedures

Each year, as part of the larger study procedures, all teachers were observed and their language arts instruction was videotaped twice per year each of the three years. The first year of the study was a baseline, observation-only year. Treatment, professional development in ISI-K, occurred in the second and third years. In addition, the students in each teacher's class were assessed three times (Fall, Winter, Spring) on standardized measures of vocabulary and word reading to facilitate growth analyses.

Assessment and training—Each year, we trained staff (graduate students) to reliably assess students. It was not possible, due to the size and complexity of the larger project, to keep assessors blind to condition. Because of this potential problem, we took steps throughout each phase of the project 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). Prior to each wave of testing, staff had to reach 98% accuracy or higher on a checklist (adapted from Sattler, 1982) evaluating the accuracy with which the assessor followed the directions in administering and scoring. Students were individually assessed by staff in quiet areas near their classrooms. For relevant subtests, compu-scoring from the commercial test producer was used to calculate standard and W-scores. W-scores, which are Rasch ability scores that provide equal-interval measurement characteristics centered at 500.

Videotaping and coding—Assessors also helped with videotaping and received additional training prior to each round of videotaping to review procedures for operating the videotaping equipment, and capturing all student instructional activities, as well as their management (i.e., teacher/child-managed, child-managed) and grouping (i.e., whole group, small group, peer, individual). After videotaping, videos were uploaded onto a dedicated computer equipped with the Noldus Observer program for later viewing and coding.

Coders included two former elementary education teachers and two trained graduate students who were trained on two coding systems, the Individualizing Student Instruction Classroom Observation System (ISI-COS; Connor et al., 2009) and an observational instrument originally developed by Haager, Gersten, Baker, and Graves (2003) that we adapted for the larger project (this instrument is described in the measures section). The

training process consisted of small group training, paired coding with a master coder, and independent coding of training videos prior to each wave of coding until a Cohen's Kappa of .80 was reached.

Data Analytic Procedures

First, to address changes in teachers' practices, we conducted three multivariate analyses of variances (MANOVA). For each multivariate analysis, we investigated whether the critical assumptions were met including homogeneity of variance (Levene's test), homogeneity of covariance (Box's test), and multivariate normality (Mahalanoblis Distances). The first MANOVA investigated the cohort differences in differentiation and in overall instructional quality (control, organization, warmth, and on task behaviors of students), the second investigated the cohort differences in the quality of literacy instruction (phonological awareness, letter-sound correspondence, decoding, vocabulary, comprehension, fluency, spelling, and writing), and the third investigated changes in the proportion of time spent in each of the dimensions of literacy instruction (teacher- or child-managed, code- or meaning-focused instruction). We chose to investigate the proportion of time because it was expected that the amount of instruction would vary by child based on his or her specific needs as guided by the A2i software. Further, the actual amount could change depending on the teacher's schedule any given year (e.g., a teacher may have had 100 minutes of uninterrupted instructional time one year and may have gone over the minimum 90 minutes, where another year she may have only had 90 minutes of uninterrupted instructional time available), thus we argue that a more realistic look at instructional time was to look at the proportion of time spent in each category. Further, although the instruction was coded at the child level, because this particular investigation focused on the teacher's instruction as a whole, we aggregated the proportional amounts to the teacher level.

Second, given the nested structure of our dataset, with students nested in classrooms, we conducted multilevel analyses (Hox, 2002) using hierarchical linear modeling (HLM). Three levels were used: time, student, and teacher. Level one represented time centered at the end of the school year (i.e., Spring testing). Level 2 represented the student and included dummy variables for cohort, with Cohort 1 as the referent group, as well as the initial score at Fall testing for the given outcome. This analysis adjusts for the pretest differences between Cohort 1 vs Cohorts 2 and 3, who began the year with lower letter word scores. Level 3 represented the teacher and included a grand mean centered variable for years of teaching. Cross-level interaction terms were constructed to determine if there was a treatment (represented by cohort) by years of teaching experience interaction. Analyses were conducted separately for Letter Word Identification and Picture Vocabulary using W-scores. The generic fully conditional model is as follows where YEARS indicates the number of years of teaching experience, C2 and C3 are dummy codes for cohort, INITIAL is the Fall score on the outcome/growth measure, and TIME is the point at which testing occurred (Fall, Winter, or Spring) centered at Spring.

$$outcome=\gamma_{000}+\gamma_{001}YEARS+\gamma_{010}C2+\gamma_{011}C2*YEARS+\gamma_{020}C3+\gamma_{021}C3*YEARS+\gamma_{030}INITIAL+\gamma_{100}TIME+\gamma_{101}YEARS*TIME+\gamma_{110}C2*TIME+\gamma_{111}C2*YEARS*TIME+\gamma_{120}C3*TIME+\gamma_{121}C3*YEARS*TIME+\gamma_{130}INITIAL*TIME+r_0+r_1TIME+u_{00}+u_{10}TIME+e$$

Results

Instructional Changes

The first question explored whether teachers changed their literacy instruction from baseline and across treatment years in terms of: differentiation, overall observed instructional quality, specific reading instructional effectiveness, and the proportion of time allotted to teacher/child managed code-focused, teacher/child managed meaning-focused, child managed code-focused or child managed meaning-focused dimensions of instruction. In terms of the differentiation and overall instructional quality, the MANOVA included ratings on differentiation, control, organization, warmth, and on task behaviors of students. The multivariate tests (Pillai's Trace) revealed a main cohort effect ($F = 2.414$, $df = 48$, $p = .02$). The tests of between-subjects effects revealed that the significant difference was only in the differentiation variable ($F = 4.610$, $df = 2$, $p = .019$). The mean rating for differentiation was 1.0 (0 = "not observed"; 1 = "not effective"; 2 = "moderately effective"; 3 = "highly effective") ($SD = .78$) for Cohort 1, 2.15 ($SD = 1.08$) for Cohort 2, and 1.90 ($SD = 1.00$) for Cohort 3; differentiation was rated significantly higher for both Cohort 2 and 3 ($p = .007$ and .032 respectively) though Cohort 3 was not significantly different than Cohort 2 ($p = .536$). There were no differences between cohorts in terms of our observations of the quality of classroom instruction. Across the cohorts, the mean rating for control was a moderately to highly effective score of 2.52 ($SD = .70$), organization was 2.42 ($SD = .30$), warmth was 2.72 ($SD = .47$), and on task behaviors was 2.45 ($SD = .59$).

In terms of the observed quality of literacy instruction, the MANOVA included ratings on the effectiveness of reading instruction on: phonological awareness, letter-sound correspondence, decoding, vocabulary, comprehension, fluency, spelling, and writing. The multivariate tests (Pillai's Trace) revealed no cohort main effect ($F = 1.194$, $df = 42$, $p = .312$). The mean ratings were 2.30 ($SD = .85$) for phonological awareness, 2.2 ($SD = .77$) for letter-sound correspondence, 2.42 ($SD = .57$) for decoding, 1.78 ($SD = .89$) for vocabulary, 2.28 ($SD = .78$) for comprehension,.8 ($SD = .87$) for fluency, 1.45 ($SD = .74$) for spelling, and 2.05 ($SD = .78$) for writing.

There were no statistically significant differences in the proportion of time spent in each dimension between cohorts ($F = .429$, $df = 50$, $p = .856$). Similarly, there were no statistically significant differences in the amount of time spent in each dimension between cohorts ($F = .469$, $df = 48$, $p = .872$).

Student Outcomes

The second research question addressed the effect of ISI-K professional development instruction on student reading and vocabulary outcomes, whether training effects accumulated, and whether the training effects were moderate by teachers' experience. Table

2 displays the student raw and standard score means at each time point. It is important to note that cohorts were significantly different in the fall on letter-word identification; Cohort 1 started significantly higher than cohorts 2 or 3 – which were not significantly different from each other. Because of the significant differences in fall scores, the initial letter-word score was included as a control variable for both end of year performance and growth for all analyses.

Letter Word Reading—We examined the impact of ISI-K professional development on students' growth and outcomes on WJ-III Letter Word Identification W-scores. The results of this analysis are presented in Table 3. The unconditional model revealed an Intra-class correlation (ICC) of 0.05 for performance and .06 for growth suggesting that 5% and 6% of variance in end of year reading performance and variance within year-growth were accounted for at the teacher level of the model, prior to adding in any predictor or control variables at any level. In the next step, initial skill was added as a control variable and was significant in predicting status as well as growth.

There was a negative value for initial skill on growth. The next model included the cohort variables. This model revealed that there was a significant cohort effect on growth and performance for both cohorts. This suggested that both Cohorts 2 and 3 had stronger outcomes and growth than Cohort 1, but were not significantly different from each other. Students in Cohort 1 ended at 391.79; in Cohort 2 students ended at 407.74, and students in Cohort 3 ended at 414.57 w-score points. These w-score points translate into to a substantial grade equivalence difference of 3 months (with Cohort 1 at a 1.2 grade equivalent and Cohort 3 at a 1.5). Notably all cohorts were performing at grade levels higher than would be expected based upon national norms despite the relatively low socioeconomic status of the sample. Students in Cohort 1 grew approximately 13.18 Letter Word Identification W-score points between testing waves, students in Cohort 2 grew approximately 20.68 W-score points, and students in Cohort 3 grew approximately 23.91 W-score points. Thus, although there was a trend that each cohort had higher achievement, findings did not indicate significant cumulative effects for the second year of training. That is, Cohort 3 was not significantly higher achieving than Cohort 2. In the last model, teaching experience was added. The model did not suggest that teaching experience had a significant main effect or cohort interaction; further, the model did not fit substantially better than the previous model.

Picture Vocabulary—Next, we analyzed the impact of teacher intervention on students' growth and outcomes on WJ-III Picture Vocabulary W scores using the same modeling procedures as were used with the Letter Word subtest. The unconditional model revealed an ICC of 0.19 for status and .02 for growth, suggesting that 19% and 2% of vocabulary performance and growth were accounted for at the teacher level. In the next step, initial skill was added as a control variable and was significant in predicting status as well as growth. The next model included the cohort variables. Unlike the letter word reading models, this model revealed that there were no significant cohort effects on growth or performance. That is Cohorts 2 and 3 had similar growth and outcomes to Cohort 1. In the last model, teaching experience was added. Unlike the letter word reading results, this model revealed a marginally significant ($p = .074$) main effect of years of teaching on end of year

performance. However, neither the model with cohort nor the model with teaching experience were significantly better fitting than the initial skill only model.

Discussion

The present study extends the literature on using professional development to enhance kindergarten reading instruction by examining the practice of effective teachers and exploring the impact of training teachers to differentiate instruction. Although our professional development used ISI-K, findings have implications for other methods of professional development that aim to improve Tier 1 instruction for successful implementation of RTI. Moreover, many kindergarten teachers are expected to implement supplemental intervention (Tier 2). If our findings were null, then schools might save considerable resources by only training less effective teachers. As a quasi-experimental study, it adds to the current limited research on differentiating Tier 1 instruction (e.g., Gersten et al., 2008) by addressing the need to understand whether effective teachers benefited from professional development and whether this is associated with stronger student reading performance. It is noteworthy that through direct observation, we confirmed that the 10 participating teachers' instruction was rated as effective in their baseline year, which was consistent with their principals' nominations. Further evidence of their efficacy was that, although a majority of students were from low SES backgrounds, and teachers provided instruction to a relatively high proportion of students with disabilities, on average, most students read on grade level, even in the baseline year. Thus, in the current context of limited resources in schools, these generally effective teachers provided a unique test of the benefits of professional development to translate knowledge into practice (Cochran-Smith & Lyttle, 1999).

The first research question examined changes in kindergarten teachers' literacy practices associated with the professional development; findings revealed that even effective teachers learned to provide more small group and differentiated instruction after treatment relative to their own baseline performance and that students' reading performance improved. However, the overall quality of instruction remained consistently "effective"; nor did we observe teachers changing the proportion of time allotted to different types of reading instruction. Clearly, findings from the present study are consistent with the randomized control trial we conducted (Al Otaiba et al., 2011) that demonstrated treatment teachers improved their differentiation of instruction relative to controls and that demonstrated students in treatment classrooms outperformed students in control classrooms. The lack of observed changes in reading instructional quality are consistent with the larger study. Our findings also converge with results of both O'Connor, et al. (2005), who observed more small group instruction after professional development and coaching than in a baseline year of instruction and with Hong and Hong (2009) who reported that homogenous differentiated small group instruction was the active ingredient in increasing student achievement.

The second research question addressed the effect of professional development instruction on student reading and vocabulary outcomes, whether training effects accumulated, and whether the training effects were moderated by teachers' experience. There was a significant effect for professional development on students' word reading. We focused on standardized

measures in the present study and we used three points in time to evaluate growth more reliably so we could use growth curve analyses. On average, students in the baseline year gained 13.8 W-score points on letter word identification compared to 20.68, and 23.91 W-scores points during the first and second year of treatment. This translates into a 49.9 % growth improvement over baseline for the first year of treatment and a 73.3% growth improvement over baseline for the second year of treatment. Thus, although there was no significant accumulation, we did observe a trend that was even stronger than was observed by Biancarosa et al. (2010), who reported gains over baseline resulting from Literacy Collaborative coaching of 16% Year one, 28% Year 2 and 32% in Year 3. This finding differs somewhat from Vaughn et al. (2008) who found students in Cohort 3 performed significantly better than students in cohort1 with no differences for Cohort 2 reported. We did not find that a second year, or more of the same professional development, led to important changes in practice, which differs from O'Conner et al., 2005, Biancarosa et al., 2010, and Vaughn et al., 2008. Furthermore, our model did not indicate that teachers' experience moderated the impact of professional development on student word reading outcomes. However, it is noteworthy that our teachers had been nominated as already effective, even though they were not previously differentiating instruction.

The vocabulary findings were not as encouraging; the vocabulary growth and outcomes were remarkably stable but, this finding may not be surprising given that the picture vocabulary measure focuses more on breadth of word knowledge and does not measure depth of vocabulary knowledge. Thus, it may not likely have been sensitive to components of instruction that were focal to ISI-K professional development.

Limitations and Directions for Future Research

As with any research, and particularly, with school-based research, there are several limitations that warrant caution about generalizing our findings, but that also warrant future research. A prior randomized control trial (Al Otaiba et al., 2011) showed that on average, students whose teachers received ISI-K training outperformed controls, but in the present study, teachers' baseline year served as their own control and we did not have a control group of children. Thus there are potential threats to internal validity. It is notable that our teachers began the study with relatively effective reading instructional practices. Thus our findings with regard to no cumulative effects of a second year of training might vary for less effective teachers. It is possible, since teachers were effective at the start of the study, that our quality ratings may not have been sensitive enough to detect change. In terms of the lack of change in the quantities of instruction, this could have been related to the consistent core reading program.

In addition, a stronger design would have ensured assessors were blind to condition and would have systematically documented coaching by research partners. Although our sample size of teachers was small, we had sufficient power to detect the large change in differentiation, but a larger and more diverse sample of teachers may shed additional light on how best to individualize professional development for teachers with a range of efficacy. We did control for students' initial status statistically, but Cohort 1 did begin with stronger letter word reading scores than did subsequent cohorts.

Acknowledgments

This work was supported by a Multidisciplinary Learning Disabilities Center Grant (P50HD052120) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development or the National Institutes of Health.

References

Al Otaiba S, Connor CM, Folsom JS, Greulich L, Meadows J, Li Z. Assessment data-informed guidance to individualize kindergarten reading instruction: Findings from a cluster-randomized control field trial. Elementary School Journal. 2011; 111(34):535–560. NIHMS ID 249051. http://dx.doi.org/10.1086/659031. [PubMed: 21818158]

Al Otaiba, S.; Clancy-Menchetti, J.; Schatschneider, C. Examining the effects of professional development designed to improve the effectiveness of early reading instruction for students at-risk for reading difficulties: Experimental and quasi-experimental approaches. In: Scruggs, T.; Mastropieri, M., editors. Advances in Learning and Behavioral Disabilities. Vol. 19. Oxford, UK: Elsevier Science/JAI Press; 2006. p. 207-243.

Bereiter, C.; Brown, A.; Campione, J.; Carruthers, I.; Case, R.; Hirshberg, J.; Treadway, GH. Open court reading. Columbus, OH: SRA McGraw-Hill; 2002.

Biancarosa G, Bryk AS, Dexter E. Assessing the value-added effects of literacy collaborative professional development on student learning. Elementary School Journal. 2010; 111(1):7–34. http://dx.doi.org/10.1086/653468.

Cochran-Smith, M.; Lytle, S. Relationship of knowledge and practice: Teacher learning in communities. In: Iran-Nejad, A.; Pearson, CD., editors. Review of Research in Education. Vol. 24. Washington, DC: AERA; 1999. p. 249-306.

Connor CM, Morrison FJ, Fishman BJ, Ponitz CC, Glasney S, Underwood PS, Schatschneider C. The ISI classroom observation system: Examining the literacy instruction provided to individual students. Educational Researcher. 2009; 38:85–99. http://dx.doi.org/10.3102/0013189X09332373.

Connor CM, Morrison FJ, Fishman BJ, Schatschneider C, Underwood P. THE EARLY YEARS: Algorithm-Guided Individualized Reading Instruction. Science. 2007; 315(5811):464–465. [PubMed: 17255498]

Connor CM, Morrison FJ, Fishman B, Crowe EC, Al Otaiba S, Schatshneider C. A longitudinal cluster-randomized control study on the accumulating effects of individualized literacy instruction on students' reading from 1st through 3rd grade. Psychological Science. 2013; 24 (8):1408–1419. [PubMed: 23785038]

Connor CM, Piasta SB, Fishman B, Glasney S, Schatschneider C, Crowe E, Morrison FJ. Individualizing student instruction precisely: Effects of child × instruction interactions on first graders' literacy development. Child Development. 2009; 80:77–100. http://dx.doi.org/10.1111/j.1467-8624.2008.01247.x. [PubMed: 19236394]

Conte KL, Hintze JM. The effects of performance feedback and goal setting on oral reading fluency within curriculum-based measurement. Diagnostique. 2000; 25:85–98.

Cronbach, LJ.; Snow, RE. A handbook for research on interactions. New York: Irvington; 1977. Aptitudes and instructional methods.

Croninger RG, Rice JK, Rathbun A, Nishio M. Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement. Economics of Education Review. 2007; 26:312–324. http://dx.doi.org/10.1016/j.econedurev.2005.05.008.

Crowe E, Connor CM, Petscher Y. Examining the core: Relations among reading curriculums, poverty, and first through third grade reading Achievement. Journal of School Psychology. 2009; 47:187–214. [PubMed: 19328286]

Denton CA, Swanson EA, Mathes PG. Assessment-based instructional coaching provided to reading intervention teachers. Reading and Writing Quarterly: Overcoming Learning Difficulties. 2007; 20:569–590. DOI: 10.1007/s11145-007-9055-0

Florida Center for Reading Research. Student Center Activities: Grades K-1. 2007. Instructional routines. Retrieved from http://www.fcrr.org/curriculum/SCAindex.shtm

Fletcher, JM.; Denton, CA.; Fuchs, L.; Vaughn, SR. Multi-tiered reading instruction: Linking general education and special education. In: Richardson, SO.; Gilger, JW., editors. Research-based education and intervention: What we need to know. Baltimore, MD: The International Dyslexia Association; 2005. p. 21-43.

Fuchs LS, Fuchs D, Hamlett CL, Allinder RM. The contribution of skills analysis to curriculum-based measurement in spelling. Exceptional Children. 1991; 57(5):443–452.

Fuchs L, Vaughn S. A Response to "Competing views: A dialogue on response to intervention": why response to intervention is necessary but not sufficient for identifying students with learning disabilities. Assessment for Effective Intervention. 2006; 32(1):58–61. http://dx.doi.org/10.1177/15345084060320010801.

Gersten, R.; Compton, D.; Connor, CM.; Dimino, J.; Santoro, L.; Linan-Thompson, S.; Tilly, WD. A practice guide. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education; 2008. Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. NCEE 2009-4045Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

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. Washington, DC: U.S. Department of Education, Institute of Education Sciences; 2008.

Gough PB, Tunmer WE. Decoding, reading, and reading disability. RASE: Remedial & Special Education. 1986; 7:6–10.

Haager, D.; Gersten, R.; Baker, S.; Graves, AW. The English-language learner classroom observation instrument for beginning readers. In: Vaughn, S.; Briggs, KL., editors. Reading in the classroom: Systems for the observation of teaching and learning. Baltimore, MD: Brookes; 2003. p. 111-144.

Harn BA, Chard DJ, Biancarosa G, Kame'enui EJ. Coordinating instructional supports to accelerate at-risk first grade readers performance: An essential mechanism for effective RTI. Elementary School Journal. 2011; 112(2):332–354. http://dx.doi.org/10.1086/661997.

Hong G, Hong Y. Reading instruction time and homogeneous grouping in kindergarten: An application of marginal mean weighting through stratification. Educational Evaluation and Policy Analysis. 2009; 31(1):54–81. http://dx.doi.org/10.3102/0162373708328259.

Hox, J. Multilevel analysis. Mahwah, NJ: Lawrence Erlbaum; 2002.

Juel C, Minden-Cupp C. Learning to read words: Linguistic units and instructional strategies. Reading Research Quarterly. 2000; 35(4):458–492. http://dx.doi.org/10.1598/RRQ.35.4.2.

La Paro KM, Hamre B, Locasale-Crouch J, Pianta RC, Bryant D, Early D, Burchinal M. Quality in kindergarten classrooms: Observational evidence for the need to increase children's learning opportunities in early education classrooms. Early Education and Development. 2009; 20(4):657–692. http://dx.doi.org/10.1080/10409280802541965.

Lou Y, Abrami PC, Spence JC, Poulsen C, Chambers B, d'Apolionia S. Within-class grouping: A meta-analysis. Review of Educational Research. 1996; 66:423–458. http://dx.doi.org/10.3102/00346543066004423.

McCoach DB, O'Connell AA, Levitt H. Ability grouping across kindergarten using an Early Childhood Longitudinal Study. Journal of Educational Research. 2006; 99(6):339–346. http://dx.doi.org/10.3200/JOER.99.6.339-346.

National Center for Education Statistics. The nation's report card: Reading 2011 (NCES 2012–457). Washington, DC: Institute of Education Sciences, U.S. Department of Education; 2011.

National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, DC: U.S. Government Printing Office; 2000.

O'Connor RE, Fulmer D, Harty KR, Bell KM. Layers of reading intervention in kindergarten through third grade: Changes in teaching and student outcomes. Journal of Learning Disabilities. 2005; 38(5):440–455. http://dx.doi.org/10.1177/00222194050380050701. [PubMed: 16329445]

Rosenthal R, Rosnow RL. Applying hamlet's question to the ethical conduct of research: A conceptual addendum. American Psychologist. 1984; 39(5):561–563. http://dx.doi.org/10.1037/0003-066X.39.5.561.

Sattler JM. Age effects on Wechsler Adult Intelligence Scale–Revised tests. Journal of Consulting and Clinical Psychology. 1982; 50(5):785. http://dx.doi.org/10.1037/0022-006X.50.5.785. [PubMed: 7142556]

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(3):346–359. http://dx.doi.org/10.1016/j.lindif.2008.05.002. [PubMed: 19325710]

Shadish, WR.; Cook, TD.; Campbell, JR. Experimental and quasi-experimental designs for generalized causal inference. New York: Houghton Mifflin Company; 2002.

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.

Vaughn S, Linan-Thompson S, Kouzekanani K, Bryant DP, Dickson S, Blozis SA. Reading instruction grouping for students with reading difficulties. Remedial and Special Education. 2003; 24(5):301–315. http://dx.doi.org/10.1177/07419325030240050501.

Vaughn, S.; Linan-Thompson, S.; Woodruff, T.; Murray, C.; Wanzek, J.; Scammaca, N., et al. Effects of professional development on improving at-risk students' performance in reading. In: Greenwood, CR.; Kratochwill, RR.; Clements, M., editors. Schoolwide prevention models: Lessons learned in elementary schools. New York: Guilford; 2008. p. 115-142.

Vaughn, S.; Wanzek, J.; Woodruff, AL.; Linan-Thompson, S. Prevention and early identification of students with reading disabilities. In: Haager, DH.; Vaughn, S.; Klingner, JK., editors. Evidence-based reading practices for response to intervention. Baltimore, MD: Brookes; 2007. p. 11-27.

Wanzek J, Vaughn S. Research-based implications from extensive early reading interventions. School Psychology Review. 2007; 36:541–561.

Wanzek J, Vaughn S. Is a three-tier reading intervention model associated with reduced placement in special education? Remedial and Special Education. 2011; 32(2):167–175. http://dx.doi.org/10.1177/0741932510361267.

Woodcock, RW.; McGrew, KS.; Mather, N. Woodcock-Johnson III tests of achievement. Itasca, IL: Riverside Publishing; 2001.

Zirkel PA, Thomas LB. State laws for RTI: An updated snapshot. Teaching Exceptional Children. 2010; 42(3):56–63.

Table 1

Student Demographics

	C1	C2	C3	Combined
N	123	143	150	416
% Male	45.5	58	54	52.9
Age in Fall	5.4	5.3	5.2	5.3
Race				
% Black	53.7	60.1	59.3	57.9
% White	34.1	37.8	31.3	34.4
% Other	12.2	2.1	9.4	7.6
% Hispanic	2.4	2.1	4.7	3.1
% FARL	52	52.4	62.7	56
% LEP	1.6	0.7	0.7	1
% in Special Education	16.3	15.4	14.7	15.4

Note. C1 = Cohort 1. C2 = Cohort 2. C3 = Cohort 3. FARL = Free and Reduced Price Lunch program. LEP = Limited English Proficiency.

Table 2

Descriptive Statistics of Student Standard Score Variables

	Cohort 1			Cohort 2			Cohort 3			ANOVA		Full Sample		
	M	***SD***	***n***	**M**	***SD***	***n***	**M**	***SD***	***n***	**F**	***p***	**M**	***SD***	***n***
Student Achievement														
Letter word Identification Standard Scores														
Fall	108.96	13.648	123	97.46	10.646	143	98.24	12.220	150	36.24	0.000	101.14	13.152	416
Winter	111.10	12.768	123	104.08	11.551	143	106.58	12.475	150	11.01	0.000	107.06	12.547	416
Spring	111.76	13.313	123	104.97	14.049	143	108.31	13.986	150	8.02	0.000	108.18	14.044	416
Picture Vocabulary Standard Scores														
Fall	101.07	9.358	123	100.69	10.080	143	100.25	9.832	150	0.24	0.788	100.64	9.764	416
Winter	106.54	16.004	123	100.27	12.722	143	101.15	13.141	150	7.74	0.010	102.44	14.138	416
Spring	102.50	9.448	123	99.36	10.299	143	99.35	10.857	150	4.08	0.018	100.29	10.341	416

Table 3

Letter Word Identification Results

	Unconditional Model					Initial Skill					Cohort + Initial Skill					Cohort, Experience, and Fall LW				
Fixed Effect	**Coef.**	**se**	**t-ratio**	***df***	***p***	**Coef.**	**se**	**t-ratio**	***df***	***p***	**Coef.**	**se**	**t-ratio**	***df***	***p***	**Coef.**	**se**	**t-ratio**	***df***	***p***
Status	405.44	2.22	182.33	9	0.000	405.60	1.63	248.84	9	0.000	391.79	2.04	191.84	9	0.000	391.69	1.99	197.23	8	0.000
Experience																0.15	0.23	0.62	8	0.551
Cohort 2											15.95	2.39	6.68	412	0.000	15.88	2.39	6.64	412	0.000
Experience																0.27	0.28	0.96	412	0.338
Cohort 3											22.78	2.37	9.60	412	0.000	22.29	2.38	9.65	412	0.000
Experience																−0.09	0.27	−0.35	412	0.729
Fall LW-ID						0.67	0.04	17.76	414	0.000	0.82	0.04	21.69	412	0.000	0.82	0.04	21.57	412	0.000
Growth	19.69	0.78	25.32	9	0.000	19.68	0.82	24.15	9	0.000	13.18	1.07	12.30	9	0.000	13.11	1.04	12.56	8	0.000
Experience																0.09	0.12	0.76	8	0.469
Cohort 2											7.50	1.29	5.81	412	0.000	7.48	1.29	5.79	412	0.000
Experience																0.13	0.15	0.84	412	0.403
Cohort 3											10.73	1.28	8.36	412	0.000	10.83	1.29	8.42	412	0.000
Experience																−0.07	0.15	−0.51	412	0.609
Fall LW-ID						−0.14	0.02	−6.99	414	0.000	−0.07	0.02	−3.47	412	0.001	−0.07	0.02	−3.43	412	0.001
Random Effect	Variance		***df***	x^2	***p***	Variance		***df***	x^2	***p***	Variance		***df***	x^2	***p***	Variance		***df***	x^2	***p***
Level 1	109.59					62.13					62.09					62.04				
Level 2																				
Status	590.96		406	3033.97	0.000	337.49		405	3045.20	0.000	268.14		403	2498.33	0.000	266.76		403	2489.03	0.000
Growth	53.28		406	799.96	0.000	77.50		405	1240.25	0.000	62.95		403	1040.15	0.000	62.49		403	1035.21	0.000
Level 3																				
Status	32.37		9	30.09	0.001	16.99		9	28.06	0.001	10.89		9	24.19	0.004	8.50		8	20.96	0.007
Growth	3.40		9	23.29	0.006	3.98		9	24.36	0.004	2.49		9	19.91	0.018	1.86		8	16.79	0.032

Note: Letter word Identification from Woodcock et al., (2001). Status refers to end of year performance.