THE UNIVERSITY OF MEMPHIS | Center for Research in Educational Policy

Efficacy of the Leveled Literacy Intervention System for K–2 Urban Students:

An Empirical Evaluation of LLI in Denver Public Schools

Study Dates: 2011–2012

THE UNIVERSITY OF MEMPHIS | Center for Research in Educational Policy

Efficacy of the Leveled Literacy Intervention System for K-2 Urban Students:

An Empirical Evaluation of LLI in Denver Public Schools

November 2013

Carolyn R. Ransford-Kaldon
Cristin L. Ross
Christine C. Lee
E. Sutton Flynt
Louis A. Franceschini
Todd A. Zoblotsky

Center for Research in Educational Policy
The University of Memphis
866-670-6147
http://www.memphis.edu/crep

Acknowledgements

This study was supported by funding from Heinemann Publishing, publisher of professional resources and a provider of educational services for teachers in kindergarten through college. We appreciate the opportunity to evaluate one of their widely used products, the Leveled Literacy Intervention system (LLI), and thank Denver Public Schools for the opportunity to work with their schools, teachers, and students.

For questions about this report or the research study, please contact Dr. Carolyn Ransford-Kaldon, Principal Investigator, at cransfrd@memphis.edu or 1-866-670-6147.

Table of Contents

List of Figures

List of Tables

Executive Summary

This report summarizes the results of an efficacy study of the Leveled Literacy Intervention system (LLI) conducted by the Center for Research in Educational Policy (CREP) in Denver Public Schools (DPS) during the 2011-2012 school year. Developed by authors Irene C. Fountas and Gay Su Pinnell and published by Heinemann, LLI is a short-term, small-group, supplemental literacy intervention system that uses a series of "leveled" texts (i.e., texts of progressing difficulty) to help students in kindergarten through second grade achieve grade-level competency in literacy. There were four key purposes of this study: (1) to determine the efficacy of the Leveled Literacy Intervention system (LLI) in increasing literacy achievement for urban K-2 students and associated student subgroups; (2) to examine LLI program implementation fidelity in urban settings; (3) to determine perceptions of the LLI system according to relevant stakeholders; and (4) to corroborate the Fountas & Pinnell Benchmark Assessment System with established literacy assessments (i.e., the Developmental Reading Assessment, 2nd Edition [DRA2] and the STAR Early Literacy Assessment).

A total of 320 K-2 students participated in this mixed-methods randomized controlled trial (RCT) that included both quantitative and qualitative data. The students were matched demographically and randomly assigned to treatment and control groups. During the study, the treatment group students participated in LLI (18 weeks for first and second grade and 12 weeks for kindergarten), while the control group students could not receive LLI until after the study was over. The control students could receive other literacy interventions, however. Treatment and control group students' pre- and posttest performance was compared on three measures of student literacy achievement: the Fountas & Pinnell Benchmark Assessment System, the DRA2, and the STAR Early Literacy Assessment. Additional DRA2 and STAR data from a comparison group of 386 students who received LLI during the 2011-2012 school year but did not participate in the RCT was also examined. Further, an assessment of LLI implementation fidelity included independent observations of LLI groups and teacher-provided data taken from the LLI Online Data Management System. The quality of the core literacy instruction was also examined using classroom literacy observations, and feedback regarding LLI and the participating schools' core literacy programs was obtained from LLI teachers, classroom teachers, principals, parents/guardians, district literacy specialists, and independent on-site researchers who collected data for the study. Results from the current study are summarized by research question below.

1. What progress in literacy achievement, if any, do urban students who receive LLI make compared to students who receive core literacy instruction alone?

The results of the current study revealed that LLI positively impacts urban students' literacy achievement in kindergarten, first grade, and second grade students. On average, kindergarten students who received LLI progressed from Level A to Level C on the Fountas & Pinnell Benchmarks and outperformed their control group counterparts (who progressed from Level A to Level B) by one benchmark level. In first grade, LLI students outperformed the control group by one benchmark level on average, progressing from Level A to Level E; control group students progressed from Level A to Level D. Finally, second grade LLI students outperformed their control counterparts by less than one benchmark level on average, with both groups starting at Level E/F and finishing around Level I. Given the progress made, though not statistically significant, post-hoc analyses were conducted in order to examine these effects in a larger sample. In the combined-sample post-hoc analyses, the treatment group gained about 4.5 levels, finishing close to Level J, while the control group only gained around 3 levels, finishing close to Level I. These results are similar to those seen when looking only at the Denver second grade students; however, with the increased combined sample size, statistical significance was also attained.

Significant positive effects were also found for LLI students on the DRA2 in kindergarten and second grade, but no results were found at any of the three grade levels on the STAR (most likely due to extremely small sample sizes). Further, demographic subgroups including males, females, Hispanic students, and ELL students were shown to benefit from LLI across the three grade levels. Fidelity of LLI implementation (i.e., the degree to which LLI was implemented as designed) was shown to have some impact on student achievement in kindergarten and first grade but not second grade. Additionally, the amount of LLI attendance – relative to the recommended amount – appeared to have minimal effects on student achievement. Finally, comparison data from DPS students who received LLI but were not part of the randomized study revealed that students who received LLI – and any additional instructional time, literacy support, or intervention services they needed – during the 2011-2012 school year made highly significant gains on the DRA2 at all three grade levels, but only second grade students made significant gains on the STAR. However, these results should be interpreted with caution because the comparison group analysis did not include a control group who did not receive LLI, so it is not possible to infer a causal relationship between comparison group students' LLI participation and their growth in literacy scores.

2. At what level of fidelity to the program model is LLI implemented by teachers participating in the study?

Overall, the observation results from the current study suggest that LLI was implemented with a high degree of fidelity to design. In most of the observations, the majority of lesson components received high fidelity ratings. Further, the on-site researchers generally concluded that the lessons they observed were delivered as designed. Additionally, the observation results revealed that overall LLI implementation was consistent across the school year, with strong fidelity scores received at both time points when the observations were conducted. These observation results were corroborated by self-report feedback from the participating LLI teachers. Finally, the LLI attendance records from the current study revealed that, on average, students received less than the model's recommended number of instructional days (i.e., 62 days instead of 90 for first and second grade, and 45 days instead of 70 for kindergarten). Although second grade students made few significant gains, kindergarten and first grade students made significant progress in their literacy achievement despite receiving less than the recommended amount of instruction. This finding suggests that LLI can still be effective during a relatively shorter timeframe, which may be valuable to districts with a large number of students to serve or limited time in which to implement early literacy interventions.

3. What are stakeholders' perceptions of the LLI system and the core literacy program?

Overall, LLI teachers, classroom teachers, principals, parents/guardians, district literacy specialists, and on-site researchers shared extremely positive perceptions of the LLI system and its impact on struggling students' literacy. Stakeholders felt that LLI has benefits for students' literacy achievement and skills as well as their enjoyment, enthusiasm, and confidence related to reading and writing. Stakeholders also reported positive perceptions of such aspects of the LLI system as its design and organization, instructional components, and materials (particularly the lesson books and take-home books). However, although stakeholders generally perceived that LLI is helpful for English Language Learner students, there was some disagreement about its benefits for students classified as special education. In general, stakeholders agreed that LLI may be too fast-paced for learners who need to spend more time on certain concepts. Additionally, stakeholders raised areas of concern including the expense of the LLI professional development and kits, the relatively small number of students that schools can serve with LLI (particularly when considering time and staffing limitations), and the fact that

it is difficult for districts and schools to achieve the recommended amount of LLI instructional time during the school year.

Regarding the core literacy instruction, stakeholders' perceptions were generally positive, although some areas of concern were identified. Stakeholders perceived that their schools are generally supportive of literacy and provide a learning environment conducive to literacy development. Further, stakeholders shared positive perceptions of the core literacy program's impact on students' achievement, enthusiasm, and confidence related to reading and writing, as well as such aspects of the program as small group instruction, guided reading, differentiated instruction, support for ELL students, and teacher flexibility and autonomy. However, a large percentage of stakeholders agreed that the core literacy instruction needs improvement. Areas for improvement suggested by stakeholders included the need for a consistent and comprehensive curriculum, increased parental involvement, improved resources, and more individualized support for students.

4. How do the results of the Fountas & Pinnell Benchmark Assessment System compare to those of the Developmental Reading Assessment, 2nd Edition (DRA2) and the STAR Early Literacy Assessment?

Grade level equivalence information was available to compare treatment and control group students' scores on the Fountas & Pinnell Benchmarks with their scores on the DRA2 but not on the STAR Early Literacy Assessment, which also did not have a sufficient sample size in the current study to support such an analysis. Further, definitive conclusions can only be drawn about the comparison between kindergarten benchmark and DRA2 scores due to the high number of unavailable DRA2 scores in first and second grades for the time period used for analysis. Overall, there was a low rate of agreement between kindergarten students' posttest instructional levels on the benchmarks and their posttest DRA2 scores. However, approximately one-quarter of these kindergarten students did not have a comparable posttest DRA2 level for their posttest benchmark level; therefore, these students were automatically considered to have "no match" because no corresponding DRA2 level was available in the grade level equivalence information provided by Heinemann. Further, kindergarten students' scores on the DRA2 were more frequently categorized as "proficient" than on the benchmarks, with agreement between the benchmarks and DRA2 regarding proficiency or non-proficiency status occurring only half of the time. Finally, although there was a low level of agreement between both first and second grade benchmark and DRA2 scores, there was a trend suggesting high agreement between the two assessments on proficiency or non-proficiency status in these grades. However, due to the extremely small sample size, the first and second grade results are highly inconclusive and must be interpreted with caution.

The current study encountered several limitations that may limit the generalizability of the findings and that prevented researchers from obtaining adequate power to draw definitive conclusions. These limitations included a small sample size of students participating in the study (particularly in second grade), the researchers' inability to control testing conditions for two of the outcome measures, the acknowledgement that control group students were allowed to receive other supplemental literacy services besides LLI while they were participating in the study, and that treatment group students did not receive the recommended amount of LLI instructional time. However, despite these limitations, the current study found significant positive effects of LLI on urban students' literacy achievement when implemented with fidelity to the LLI model. Further, stakeholders in Denver Public Schools – including teachers, administrators, and parents/guardians – were supportive of LLI and perceived positive benefits of the LLI system for their students. Altogether, the results from this evaluation allow us to conclude

that LLI positively impacts urban students' literacy skills, particularly in kindergarten and first grade. These results also suggest that continued implementation of LLI would be beneficial in DPS and offer an opportunity for research-based recommendations that may enhance the system, future research, and ultimately student achievement. A list of these recommendations – including items related to LLI design, implementation, and professional development, as well as future directions for LLI research – may be found in the main body of this report.

Introduction

This report summarizes the results of an efficacy study of the Leveled Literacy Intervention system (LLI) conducted by the Center for Research in Educational Policy (CREP) in Denver Public Schools (DPS) during the 2011-2012 school year. Schools within DPS have widely adopted the targeted, small-group implementation model of LLI with support from Heinemann consultants providing LLI professional development, along with continuing support and development provided by trained DPS staff. This report focuses on the implementation and impact of this model in a sample of 13 elementary and K-8 DPS schools who voluntarily adopted the LLI system.

This study was designed to replicate and extend the findings of a previous study conducted by CREP during the 2009-2010 school year in rural and suburban settings. While the efficacy of LLI was established in the prior study for the students in these settings, additional research was needed to establish LLI's effectiveness in urban school districts, which typically have limited resources, high rates of student mobility, issues relating to teacher retention, and students who come from high-risk neighborhoods into high-risk schools. Further, urban school districts have a great need for research-based programs that clearly demonstrate a positive impact on student achievement. The goal of this study was to examine the extent to which participation in LLI influenced urban student literacy achievement and teachers' instructional practices regarding literacy. Additionally, this study was designed to determine the strengths and weaknesses of LLI according to relevant stakeholders.

CREP is a State of Tennessee Center of Excellence, located at the University of Memphis, whose mission is to implement a research agenda associated with educational policies and practices in preK-12 public schools and to provide a knowledge base for use by educational practitioners and policymakers. Since 1989, the Center has served as a mechanism for mobilizing community and university resources to address educational problems and to meet the University's commitment to primary and secondary schools. Functioning as a part of the College of Education, Health, and Human Sciences, CREP seeks to accomplish its mission through a series of investigations conducted by Center personnel, college and university faculty, and graduate students.

Theoretical Framework

Developed by authors Irene C. Fountas and Gay Su Pinnell (2008) and published by Heinemann, LLI is a short-term, small-group, supplemental literacy intervention system designed for students in kindergarten through second grade who struggle with literacy. The goal of LLI is to provide intensive support to help these early learners quickly achieve grade-level competency. The LLI system has its roots in the theoretical and empirical work of Marie Clay (1991) and of Fountas and Pinnell (1996, 2006), and its lesson design draws from empirical research on reading acquisition and reading difficulties, language learning, and student motivation (e.g., Armbruster, Lehr, & Osborn, 2001; National Institute of Child Health and Human Development, 2000a; National Institute of Child Health and Human Development, 2000b). The LLI materials are based around a series of "leveled" texts (i.e., texts of progressing difficulty) with difficulty measured by the Fountas & Pinnell Text Level Gradient™ (Fountas & Pinnell, 2007). LLI emphasizes systematic and explicit instruction in phonological awareness, phonics, fluency, comprehension, and the expansion of oral language skills, including vocabulary.

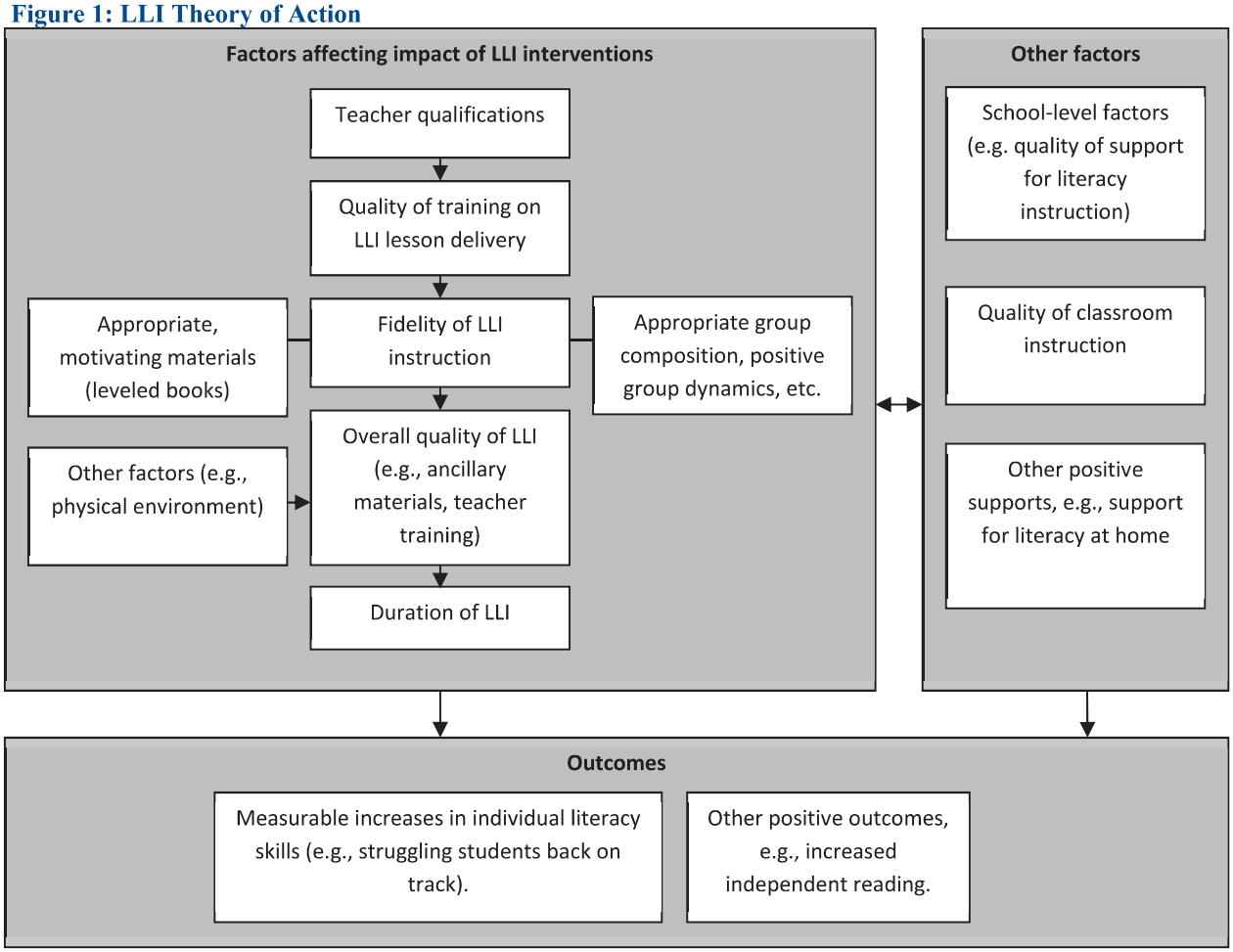
Research suggests that children with poor early reading skills continue to struggle with reading and writing in the later grades and are more likely to drop out of school (Alexander, Entwisle, & Horsey,

1997; Juel, 1988; Tabors, Snow, & Dickinson, 2001). However, there is evidence that quality early intervention programs can prevent the development of long-term reading deficiencies (Heibert & Taylor, 1994; Wanzek & Vaughn, 2007). Previous studies by Harrison, Peterman, Grehan, Ross, Dexter, and Inan (2008) and Peterman, Grehan, Ross, Gallagher, and Dexter (2009) showed that K-2 students enrolled in LLI made significant gains on the Gates-MacGinitie Reading Test, with 25 to 44% of students reading at or above average by the end of the program. Further, Ransford-Kaldon, Flynt, Ross, Franceschini, Zoblotsky, Huang, and Gallagher (2010) conducted a mixed-methods randomized controlled trial (RCT) and found that rural and suburban K-2 students who received LLI significantly outperformed their matched counterparts in the control group by an average of 1.5 levels in the Fountas & Pinnell Benchmark Assessment System. The current study – also utilizing a mixed-methods RCT – expanded on prior research findings and examined whether urban students receiving LLI achieved greater literacy gains than students receiving core literacy instruction alone.

Theory of Action

Figure 1, below, represents our preliminary identification of some of the key factors that impact the quality of LLI and its overall impact on student learning. Generally, the model states that measurable increases in student literacy growth and other positive outcomes will result upon completion of a certain number of intervention sessions as well as from a combination of factors, including those directly related to the intervention itself—*and* other non-LLI factors such as the quality of the students' core literacy instruction and support they receive for literacy at home.

Figure 1: LLI Theory of Action



LLI Factors

Factors that we propose that may directly affect the quality of LLI include: teacher qualifications and skills; the quality of teachers' LLI training; the level of materials matched appropriately to students' reading level and progress (i.e., teachers select the appropriate sequence of leveled books at the students' reading level); the overall quality and fidelity of instruction; composition of the student group (i.e., students at more or less the same reading level or combinations that include students with special needs and ELL students); and other factors such as the learning environment and the duration of the intervention. We discuss each of these factors in turn, along with how these factors may be measured.

Teacher Qualifications and Training

A cluster of teacher factors may affect the overall quality of LLI. For example, teachers who already have a certain level of experience and skill in delivering literacy interventions may be more likely to benefit from LLI training and more likely to make good use of the materials than teachers who are less experienced. It was, therefore, important to have background information on the teachers providing the intervention—including years of experience, degrees attained, and other relevant training and job experience. Finally, the level and quality of professional development provided to LLI teachers, subsequent to their selection, is critical to LLI implementation.

Appropriate Selection of Materials

The LLI system depends heavily on the use of leveled, high-interest texts that are selected after assessing students with the Fountas & Pinnell Benchmark Assessment System in order to determine each student's beginning instructional reading level and independent reading level. As the intervention progresses, teachers select the progress or sequence of the leveled texts that students read. Therefore, it was important to evaluate the match between the leveled texts used for instruction, at the beginning and throughout LLI. It was also important to assess students' degree of engagement in using the materials. Careful observations of the interventions themselves, along with teacher surveys and logs from the LLI Data Management System, were used to measure K-2 students' achievement and the level of student engagement during lesson instruction.

Implementation Fidelity

High-fidelity implementation of LLI depends in part on the amount and quality of professional training provided, the support of school administrators, and a commitment from the teachers selected to be LLI interventionists. We wanted to know how much training had been provided in the district—as well as the quality and relevance of the training from the teachers' perspective. Relevant teacher demographics and perceptions regarding the LLI system and training were therefore obtained from the participating K-2 teachers.

Group Composition and Behavior

Since LLI is a small-group intervention, it was important to know about the individual characteristics of each group. For example, if students have different needs (e.g., students can be struggling at the same instructional reading level for different reasons), it could be difficult for a teacher to provide instruction that meets the needs of the group as a whole. Also, if one student is unmotivated or disruptive, this could presumably impact the social dynamics of the group—and ultimately affect the

success of the intervention. It was also important to know whether some or all of the students in the group had been identified as students with special needs or ELL students. With this in mind, we ascertained each group's demographics and characteristics (i.e., literacy level).

Other Intervention Factors

The impact of the LLI system is directly affected by other factors, such as the duration of the intervention. We utilized part of the LLI Data Management System to obtain a record of the actual number of days of intervention for each group (14 weeks is recommended for kindergarten and 18 weeks is recommended for first and second grades). In addition, data was collected relative to adherence to the recommended 30-minutes-a-day, 5-days-a-week instructional cycle. Finally, we conducted surveys and structured focus groups with on-site researchers, LLI teachers, and other stakeholders to gather additional qualitative data related to instructional time, student absences and mobility, materials, and other pertinent factors.

Non-LLI Factors

Non-LLI factors include school-level variables, such as the overall support for literacy in the school and the quality of literacy instruction in the child's regular classroom, as well as such external variables as the amount of literacy support that students receive at home.

School-Level Variables

A full understanding of how the LLI system works in a particular context—and why it may be more or less successful from one school to another—will be usefully informed by understanding school-level factors, such as overall support for literacy in the school (e.g., literacy may receive more emphasis and resources in some schools than in others) and school-level attention to the needs of struggling students. Certain schools, in other words, may provide contexts that tend to promote a high-quality implementation of LLI. In the current study, we measured these factors through surveys of both LLI teachers and regular classroom teachers, as well as surveys and/or focus groups with administrative staff.

Quality of Core Literacy Instruction

Students receiving LLI in the current study were also receiving literacy instruction from their regular classroom teacher, and some portion of any measured gains in literacy skills over the period of the study may be attributable to the quality of literacy instruction that the children received in their regular classroom. Students who are receiving high-quality literacy instruction in the classroom *and* high-quality literacy intervention are more likely to show progress than students who receive the same quality of intervention but lower-quality classroom instruction. We used classroom observations and a classroom teacher survey regarding the school's K-2 literacy program as a measure of the nature of core literacy instruction that the students received.

Home Literacy Support

The literacy instruction received by students in the current study – whether core literacy instruction alone or combined with the LLI intervention – could have been supplemented by positive supports for literacy at home. These supports may have included a parent/guardian, sibling, or other

relative reading to the child, listening to the child read aloud, or helping the child practice such strategies as sounding out words, identifying high-frequency words, or writing. Additionally, family members may engage in varying degrees of involvement with literacy activities at the child's school, such as volunteering to read in the classroom or communicating with the child's teacher regarding his/her progress. In order to ascertain the amount of literacy support that students in the study were receiving at home, we surveyed parents/guardians or other family members regarding home literacy practices.

Research Questions

There were four key purposes of this study: (1) to determine the efficacy of the Leveled Literacy Intervention system (LLI) in increasing literacy achievement for urban K-2 students and associated student subgroups; (2) to examine LLI program implementation fidelity in urban settings; (3) to determine perceptions of the LLI system according to relevant stakeholders; and (4) to corroborate the Fountas & Pinnell Benchmark Assessment System with established literacy assessments (i.e., the Developmental Reading Assessment, 2nd Edition [DRA2] and the STAR Early Literacy Assessment). This study focused on one U.S. school district and comprised 320 K-2 students who were matched demographically and randomly assigned to treatment and control groups. The evaluation used a mixed-methods design to address the following confirmatory and exploratory research questions:

Confirmatory Questions

1. What progress in literacy achievement, if any, do urban students who receive LLI make compared to students who receive core literacy instruction alone?

    a. Does the effectiveness of LLI vary by the following subgroups: English Language Learners, students with a special education designation, and ethnic minorities (i.e., African-American and Hispanic students)?

2. At what level of fidelity to the program model is LLI implemented by teachers participating in the study?

3. What are stakeholders' perceptions of the LLI system and the core literacy program?

Exploratory Question

4. How do the results of the Fountas & Pinnell Benchmark Assessment System compare to those of the Developmental Reading Assessment, 2nd Edition (DRA2) and the STAR Early Literacy Assessment?

Method

The present study of the LLI system employed a mixed-methods randomized controlled trial (RCT) that included both quantitative and qualitative data. A matched-pair design was utilized to ensure equivalency between treatment and control groups, and pre-post comparisons of student achievement in literacy were conducted. In addition, an assessment of fidelity of implementation – including LLI and classroom observations as well as feedback from teachers, parents/guardians, school- and district-level administrators, and independent on-site researchers – yielded both observational and self-reported data.

Multiple instruments were utilized in the evaluation, including three measures of reading achievement for evaluating students' progress in literacy; two observational tools for assessing LLI and classroom teachers' instructional practices; and four surveys – along with three focus groups – to obtain feedback on LLI and the core literacy program from LLI and classroom teachers, parents/guardians, principals, district literacy specialists, and on-site researchers. Details of each instrument will be discussed later in this section.

System Description: Leveled Literacy Intervention (LLI)

The Leveled Literacy Intervention system (Fountas & Pinnell, 2008) is a short-term, intensive, small-group intervention designed for children in kindergarten through second grade who are having difficulty learning early reading and writing skills. The goal of the system is to accelerate these children's progress in order to bring their skills up to grade level so their early literacy difficulties do not become long-term deficits. The system is appropriate for struggling regular education students and students with special needs, and there are minor modifications for English language learners (ELL students).

Children enrolled in LLI meet in small groups (ideally three students) for daily 30-minute lessons, and the intervention lasts a maximum of 18 weeks, depending on the progress of the individual child. According to the developers (Heinemann, 2008), LLI emphasizes the development of oral language skills as a foundation for reading and the five components of reading instruction identified by the National Reading Panel (National Institute of Child Health and Human Development, 2000a, 2000b): phonological awareness and phonics, fluency, vocabulary, and comprehension. Phonics instruction is systematic, explicit, and follows a prescribed sequence of sound-letter relationships and spelling patterns. Additionally, reading comprehension skills are taught through intensive interactions between the teacher and the students and amongst students. LLI also is designed to develop students' motivation and interest in reading and writing.

An underlying premise of LLI is that children benefit from experience with texts that they can read without difficulty at their "independent level," as well as with more challenging texts written at their "instructional level" (Heinemann, 2008). The LLI system provides students with both kinds of reading experiences, alternating between easier texts and more challenging ones. Easier texts build fluency and give students success at reading that builds confidence and positive self-esteem. More challenging texts, which students read with scaffolding and support from the LLI teacher, give children the opportunity to develop more sophisticated reading skills. LLI materials specify concepts that teachers can emphasize when discussing each book in the sequence. Other key ideas underlying the design of LLI are the following:

- Struggling children learn best when lessons follow a predictable sequence. All LLI lessons have the same basic structure, allowing children to focus most of their processing attention on reading, writing, phonics, and word study activities.

- Children who are struggling with reading and writing need to learn fast, automatic processing of oral and written language. For this reason, LLI lessons are designed to be fast-paced, with a specified set of literacy activities for each day of the intervention. The fast pace promotes rapid processing and keeps children engaged in the lessons and motivated to participate in the literacy activities and discussion.

- Literacy interventions should be linked to classroom instruction and the home environment. Children take LLI books home to read aloud to their parents, along with simple homework assignments, and they also may take books back to the classroom.

- A system of ongoing formative assessments conducted during the 18 weeks gives teachers information about student learning that can inform their instructional decision-making.

Literacy teachers selected to be LLI teachers receive eight days of professional development focused on how to implement the LLI instructional program. They also receive the necessary LLI materials and a detailed teaching guide. Additional professional development is provided throughout implementation, including training on how best to facilitate comprehension skills through teacher-student and student-student interactions.

In addition to the professional development regarding LLI materials and instructional strategies, DPS educators also received access to the LLI online data management system, used to track student progress and attendance. Further, for the purposes of the study, the district voluntarily agreed to provide the LLI system as specifically designed by the developers. The following guidelines for implementation, which were jointly agreed upon by the researchers, developers, and district, were utilized:

- Treatment group students could not receive any supplemental literacy interventions other than LLI for the duration of the study. Control group students could only receive such services if their total amount of literacy instructional time did not exceed the amount for treatment group students; however, they could not receive LLI for the duration of the study. (Example: If treatment group students at a particular school received two hours of core literacy instruction and half an hour of LLI instruction, control group students at that school could also receive half an hour of non-LLI supplemental instruction – e.g., the Voyager Passport intervention – in addition to their two hours of core literacy instruction.)

- The schools attempted to provide the maximum amount of LLI instructional time to treatment group students (i.e., 18 weeks for first and second graders and 14 weeks for kindergarteners, according to LLI program recommendations).

- The schools agreed to limit the LLI groups containing treatment group students to three students per group, according to LLI program recommendations. (In order to serve more students, several schools had adopted the practice of offering four-student groups.)

- The district agreed to a designated start and end date for all LLI groups containing treatment group students across all 13 schools participating in the study. These dates reflected, as closely as possible, the recommended amount of LLI instructional time mentioned above.

Setting and Population of Participants

Thirteen elementary or K-8 schools in Denver Public Schools (DPS) in Denver, Colorado, volunteered to participate in the study. DPS is a large urban district that served 81,870 students and employed 4,555 teachers during the 2011-2012 school year. There are 162 schools in the district, including 73 elementary schools and 16 K-8 schools. The majority of students are Hispanic or White (58.0% and 20.3%, respectively), with nearly three-quarters of students eligible for free or reduced lunch (72.5%). Additionally, a third of students are English Language Learners (33.9%), and nearly two-fifths are Spanish speaking (37.9%); other languages most frequently spoken by students include Vietnamese, Arabic, Karen/Burmese, and Somali. Table 1 summarizes the overall demographic characteristics of the district.

Table 1: Demographic Overview of DPS Schools for the 2011-2012 School Year

District-Wide Population				District-Wide Student Demographics							
Grade Levels	Students	Teachers	Student/ Teacher Ratio	% Asian	% Black	% Hispanic	% White	% Other	% Free/Reduced Lunch	% Special Education	% English Language Learners
ECE-12	81,870	4,555	18.0	3.3	14.5	58.0	20.3	3.7	72.5	10.9*	33.9

Source: DPS Communications Office website (http://communications.dpsk12.org/newsroom/73/55/)
*This information was obtained from the DPS Department of Accountability, Research and Evaluation's 2011 School Performance Framework (http://communications.dpsk12.org/initiatives/school-performance-framework/).

Teacher Demographics

A total of 25 LLI teachers and 78 classroom teachers participated in this study. According to data obtained from a survey of participating LLI teachers (n = 21), the majority of LLI teachers in the study taught first grade (85.7%), followed by second grade (76.2%) and then kindergarten (71.4%). Over half of the teachers had been at their current school for one to five years (57.1%) and had been teaching for 11 or more years (52.4%). In addition, the majority of participating LLI teachers had obtained a Master's or Master's plus 30 degree (71.4%), had acquired a regular/professional teaching certificate (85.7%), and were fully trained in LLI (85.7%). Almost all of the teachers were female (95.2%), and most of them were White (85.7%). Overall, these teachers had a solid background of teaching experience at their current school and teaching in general, and nearly three-quarters of them had pursued advanced degrees and continuing education in their field. Taken together, they appear to have been well positioned to implement the LLI curriculum. Table 2 summarizes the demographic characteristics of the LLI teachers in the study, as reported on the LLI teacher survey.

Table 2: Demographic Characteristics of Participating LLI Teachers (*n* = 21)

Item	Percent Responded
Grade level(s) taught	
Kindergarten	71.4
First	85.7
Second	76.2
Years of teaching experience at current school	
Less than 1 year	19.0
1-5 years	57.1
6-10 years	19.0
11-15 years	4.8
More than 15 years	0.0
Years of teaching experience at any school	
Less than 1 year	0.0
1-5 years	28.6
6-10 years	19.0
11-15 years	28.6
More than 15 years	23.8
Highest level of education completed	
Bachelor's degree	23.8
Master's degree	38.1
Master's plus 30 hours	33.3
Education Specialist degree	4.8
Doctoral degree	0.0
Ethnicity	
Asian or Pacific Islander	4.8
American Indian or Alaskan Native	0.0
African-American/Black	0.0
Hispanic	9.5
White, not of Hispanic origin	85.7
Multi-racial/Other	0.0
Gender	
Male	0.0
Female	95.2
Age group	
29 years or less	14.3
30-39 years	19.0
40-49 years	42.9
50-59 years	9.5
60 years or older	14.3
Level of LLI training	
Completed training	85.7
Partially trained	9.5
None	4.8
Teacher certification level	
Paraprofessional	4.8
Alternative certificate	0.0
Initial/apprentice certificate	9.5
Regular/professional certificate	85.7

Note: Item percentages may not total 100% due to missing input or multiple responses from some participants.

According to data obtained from a survey of 51 of the 78 participating classroom teachers, the classroom teachers in the current study primarily taught kindergarten (39.2%), followed by first (35.5%) and second grades (25.5%). Similar to the LLI teachers, approximately half of the classroom teachers had been at their current school between one and five years (47.1%) and had been teaching for at least 11 years (49.1%). Further, over half had obtained a Master's or Master's plus 30 degree (56.8%), and the majority held a regular/professional teaching certificate (88.2%). Almost all of the participating classroom teachers were female (98.0%), and the majority were White (76.5%), followed by Hispanic (15.7%). Overall, the participating classroom teachers were generally well-qualified and had a substantial amount of teaching experience. Table 3 summarizes the demographic characteristics of the K-2 classroom teachers in the study, as reported on the classroom teacher survey.

Table 3: Demographic Characteristics of Participating K-2 Classroom Teachers (n = 51)

Item	Percent Responded
Grade level(s) taught	
Kindergarten	39.2
First	35.3
Second	25.5
Years of teaching experience at current school	
Less than 1 year	11.8
1-5 years	47.1
6-10 years	21.6
11-15 years	7.8
More than 15 years	5.9
Years of teaching experience at any school	
Less than 1 year	7.8
1-5 years	15.7
6-10 years	25.5
11-15 years	21.6
More than 15 years	27.5
Highest level of education completed	
Bachelor's degree	35.3
Master's degree	33.3
Master's plus 30 hours	23.5
Education Specialist degree	5.9
Doctoral degree	0.0
Ethnicity	
Asian or Pacific Islander	0.0
American Indian or Alaskan Native	2.0
African-American/Black	0.0
Hispanic	15.7
White, not of Hispanic origin	76.5
Multi-racial/Other	3.9
Gender	
Male	0.0
Female	98.0
Age group	
29 years or less	17.6
30-39 years	25.5
40-49 years	33.3
50-59 years	15.7
60 years or older	5.9

Table 3, continued

Item	Percent Responded
Teacher certification level	
Paraprofessional	0.0
Alternative certificate	3.9
Initial/apprentice certificate	3.9
Regular/professional certificate	88.2

Note: Item percentages may not total 100% due to missing input or multiple responses from some participants.

Student Demographics

Across the 13 participating schools in DPS, there were a total of 320 students who participated in the RCT study. Of these students, 110 were in kindergarten, 140 were in first grade, and 70 were in second grade. A total of 163 students comprised the randomly assigned treatment group for the study, while 157 students made up the control group; the slight discrepancy in group size is attributable to student attrition that occurred after randomization was completed. On average across the participating schools, 69.4% of students in the sample were Hispanic, 14.7% were White, 7.5% were Black, 4.4% were Asian, and 4.1% were another or mixed ethnicity. Over half of the students were male (53.1%), while slightly less than half were female (46.9%). In addition, over a third of the participating students were English Language Learners (34.1%), and 9.4% had a special education designation. Finally, in 11 of the 13 participating schools, the majority of students (between 72.3% and 97.1%) qualified for free or reduced lunch, while only half of the students at another school qualified for this program (47.9%). Table 4 summarizes the demographic characteristics of the participating K-2 students in DPS.

Table 4: Demographic Characteristics of Participating K-2 Students (n = 320)

Grade Levels	Students	% American Indian	% Asian	% Black	% Hispanic	% White	% Multi/Other	% Male	% Female	% English Language Learners	% Special Education
K-2	320	1.6	4.4	7.5	69.4	14.7	2.5	53.1	46.9	34.1	9.4

Source: Student records provided by the DPS Department of Accountability, Research and Evaluation

Comparison Group

In addition to the 320 students who participated in the RCT, DPS provided CREP with demographic and achievement (i.e., DRA2 and STAR Early Literacy) data for an additional 386 K-2 students who received LLI during the 2011-2012 school year but did not participate in the RCT. Reasons for non-participation may have included such factors as attending a DPS school that did not volunteer to participate in the RCT, not meeting the eligibility criteria to participate in the RCT, or not receiving parental consent to participate in the RCT. These students were not matched and randomly assigned to treatment and control groups, and they may have received LLI under various conditions (e.g., as part of their ELL services, in a group of more than three students, in a classroom setting, etc.). Also, different from the students in the RCT, they may have received additional instructional time and additional literacy support or intervention services as needed. However, a comparison of these students' achievement scores before and after receiving LLI provided a general measure of their growth in literacy and was used to supplement the findings from the RCT, which had a relatively small sample size due to

district limitations. The demographic characteristics of the comparison group students are summarized in Table 5, below.

Table 5: Demographic Characteristics of Comparison Group Students (n = 386)

Grade Levels	Students	% American Indian	% Asian	% Black	% Hispanic	% White	% Multi/Other	% Male	% Female	% English Language Learners	% Special Education
K-2	386	0.3	2.8	11.9	73.6	9.8	1.6	50.3	49.7	47.7	15.5

Source: Student records provided by the DPS Department of Accountability, Research and Evaluation

Instrumentation

Both quantitative and qualitative data were collected in this evaluation. CREP researchers used three measures of reading achievement for evaluating students' progress in literacy: the Fountas & Pinnell Benchmark Assessment System was administered by local on-site researchers trained by CREP, while the Developmental Reading Assessment, 2nd edition (DRA2) and the STAR Early Literacy Assessment are both routinely administered by the district. Two observational tools developed by CREP – the Leveled Literacy Intervention Observation Tool (LLIOT) and the Literacy Observation Tool (LOT) – were used to evaluate LLI and classroom teachers' literacy practices and instructional strategies in the classroom. CREP also developed two teacher surveys, the Leveled Literacy Intervention Teacher Questionnaire – Revised (LLITQ-R) and the Classroom Teacher Literacy Instruction Questionnaire (CTLIQ), as well as a principal survey (the Leveled Literacy Intervention Principal Questionnaire, or LLIPQ) and a parent/guardian survey (the Home Literacy Support Questionnaire, or HLSQ), to ascertain these stakeholders' feedback on LLI and core literacy classroom instruction. Finally, structured focus groups were conducted with district literacy specialists, LLI teachers, and on-site researchers to gather additional qualitative feedback regarding LLI. Details of each instrument are discussed below.

Student Literacy Achievement

Fountas & Pinnell Benchmark Assessment System

The Fountas & Pinnell Benchmark Assessment System 1, 2nd Edition (2010) was used to measure the following literacy skills: phonemic awareness, letter-sound relationships (decoding), vocabulary, comprehension, fluency, and writing. Both treatment and control group students in the study were tested by independent on-site researchers at the beginning and end of LLI. This data was used to measure individual student gains as well as the composition of the groups in respect to homogeneity of student needs.

The Fountas & Pinnell Benchmark Assessment System (BAS) is an individually administered assessment tool designed by the developers of LLI to reliably place students on the Fountas & Pinnell Text Level Gradient™ (Fountas & Pinnell, 2007), an A-Z gradient of text difficulty. LLI is comprised of four systems: Levels A-C are in the Orange System; Levels A-J are in the Green System; Levels C-N are in the Blue System; and Levels L-Q are in the Red System. The Orange System is generally used in kindergarten, the Green System in first grade, the Blue System in second grade, and the Red System in third grade. The goal of the LLI system is to bring children up to their current grade level in reading, starting from the earliest Level A (mid-kindergarten) to Level Q (early fourth grade). System 1 of the BAS, which is designed for grades K-2 (Levels A-N), uses both fiction and nonfiction texts to determine

an independent and an instructional reading level for the student. The BAS demonstrates high test-retest reliability (0.93 for System 1; 0.97 overall), and convergent validity was established between the reading accuracy rates of BAS System 1 books and those of Reading Recovery assessment texts (0.94 for fiction, 0.93 for nonfiction; Heinemann, 2007).

Developmental Reading Assessment, 2nd Edition (DRA2)

The Developmental Reading Assessment, 2nd Edition (DRA2) for grades K-3 was used to corroborate the Fountas & Pinnell benchmarks and provide a distal measure of student literacy achievement, since it is not as closely aligned with the LLI curriculum as the benchmarks. The DRA2 is an individually administered assessment tool designed to determine a student's reading level on a continuum from A to 40. The DRA2 evaluates three components of reading (reading engagement, oral reading fluency, and comprehension) and is designed for use by classroom teachers to connect assessment information with individualized instruction. Its levels have been correlated with the Fountas & Pinnell Text Level Gradient™ (Fountas & Pinnell, 2007), and it demonstrates moderate to high reliability (i.e., internal consistency, parallel equivalency, test-retest, and inter-rater reliability) and validity (i.e., content-related, criterion-related, and construct validity; Beaver & Carter, 2009). Trained classroom teachers administer the DRA2 each spring to all K-2 students in DPS. Fall DRA2 testing is also conducted in first and second grades for new students or students with individual learning plans (ILPs), while midyear testing is conducted for kindergarteners (who also complete DRA2 Word Analysis Tasks 1-7 in the fall).

STAR Early Literacy Assessment

The STAR Early Literacy Assessment is a computer-adaptive diagnostic reading assessment for students in prekindergarten through third grade. STAR Early Literacy is very brief (25 items that take 10 minutes or less to complete) and is self-administered by computer. The seven domains measured include general readiness, graphophonemic knowledge, phonemic awareness, comprehension, phonics, vocabulary, and structural analysis. STAR Early Literacy is designed for repeated administration throughout the school year in order to provide teachers with immediate feedback regarding student understanding of literacy concepts and enable them to target literacy instruction to student needs. Thus, it can easily be used for individualized goal setting, progress monitoring, and outcome assessment. STAR Early Literacy has been shown to have a split-half reliability of 0.91, a test-retest reliability of 0.86, and an overall validity estimate of 0.60, with a standard error of 0.02 and a 95% confidence interval ranging from 0.57 to 0.62; additionally, the assessment was normed using a nationally representative sample of 9,038 students from 50 states (Renaissance Learning, 2009). STAR Early Literacy, which is administered by classroom teachers in DPS up to four times per year, was used in the current study as an additional distal measure of student literacy achievement.

Intervention Fidelity

Leveled Literacy Intervention Observation Tool (LLIOT)

The Leveled Literacy Intervention Observation Tool (LLIOT), developed by CREP researchers for a previous evaluation of LLI, involves a targeted, 30-minute observation of an LLI group completing a randomly selected LLI lesson. The LLIOT is used to rate LLI teachers' fidelity to the LLI model as well as the quality of their literacy instructional strategies and the learning environment of the lesson. Ratings are provided using a 4-point scale that ranges from 0 (Not Observed) to 3 (Excellent). Containing 23

items, the LLIOT is comprised of three subscales: Quality of LLI Implementation (10 items; α = .83), which is designed to measure LLI teachers' implementation of the 10 main LLI lesson components; Literacy Instructional Strategies (6 items; α = .90), which is designed to assess LLI teachers' use of general teaching strategies that should be present in a successful literacy intervention; and Learning Environment (7 items; α = .83), which is designed to assess the quality of lesson factors such as organization, pacing, and the availability of materials.

On-site researchers trained by CREP conducted observations of two intervention sessions with each participating LLI group, one near the beginning of LLI and one near the end, using the LLIOT. This observation data contributed to the evaluation of fidelity to the LLI model and to gauge the level of literacy instruction provided in these groups. To ensure the reliability of data, observers received a manual which provided definitions of terms, examples and explanations of target strategies, and a description of procedures for completing the instruments. Observers also received training on the instrument in a group session and monitoring by CREP researchers throughout the observations. A copy of the LLIOT can be found in Appendix A.

LLI Online Data Management System Intervention Record

The LLI Online Data Management System (ODMS) is a tool developed by Heinemann to allow teachers to enter and track data for their LLI groups and individual students, including demographic information, entry and exit benchmark scores, Weekly Reading Record scores, attendance, lessons completed, and current reading level. This data management tool allows teachers and administrators to create individual, group, or school-level reports to monitor students' progress. The Intervention Record in the ODMS was used for tracking student and teacher attendance, reasons for absence, student reading selections, and achievement levels. CREP utilized these intervention records to provide an additional measure of the LLI implementation fidelity at each school, particularly with regard to the 30-minutes-a-day, 5-days-a-week instructional cycle.

Quality of Core Literacy Instruction

Literacy Observation Tool (LOT)

The Literacy Observation Tool (LOT) was developed by researchers at CREP to serve as an instrument for observing in elementary classrooms where teachers are engaged in teaching reading and other literacy-related practices. The LOT has been aligned to the National Reading Panel and National Research Council findings. It captures explicit instruction in the five essential components of reading identified by the National Reading Panel as important in achieving effective reading instruction: Phonemic Awareness, Phonics, Fluency, Text Comprehension, and Vocabulary. Standard use of the LOT involves multiple classroom observations during a designated literacy block (typically 1.5 to 2 hours), with seven to nine classrooms each observed for 10 minutes. In a study of 70 schools across Tennessee, strong evidence was established for the reliability of the LOT, with a phi coefficient of .75 for five observations and .82 for eight observations at a school (Sterbinsky & Ross, 2003).

Twice during the 2011-2012 school year (once at the beginning and once at the end), on-site researchers trained by CREP conducted a set of seven to nine 10-minute LOT observations in the regular K-2 classrooms at each participating school. Each set of observations was conducted in one day during the school's literacy block, and the ratings from the seven to nine individual classroom observations were combined to form a single LOT composite for that school. Therefore, the LOT was used to obtain a

measure of the quality of the regular classroom literacy instruction received by students in the study by taking a "snapshot" of each school's core literacy instruction. To ensure that the identifying and coding of literacy instructional variables occurred in a consistent manner, observers received formal training, user's manuals, and monitoring by CREP researchers. A copy of the LOT can be found in Appendix B.

School and Home Support for Literacy

Leveled Literacy Intervention Teacher Questionnaire – Revised (LLITQ-R)

The Leveled Literacy Intervention Teacher Questionnaire – Revised (LLITQ-R), developed by CREP researchers for a previous evaluation of LLI, was used in this study as a measure of the participating LLI teachers' views of the efficacy of LLI, their implementation of the LLI model, and their students' progress in literacy, as well as the overall support for literacy and LLI in their schools. The LLITQ-R consists of 26 items on a 5-point scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), five items on a 4-point scale ranging from 0 (Not At All) to 3 (Extensively), five items on a 5-point scale ranging from 0 (Not At All/Never) to 4 (Regularly/Every Day), and three open-ended items regarding LLI's strengths and areas for improvement as well as reasons to continue or not continue using the LLI system. The LLITQ-R was administered to participating LLI teachers at the end of the school year. A copy of the LLITQ-R can be found in Appendix C.

Classroom Teacher Literacy Instruction Questionnaire (CTLIQ)

The Classroom Teacher Literacy Instruction Questionnaire (CTLIQ), also developed by CREP for a previous evaluation of LLI, was used in the current study to measure the overall support for literacy in the participating schools and the nature of the regular classroom literacy instruction received by the students in the study. The CTLIQ assessed K-2 classroom teachers' self-reported literacy instructional practices and their perceptions of the core literacy program at their schools, as well as their perceptions of LLI. The CTLIQ consists of 24 items on a 5-point scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), five items on a 4-point scale ranging from 0 (Not At All) to 3 (Extensively), 10 items on a 5-point scale ranging from 0 (Not At All/Never) to 4 (Regularly/Every Day), and 3 open-ended items regarding the core literacy program's strengths and areas for improvement as well as reasons to continue or not continue the program. The CTLIQ was administered to K-2 classroom teachers at the participating schools at the end of the school year. A copy of the CTLIQ can be found in Appendix D.

Leveled Literacy Intervention Principal Questionnaire (LLIPQ)

CREP researchers developed the Leveled Literacy Intervention Principal Questionnaire (LLIPQ) for the current study as a measure of school support for literacy at the administrative level as well as principals' support for LLI specifically. The LLIPQ assessed principals' perceptions of their schools' core literacy program, their understanding of and familiarity with the LLI system, and their perceptions of LLI's implementation and efficacy at their schools. The LLIPQ consists of 26 items on a 5-point scale ranging from 0 (Strongly Disagree) to 4 (Strongly Agree) and three items on a 4-point scale ranging from 0 (Not At All) to 3 (Extensively). In addition, the survey contains five open-ended items designed to ascertain principals' perceptions of the strengths and areas for improvement of the LLI system, reasons to continue or not continue using the system, challenges to LLI implementation, additional resources needed, and efficacy of LLI in comparison with other supplemental literacy interventions at the principals' schools. The LLIPQ was administered to principals of the schools participating in the study at the end of the school year. A copy of the LLIPQ can be found in Appendix E.

Home Literacy Support Questionnaire (HLSQ)

In order to measure the amount of support for literacy received by participating students at home, CREP researchers developed the Home Literacy Support Questionnaire (HLSQ). The HLSQ is a brief survey that asks parents/guardians (or other caretaking family members) about literacy activities in which their child may engage at home, as well as their own involvement in and encouragement of these activities. Additionally, the HLSQ assesses respondents' feedback on LLI, if applicable. The HLSQ is comprised of 15 items on a 5-point scale ranging from 0 (Strongly Disagree) to 4 (Strongly Agree) and two open-ended items regarding respondents' perceptions of the strengths and areas for improvement of their child's literacy instruction at school. The HLSQ was administered to parents/guardians of both treatment and control group students in the study at the end of the school year. A copy of the HLSQ can be found in Appendix F.

Stakeholder Feedback

District Literacy Specialist, LLI Teacher, and On-Site Researcher Focus Groups

Voluntary, structured focus groups were conducted with district-level literacy specialists, LLI teachers, and on-site researchers at the end of the school year to provide more information about the overall climate for literacy instruction in the district as well as additional feedback on the LLI system. DPS literacy specialists, including LLI coordinators, discussed LLI's implementation and impact throughout the district as well as strengths, areas for improvement, challenges to LLI implementation, additional resources and success factors needed, support within the district for LLI, and LLI's efficacy in comparison with other district literacy initiatives – particularly other Tier II interventions. Participating LLI teachers responded to questions regarding LLI's strengths and areas for improvement, the efficacy of LLI in meeting students' needs, support within their schools for LLI, and their opinion of the LLI professional development they received. Finally, on-site researchers discussed their perceptions of LLI's strengths and areas for improvement, students' response to LLI, the quality of LLI and core literacy instruction received by students in the study, and their opinion of the training and use of the data collection instruments for the study. The focus group protocols can be found in Appendices G through I.

Instrumentation Summary

Table 6 summarizes each of the research questions and the participants and provides the data sources and methodology used to investigate each question.

Table 6: Summary of Data Sources and Participants by Research Question

Research Questions	Participants	Data Sources	Method
1) What progress in literacy achievement, if any, do urban students who receive LLI make compared to students who receive core literacy instruction alone?	• LLI treatment and control students	• Fountas & Pinnell Benchmarks • DRA2 • STAR Early Literacy	• Quantitative assessments of student progress in reading achievement
a) Does the effectiveness of LLI vary by the following subgroups: English Language Learners, students with a special education designation, and ethnic minorities (i.e., African-American and Hispanic students)?	• LLI and classroom teachers	• LLI teacher survey (LLITQ-R) • LLI teacher focus group • Classroom teacher survey (CTLIQ) • Classroom literacy observations (LOT)	• Qualitative assessment of student progress through teacher and administrator feedback • Quantitative data regarding regular classroom literacy instruction to establish a baseline for student progress
	• Principals and district literacy specialists	• Principal survey (LLIPQ) • District literacy specialist focus group	
2) At what level of fidelity to the program model is LLI implemented by teachers participating in the study?	• LLI teachers	• LLI observations (LLIOT) • LLI Online Data Management System Intervention Records • LLI teacher survey (LLITQ-R) • LLI teacher focus groups	• Quantitative and qualitative assessments of LLI instructional strategies and delivery
	• On-site researchers	• On-site researcher focus groups	
	• District literacy specialists	• District literacy specialist focus group	
3) What are stakeholders' perceptions of the LLI system and the core literacy program?	• LLI teachers	• LLI teacher survey (LLITQ-R) • LLI teacher focus groups	• Quantitative and qualitative assessment of LLI teachers' perceptions regarding LLI's impact on their instruction and their students' literacy, as well as classroom teachers', parents'/guardians', principals', and district literacy specialists' perceptions of LLI and the core literacy program in general
	• Classroom teachers	• Classroom teacher survey (CTLIQ)	
	• Parents/guardians	• Parent/guardian survey (HLSQ)	
	• Principals	• Principal survey (LLIPQ)	
	• District literacy specialists	• District literacy specialist focus group	
4) How do the results of the Fountas & Pinnell Benchmark Assessment System compare to those of the Developmental Reading Assessment, 2nd Edition (DRA2) and the STAR Early Literacy Assessment?	• LLI treatment and control students	• Fountas & Pinnell Benchmarks • DRA2 • STAR Early Literacy	• Quantitative comparison of assessment scores

Procedure

The current study extended from January 2011 through July 2012. In the winter and spring of 2011, CREP researchers worked with the sponsor to develop an evaluation plan and select a school district as a site for the evaluation. DPS volunteered to participate and was chosen due to its urban location and student population (i.e., a high percentage of ELL, minority, and economically disadvantaged students), its ability to comply with the study protocol, and its established relationship with the sponsor and with LLI. DPS agreed to allow randomization of eligible students to treatment and control groups, deliver LLI as designed by the developers, allow the Fountas & Pinnell benchmarks to be administered to students in the study, and provide the researchers with individual student-level data (e.g., demographic information, DRA2 and STAR Early Literacy scores).

In May 2011, CREP researchers conducted a site visit in Denver to meet with key district-level administrators as well as principals and teachers at schools interested in participating in the study. The research team provided an overview of the study requirements and the incentives to participate, which included – for each school – a set of Fountas & Pinnell Benchmark Assessment System 1 kits and Heinemann professional development on the benchmarks, a $500 stipend, a $500 gift certificate for Heinemann professional resources, complimentary use of the LLI Online Data Management System, and a school-level report of results. Although DPS had agreed to take part in the study, the district decided to allow schools to participate on a voluntary basis; therefore, following these initial meetings, a total of 13 schools elected to participate.

During the summer of 2011, CREP researchers developed and refined the data collection instruments, obtained IRB approval for the study, and worked with DPS staff to identify school coordinators and on-site researchers to conduct ongoing local work for the study. One school coordinator was identified from the team of LLI teachers at each participating school to coordinate data collection activities with CREP and help ensure smooth LLI implementation. Additionally, on-site researcher applicants were selected from a pool of local-area educators, primarily retired DPS teachers; these applicants formed a team of 18 on-site researchers to collect data for the study (i.e., Fountas & Pinnell benchmarks, LOT and LLIOT observations) throughout the school year. Finally, a recently retired DPS LLI coordinator agreed to serve as a district-level liaison for the study.

During the first full week of school, in August 2011, CREP researchers returned to Denver to meet with district personnel and teachers and finalize the study timeline and logistics, including plans for pretesting the first and second graders on the Fountas & Pinnell benchmarks in late August and early September (kindergarteners were tested and received LLI in winter and spring 2012). The CREP research team also worked with Heinemann consultants and the district professional development team to provide four days of training to the on-site researchers and LLI teachers who would be participating in the study, which included the following: (1) an LLI refresher training for experienced LLI teachers as well as the on-site researchers, to familiarize them with the LLI system; (2) two full days of LLI training for new teachers (a follow-up training day was conducted in late October); (3) a full-day training on the Fountas & Pinnell Benchmark Assessment System for LLI teachers and on-site researchers; and (4) training for on-site researchers on the LLI and classroom observation tools (i.e., LLIOT and LOT). Two Heinemann consultants conducted the LLI and benchmark trainings, while four CREP researchers conducted the observation training.

At this time, the CREP research team also worked with the district coordinator and school coordinators to develop a list of first and second grade students eligible to participate in the study

(kindergarteners were identified in winter 2012). Selection criteria included students who would be able to receive delayed literacy intervention services if assigned to the control group, could receive instruction in English, were not known to demonstrate high absenteeism, were below grade level on their Spring 2011 DRA2 independent reading level (i.e., Level 3 or below for first graders, Level 14 or below for second graders), and – in certain schools – were not eligible for Read to Achieve (RTA) services (i.e., scoring in the "Strategic" range on the Dynamic Indicators of Basic Early Literacy Skills [DIBELS] assessment), as these students were required to receive literacy services immediately as part of the RTA state grant program.[1] CREP researchers also worked with the district coordinator and school coordinators to obtain active parental consent from students eligible to participate in the study, as well as from LLI and classroom teachers who would be taking part in the study. In total, consent was obtained for 25 LLI teachers, 78 K-2 classroom teachers, and 1,018 K-2 students, including all 320 students who completed the study.

Once eligible students were identified and parental consent was received, pretesting of these students with the Fountas & Pinnell benchmarks was conducted by the on-site researchers and LLI teachers. Paired assessments were also conducted to establish inter-rater reliability on the benchmarks for the study. Subsequently, CREP conducted the randomization of the matched pairs of first and second graders based on demographic characteristics (i.e., gender, ethnicity, ELL status, and special education status), Spring 2011 DRA2 scores, and pretest Fountas & Pinnell benchmark scores of instructional reading level. Students in the treatment group were then placed in LLI groups by LLI teachers, and the planned 18 weeks of LLI instruction for first and second graders began. Although the established LLI start date for the study was September 12, the start date varied across the participating schools due to the length of time needed to identify eligible students, obtain consent, administer the benchmarks, complete randomization, and organize and schedule the LLI groups; however, all groups began LLI before September 22.

Consistent with LLI program recommendations, the first and second grade LLI evaluation period lasted 18 calendar weeks, excluding the two weeks DPS was out of session for the winter holidays. During this 18-week period, control students did not receive LLI; however, they could receive it after the evaluation period if they still needed it (e.g., according to teacher judgment or post-benchmark scores). On-site researchers used the LLIOT to conduct two random observations of each first and second grade LLI group, one in November 2011 and one in January 2012; additionally, they conducted the first set of K-2 LOT (classroom literacy) observations in late October 2011. Both CREP researchers and on-site researchers conducted paired observations using the LLIOT during the November round of observations, in order to allow an assessment of inter-rater reliability. Posttests with the Fountas & Pinnell benchmarks for the first and second grade students were completed in February 2012, following the LLI end date of February 3.

CREP researchers, along with a Heinemann consultant, returned to DPS in January 2012 to provide booster trainings on the Fountas & Pinnell benchmarks (for all LLI teachers in the study and on-site researchers), the LOT and LLIOT observations (for on-site researchers), and the LLI Orange Kit (for all LLI teachers who would teach kindergarten groups in the spring). These refresher trainings occurred prior to the second round of LLIOT and LOT observations, the benchmark post-testing for first and

[1] Out of the 13 schools participating in the study, eight were awarded the RTA grant in August 2011, resulting in a smaller sample size for the study than anticipated (i.e., 600 students). In these eight schools, approximately 176 students who were eligible for the study could not participate due to issues related to RTA (i.e., they qualified for RTA services and could not be randomized to the control group due to the need to receive services immediately, or the LLI teachers had limited capacity to serve non-RTA students due to the need to prioritize RTA student services).

second grade, the benchmark pre-testing for kindergarteners, and the implementation of kindergarten LLI. Also in January 2012, CREP worked with the school and district coordinators to identify kindergarteners who were eligible to participate in the study and to obtain parental consent for these students. The kindergarten selection criteria were identical to the first and second grade selection criteria, except kindergarteners did not have Spring 2011 DRA2 scores; therefore, to qualify for the study, kindergarteners were required to score in the "No/Little Control" or "Some Control" ranges on three of the seven DRA2 Word Analysis Tasks administered by DPS in fall 2011.

Pretesting of the identified kindergarten students with the Fountas & Pinnell benchmarks began in late January 2012. Subsequently, these students were randomly assigned to treatment or control groups using the same randomization procedure that was utilized for first and second grade. Kindergartners in the treatment group received LLI beginning February 13. (It was necessary for the first and second grade LLI evaluation period to conclude before the kindergarten evaluation period began, in order for the participating LLI teachers to have the capacity to serve kindergarten students.) The kindergarten LLI evaluation period lasted 12 calendar weeks; it was not possible to provide the recommended 14 weeks due to the fact that the evaluation period needed to end two weeks prior to the last day of school in order to allow time for post-benchmarking. As with the first and second grade groups, two random LLIOT observations were conducted for each kindergarten LLI group, one in March 2012 and one in late April or early May 2012. The second round of K-2 LOT observations was also conducted between late April and early May 2012. Posttesting of kindergarteners on the Fountas & Pinnell benchmarks was conducted in mid-May 2012, following the LLI end date of May 11.

At the end of the school year, in May 2012, the school and district coordinators were asked to encourage all K-2 LLI and classroom teachers with students in the study to complete the online LLI Teacher Questionnaire – Revised (LLITQ-R) or Classroom Teacher Literacy Instruction Questionnaire (CTLIQ), as applicable. Additionally, the principals at the participating schools were asked to complete the online LLI Principal Questionnaire (LLIPQ). CREP assisted in the online survey process by providing instructions and log-in information to all participants. CREP also distributed paper copies of the Home Literacy Support Questionnaire (HLSQ) to the school coordinators, who sent them home with both treatment and control group students. Parents/guardians could complete the survey and return it to the child's school, where it was collected by the school coordinator, or directly to CREP via mail.

Also in May 2012, the CREP research team held end-of-year meetings with DPS district literacy specialists, participating LLI teachers, and on-site researchers to debrief them, discuss any remaining issues, and conduct structured focus groups. The purpose of the focus groups was to collect qualitative data related to the study, the LLI materials, and participants' individual and collective views of LLI. CREP researchers also met individually with LLI teachers, as needed, to verify teachers' group and student data taken from the LLI Online Data Management System and to request missing information. Finally, CREP researchers met with district-level administrators and representatives from the DPS Department of Accountability, Research and Evaluation to discuss obtaining pre and post DRA2 and STAR Early Literacy scores for treatment and control group students, as well as comparison group data for students who did not participate in the RCT but received LLI during the 2011-2012 school year. Because the sample size for the study was smaller than expected due to the participating schools' compliance with RTA grant stipulations, CREP researchers requested this comparison group data in order to examine the general literacy growth of LLI students and supplement the RCT findings. Demographic and achievement data (i.e., spring 2011 through spring 2012 DRA2 and STAR Early Literacy scores) for a total of 706 students (163 treatment, 157 control, 386 comparison) were received electronically from DPS in July 2012.

Table 7 provides a summary of data collection procedures, including the instruments organized by type, a general timeline and description of the data collection process, and the number received for each instrument.

Table 7: Data Collection Summary

Type of Measure	Instrument	Timeline	Number Collected	Description
Student Achievement Measures	• Fountas & Pinnell Benchmarks	• August-September 2011 (grades 1-2) • January-February 2012 (K-2) • May 2012 (K)	• 140 1st grade pre/post • 70 2nd grade pre/post • 110 K pre/post	• The Fountas & Pinnell Benchmarks were administered to 1st and 2nd graders in both treatment and control groups as a pretest in fall 2011 and a posttest in Winter 2012. This assessment was administered to kindergarteners in both treatment and control groups as a pretest in Winter 2012 and a posttest in Spring 2012.
	• DRA2	• Spring 2011 and Spring 2012 (grades 1-2) • Winter 2012 and Spring 2012 (K)	• 116 1st grade pre/post • 60 2nd grade pre/post • 102 K pre/post	• Scores from the Spring 2011 and Spring 2012 DPS administrations of the DRA2 were collected as pre- and posttests, respectively, for the 1st and 2nd graders in the study for whom scores were available. Scores from Winter 2012 and Spring 2012 were collected as pre- and posttests, respectively, for kindergarteners in the study for whom scores were available.*
	• STAR Early Literacy	• Fall 2011 and Winter 2012 (grades 1-2) • Winter 2012 and Spring 2012 (K)	• 72 1st grade pre/post • 31 2nd grade pre/post • 37 K pre/post	• Scores from the fall 2011 and winter 2012 DPS administrations of STAR Early Literacy were collected as pre- and posttests, respectively, for 1st and 2nd graders in the study for whom scores were available. Scores from Winter 2012 and Spring 2012 were collected as pre- and posttests, respectively, for kindergarteners in the study for whom scores were available.
Observations	• LLIOT	• November 2011 and January 2012 (grades 1-2) • March 2012 and April-May 2012 (K)	• 120 1st and 2nd grade LLIOTs • 44 K LLIOTs	• Trained on-site researchers observed all 1st and 2nd grade LLI groups twice in fall 2011/winter 2012, and all K LLI groups twice in spring 2012.
	• LOT	• October 2011 (K-2) • April-May 2012 (K-2)	• 26 K-2 LOTs	• Trained on-site researchers observed randomly selected K-2 classrooms during the literacy block at the beginning and end of the school year.
Surveys	• LLITQ • CTLIQ • LLIPQ • HLSQ	• May 2012 • May 2012 • May 2012 • April-May 2012	• 21 LLITQs • 51 CTLIQs • 7 LLIPQs • 137 HLSQs	• Surveys were completed at the end of the school year to obtain feedback from LLI teachers, K-2 classroom teachers, principals, and parents/guardians.

Table 7, continued

Type of Measure	Instrument	Timeline	Number Collected	Description
Structured Focus Groups	• LLI teacher focus group • On-site researcher focus group • District literacy specialist focus group	• May 2012 • May 2012 • May 2012	• Approximately 20 LLI teachers • Approximately 15 on-site researchers • 4 district literacy specialists	• Focus groups were held at the end of the school year to obtain qualitative feedback about LLI and students' progress.

*Spring is the only DPS assessment window during which all first and second graders take the DRA2. All kindergarteners take the DRA2 in winter and spring.

Number of Days of LLI Instruction

Across all 13 participating DPS schools, the first and second grade treatment group students received, on average, 62 days of LLI instruction over 18 weeks between September 2011 and February 2012, with a range of 43 to 75 days of instruction. The recommended amount of LLI instruction for first and second grade students, according to the LLI program guide, is 18 weeks (approximately 90 days).

Between February and May 2012, the kindergarten treatment group students received, on average, 45 days of LLI instruction over 12 weeks, with individual students ranging in their attendance from 27 to 69 days. According to program recommendations, kindergarteners should receive 14 weeks (approximately 70 days) of LLI instruction.

Due to student-level factors (e.g., individual absences or unavailability during LLI group time) as well as school or district limitations (e.g., holidays, assessment windows during which LLI teachers and/or students were pulled during LLI group time, delays in starting LLI due to scheduling conflicts or difficulty accessing student data), treatment group students in the study received less than the recommended amount of LLI instruction. All treatment group students (K-2) received approximately five fewer weeks of instruction than recommended – first and second graders received the equivalent of 12.4 weeks on average, while kindergarteners received the equivalent of 9 weeks on average.

Results

The following section presents the results of the evaluation, discussed in relation to each instrument and each grade level, as appropriate. First, a summary of the quantitative and qualitative results will be presented, and the conclusion section will further discuss these results as they pertain to each of the research questions in the present study.

Inter-Rater Reliability

Fountas & Pinnell Benchmark Assessment System

In order to establish inter-rater reliability for the Fountas & Pinnell benchmarks in the current study, 48 paired assessments were conducted by 14 pairs of on-site researchers during the pretest benchmark administration for first and second graders in Fall 2011. Inter-rater reliability between the two raters on their independent determinations of students' instructional levels was assessed using Cohen's Kappa, which determines whether the extent of agreement between two observers is above expected levels (i.e., provides the proportion of agreement greater than that expected by chance; Sim & Wright, 2005). The logical ordering for scoring of the data analyzed enables one to discriminate degrees of rater agreement and non-agreement, resulting in the computation and reporting of a "weighted" Kappa statistic. The weighted Kappa statistic is particularly appropriate when ratings are provided in ordered-categorical form, as in this case, and where raters scored on a continuum with nine levels ranging from 0 to 8 (i.e., Pre-A = 0, A = 1, B = 2, C = 3, D = 4, F = 5, G = 6, H = 7, I = 8). With respect to items like these, the weighted Kappa would assign more "weight" in those instances where ratings were closer together and less "weight" in those instances where ratings were farther apart.

With respect to the interpretation of the Kappa statistic, whether weighted or un-weighted, values between 0.41 and 0.60 are conventionally interpreted as an indication of "moderate" agreement between two raters, 0.61 to 0.80 are an indication of "substantial" agreement, while values of 0.81 or higher are conventionally interpreted as signs of "almost perfect" agreement (Landis & Koch, 1977).

For the group of paired ratings scores (n = 48), the weighted Kappa was 0.95 with a 95% Confidence Interval of (0.88 – 1.00), $p < .001$, indicating that almost perfect agreement was reached between raters. The 95% Confidence Interval is interpreted as meaning that we are 95% confident that the true Kappa value ranged from a low of 0.88 to a high of 1.00, as the value of 0.95 obtained from the data is only an estimate (i.e., a point estimate). However, Kappa statistics should be interpreted based on pairs of raters who are judging the same set of data, which is not the case when all 48 paired scores are treated as one dataset. When ratings were grouped by pair (n = 14), only one pair (Pair 4) had less than perfect agreement between raters. The weighted Kappa was 0.81 with a 95% Confidence Interval of (0.59 – 1.00), $p < .001$, indicating almost perfect agreement between raters. However, because the sample sizes (i.e., the number of ratings by pair) were very low, the interpretation of the Kappa statistics should be treated with caution (see Table 8).

Table 8: Inter-Rater Reliability Statistics by Pair for the Fountas & Pinnell Benchmarks

Pair	Number of Ratings	Weighted Kappa	Lower Level 95% Confidence Interval	Upper Level 95% Confidence Interval	*p*
1	1	NA	NA	NA	NA
2	2	1.00	1.00	1.00	0.157
3	9	1.00	1.00	1.00	< 0.001
4	9	0.81	0.59	1.00	< 0.001
5	1	NA	NA	NA	NA
6	9	1.00	1.00	1.00	< 0.001
7	2	1.00	1.00	1.00	0.157
8	2	NA	NA	NA	NA
9	3	1.00	1.00	1.00	0.083
10	1	NA	NA	NA	NA
11	1	NA	NA	NA	NA
12	2	1.00	1.00	1.00	0.157
13	1	NA	NA	NA	NA
14	5	1.00	1.00	1.00	< 0.001
All Pairs	48	0.95	0.88	1.00	< 0.001

NA = Raters had to use a minimum of 2 of the available 9 rating categories to calculate Kappa. In these cases, raters used only 1 of the available rating categories.

Leveled Literacy Observation Tool (LLIOT)

A total of 58 paired observations using the LLIOT were conducted by 20 pairs of observers (including both on-site researchers and members of the CREP research team) in Fall 2011, during the first round of first and second grade LLIOTs. For each of the 23 items, the percentage of observations in which the two independent raters exhibited either perfect agreement or agreement within one category (i.e., "Not Observed," "Needs Improvement," "Acceptable," or "Excellent") was calculated. To demonstrate perfect agreement, both raters had to select the same category (e.g., "Acceptable"). To demonstrate agreement within one category, both raters had to select the same category, or Rater 2 had to select one category above or below Rater 1 (e.g., Rater 1 selected "Acceptable" and Rater 2 selected "Needs Improvement" or "Excellent").

As shown in Table 9, raters exhibited agreement within one category over 75.0% of the time on all items but one ("Classroom connection"). Further, on the majority of these items, raters demonstrated agreement within one category over 90.0% of the time. The items with the highest percentage of agreement included "Students are actively engaged" (100.0%) and "Lesson is well organized" (98.3%), while the items with the lowest percentage of agreement included "Classroom connection" (58.6%) and "Instructional modifications are observed when needed" (77.6%). Overall, raters exhibited perfect agreement over half of the time on 15 out of the 23 items.

Table 9: Rater Agreement for LLIOT Items

Item	Percentage of Observations (n = 58)	
	Perfect Agreement	Agreement Within One Category
Quality of LLI Implementation		
Rereading (shared or independent)	53.4	81.0
Assessment using Reading Record	81.0	94.8
Phonics/word work (e.g., parts of words; single letter sounds; letter, word, or picture cards; magnetic letters)	65.5	94.8
Writing about reading (interactive, dictated, or independent)	77.6	94.8
New book - Introducing new text	72.4	94.8
New book - Reading (shared or independent)	56.9	94.8
New book - Monitoring and supporting students as needed	69.0	94.8
New book - Discussing and revisiting the story	46.6	82.8
Classroom connection	50.0	58.6
Home connection	46.6	87.9
Literacy Instructional Strategies		
Teacher models, encourages, and provides opportunities for fluent oral reading.	48.3	82.8
Teacher introduces vocabulary words (e.g., high frequency, story-specific words).	46.6	86.2
Teacher emphasizes understanding/comprehension of what is read.	48.3	93.1
Teacher models and encourages students to use appropriate reading strategies (e.g., phonemic awareness).	48.3	94.8
Teacher engages students in conversation about the text.	58.6	91.4
Teacher assists students in problem-solving.	53.4	93.1
Learning Environment		
Lesson is well organized.	75.9	98.3
Teacher appropriately paces lesson components.	44.8	91.4
Teacher engages in ongoing assessment of student learning (e.g., questioning, providing feedback/corrective instruction, checking responses).	74.1	94.8
Students are actively engaged.	70.7	100.0
Instructional modifications are observed when needed.	36.2	77.6
Instructional materials needed to implement lesson are readily available.	74.1	96.6
The lesson is delivered as designed.	63.8	96.6

Student Literacy Achievement

To determine whether LLI students' progress in literacy was statistically significantly different from that of their control group counterparts, analyses of pretest to posttest gains were conducted at Kindergarten and at Grades One and Two with respect to three measures of literacy achievement: the Fountas & Pinnell benchmarks, DRA2, and STAR Early Literacy assessment. At all three grades, benchmark gains were computed by obtaining a numeric equivalent for students' beginning and ending instructional levels (e.g., A = 1, B = 2, and so on) and then computing the difference between the posttest and pretest scores. Students who did not score as high as the lowest level, A, were assigned a level of "Pre-A" with a corresponding score of 0. Literacy gains on the DRA2 and STAR were computed based upon the timing of the test's administration: Winter 2011-2012 (pretest) to Spring 2012 (posttest) for Kindergarten students; Spring 2011 (pretest) to Spring 2012 (posttest) for students in Grades One

and Two on the DRA2[2]; and Fall 2011 (pretest) to Winter 2011-2012 (posttest) for Grades One and Two on the STAR. In a manner similar to computing benchmark gains, DRA2 gains were derived by subtracting posttest from pretest "levels." However, STAR gains were computed by transforming pretest and posttest Percentile Ranks into Normal Curve Equivalents (NCEs) via a standard formula and then computing a posttest to pretest NCE gain score.

For the aforementioned three measures, independent *t*-test analyses involving all treatment and control group students at a grade level were regarded as confirmatory in nature, with the Benjamini-Hochberg method applied to correct for multiple comparisons. While gain-score differences between treatment and control subgroups of students by such demographic characteristics as gender, ethnicity, ELL status, and special education status were likewise compared via multiple independent *t*-tests, no correction was applied to the probability level of the statistical outcomes as these analyses were regarded as exploratory.

In addition to the analyses of achievement gains involving students' demographic characteristics, other analyses of test score gains concerning conditions of program administration were investigated. At each grade level, differences in the fidelity of program implementation were explored by classifying student gains as a function of "no" LLI implementation (control) or of either "low" or "high" LLI implementation based on teachers' scores on the LLIOT. Because three groups were involved in these analyses, the mean gains for these groups were statistically tested using the Analysis of Variance (ANOVA), followed by post-hoc testing of all possible pairs of means. Similarly, for all gains by grade level, the role played by the level of student attendance in the program was examined by calculating the ratio of the percentage of LLI students' "actual days attended" versus their "ideal days attended" and then employing Pearson product-moment correlations to link the results with students' benchmark, DRA2, and STAR gains. For Kindergarten students, the "ideal" number of days of program attendance was given as 70; for students in Grades One and Two, the "ideal" number of days of program attendance was given as 90.

Kindergarten Gains

Controlling for multiple comparisons, statistically significant differences were observed for Kindergarten students with respect to both benchmark gains ($t(108)= 3.99$, $p < .001$, $g = 0.76$) and DRA2 gains ($t(105) = 2.69$, $p < .01$, $g = 0.52$), with the results in both instances favoring the treatment group over the control group (see Table 10). The effect sizes in both cases were "substantively important" (i.e., $g \geq 0.25$) according to What Works Clearinghouse guidelines (U.S. Department of Education, 2011). While the gains observed for the STAR also appeared to show some advantage for the treatment group, the difference was not statistically significant, given a much smaller number of paired scores and a large standard deviation relative to the mean difference observed between groups ($t(35) = 0.40$, *ns*, $g = 0.13$). In addition, the effect size for the STAR mean difference was not substantively important.

Among the exploratory analyses conducted by students' demographic characteristics and also shown in Table 10, statistically significant differences with respect to benchmark gains were observed for almost all treatment subgroups, including Kindergarten treatment group students reported as being male ($t(51) = 2.33$, $p < .05$, $g = 0.63$), female ($t(55) = 3.31$, $p < .01$, $g = 0.86$), not of Hispanic origin ($t(32) =$

[2] Although first and second grade treatment group students participated in LLI from Fall 2011 to Winter 2011-2012, a sufficient number of scores for these analyses were only available for Spring 2011 and Spring 2012 because DPS administers the DRA2 to all K-3 students in the spring. Kindergarteners also complete the DRA2 at midyear.

2.90, $p < .01$, $g = 0.97$), of Hispanic origin ($t(74) = 2.92$, $p < .01$, $g = 0.67$), not classified as receiving ELL services ($t(66) = 3.28$, $p < .01$, $g = 0.79$), classified as receiving ELL services ($t(40) = 2.27$, $p < .05$, $g = 0.69$), and not classified as receiving special education services ($t(101) = 3.97$, $p < .001$, $g = 0.78$). All effect sizes for the benchmark analyses, including the analysis for students classified as receiving special education services (which was not statistically significant), were substantively important.

With respect to DRA2 gains, exploratory analyses of subgroup differences indicated statistically significant advantages for Kindergarten students in the treatment condition who were classified as female ($t(55) = 2.87$, $p < .01$, $g = 0.75$), not of Hispanic origin ($t(31) = 3.80$, $p < .001$, $g = 1.29$), not classified as receiving special education services ($t(98) = 2.76$, $p < .01$, $g = 0.55$), and, at a liberal level of statistical significance (that is, $p < .10$), for Kindergarten students not classified as receiving ELL services ($t(65) = 1.91$, $p < .10$, $g = 0.46$) and classified as receiving ELL services ($t(38) = 1.83$, $p < .10$, $g = 0.57$). Regardless of the level of statistical significance, the effect sizes obtained for all five of the aforementioned subgroup comparisons were substantively important. It should be noted that the sample sizes for some subgroups, particularly for students designated as receiving special education services, were very small, limiting the statistical power of the test and the generalizability of the findings.

Table 10: Control and Treatment Group Gains for Kindergarten Students across Three Measures of Literacy Achievement

Achievement Measure	Control			Treatment			t		g
	n	*M*	*SD*	*n*	*M*	*SD*			
All Students									
Fountas & Pinnell Benchmarks	55	1.3	1.2	55	2.3	1.5	3.99	***	0.76
DRA2	52	1.3	1.5	55	2.1	1.4	2.69	**	0.52
STAR Early Literacy	20	-3.3	24.2	17	0.0	25.4	0.40		0.13
Males Only									
Fountas & Pinnell Benchmarks	27	1.1	1.1	26	1.9	1.5	2.33	*	0.63
DRA2	24	1.5	2.0	26	1.9	1.0	0.97		0.27
STAR Early Literacy	10	2.0	26.5	8	3.7	32.4	0.12		0.06
Females Only									
Fountas & Pinnell Benchmarks	28	1.4	1.3	29	2.6	1.4	3.31	**	0.86
DRA2	28	1.2	1.0	29	2.2	1.7	2.87	**	0.75
STAR Early Literacy	10	-8.6	21.8	9	-3.4	18.5	0.56		0.25
Non-Hispanic Only									
Fountas & Pinnell Benchmarks	18	1.1	1.1	16	2.2	1.2	2.90	**	0.97
DRA2	17	1.2	0.8	16	2.3	0.9	3.80	***	1.29
STAR Early Literacy	5	1.3	16.3	6	-0.7	36.6	-0.11		-0.06
Hispanic Only									
Fountas & Pinnell Benchmarks	37	1.4	1.3	39	2.3	1.6	2.92	**	0.67
DRA2	35	1.3	1.8	39	2.0	1.6	1.61		0.37
STAR Early Literacy	15	-4.8	26.7	11	0.3	19.0	0.54		0.22
Non-ELL Only									
Fountas & Pinnell Benchmarks	35	1.3	1.2	33	2.4	1.6	3.28	**	0.79
DRA2	34	1.3	1.8	33	2.0	1.0	1.91	†	0.46
STAR Early Literacy	20	-3.3	24.2	17	0.0	25.4	0.40		0.13
ELL Only									
Fountas & Pinnell Benchmarks	20	1.3	1.2	22	2.1	1.2	2.27	*	0.69
DRA2	18	1.3	1.0	22	2.2	1.9	1.83	†	0.57
STAR Early Literacy	0			0					

Table 10, continued

Achievement Measure	Control			Treatment			t		g
	n	*M*	*SD*	*n*	*M*	*SD*			
Non-SPED Only									
Fountas & Pinnell Benchmarks	52	1.3	1.2	51	2.3	1.4	3.97	***	0.78
DRA2	49	1.3	1.6	51	2.1	1.4	2.76	**	0.55
STAR Early Literacy	19	-4.4	24.4	17	0.0	25.4	0.53		0.17
SPED Only									
Fountas & Pinnell Benchmarks	3	0.7	0.6	4	1.5	1.7	0.78		0.50
DRA2	3	1.7	0.6	4	1.5	0.6	-0.38		-0.24
STAR Early Literacy	1	17.6		0					

*p < .05. **p < .01. ***p < .001.

Grade One Gains

For the Grade One treatment and control groups, no statistically significant differences were observed between the two groups in the aggregate with respect to either DRA2 gain scores or STAR gain scores (see Table 11). While most subgroup comparisons on these two measures also proved not to be statistically significant, an exception emerged as regards a single subgroup on one of the two measures. Among Grade One males, the DRA2 gains of the control group appeared to outpace those of the treatment group ($t(62) = -2.12$, $p < .05$, $g = -0.52$), with a substantively important effect size.

As contrasted with Grade One DRA2 and STAR outcomes, not only were statistically significant differences in benchmark level gains more often observed, such gains also tended systematically to favor the treatment group. Further inspection of Table 11 reveals that the benchmark gains made by the Grade One treatment group were observed to be statistically significantly higher than those of their control group counterparts with substantively important effect sizes, both across all students ($t(138) = 2.83$, $p < .01$, $g = 0.48$), as well as for the following demographic subgroups: treatment group females ($t(61) = 2.86$, $p < .01$, $g = 0.71$), treatment group Hispanic students ($t(94) = 2.89$, $p < .01$, $g = 0.59$), treatment group students not classified as receiving ELL services ($t(88) = 2.05$, $p < .05$, $g = 0.43$), treatment group students classified as receiving ELL services ($t(48) = 2.28$, $p < .05$, $g = 0.64$), and treatment group students not classified as receiving special education services ($t(122) = 3.02$, $p < .01$, $g = 0.54$). It should be noted that the sample sizes for many subgroups, particularly for students designated as receiving special education services, were very small, limiting the statistical power of the test and the generalizability of the findings.

Table 11: Control and Treatment Group Gains for Grade One Students across Three Measures of Literacy Achievement

Achievement Measure	Control			Treatment			t		g
	n	*M*	*SD*	*n*	*M*	*SD*			
All Students									
Fountas & Pinnell Benchmarks	69	3.4	1.8	71	4.4	2.2	2.83	**	0.48
DRA2	59	10.6	3.4	59	9.8	3.9	-1.23		-0.23
STAR Early Literacy	37	5.7	23.9	35	7.2	28.5	0.24		0.06
Males Only									
Fountas & Pinnell Benchmarks	37	3.5	1.8	40	4.1	2.2	1.26		0.28
DRA2	31	11.1	3.0	33	9.1	4.3	-2.12	*	-0.52
STAR Early Literacy	21	-0.2	21.3	19	7.7	27.8	1.02		0.32

Table 11, continued

Achievement Measure	Control			Treatment			t		g
	n	*M*	*SD*	*n*	*M*	*SD*			
Females Only									
Fountas & Pinnell Benchmarks	32	3.3	1.9	31	4.7	2.1	2.86	**	0.71
DRA2	28	10.1	3.9	26	10.7	3.1	0.57		0.15
STAR Early Literacy	16	13.6	25.6	16	6.6	30.3	-0.70		-0.24
Non-Hispanic Only									
Fountas & Pinnell Benchmarks	21	3.9	1.9	23	4.3	2.1	0.74		0.22
DRA2	17	11.3	3.4	21	11.1	3.3	-0.18		-0.06
STAR Early Literacy	13	14.5	33.9	14	4.9	30.6	-0.78		-0.29
Hispanic Only									
Fountas & Pinnell Benchmarks	48	3.2	1.8	48	4.4	2.2	2.89	**	0.59
DRA2	42	10.4	3.5	38	9.1	4.0	-1.53		-0.34
STAR Early Literacy	24	1.0	15.2	21	8.8	27.7	1.18		0.35
Non-ELL Only									
Fountas & Pinnell Benchmarks	47	3.7	1.8	43	4.6	2.2	2.05	*	0.43
DRA2	40	10.9	3.1	33	10.1	4.0	-0.91		-0.21
STAR Early Literacy	34	6.2	24.4	33	5.5	28.5	-0.11		-0.03
ELL Only									
Fountas & Pinnell Benchmarks	22	2.8	1.7	28	4.1	2.2	2.28	*	0.64
DRA2	19	10.2	4.1	26	9.4	3.7	-0.63		-0.19
STAR Early Literacy	3	0.5	21.7	2	35.3	6.5	2.10		1.39
Non-SPED Only									
Fountas & Pinnell Benchmarks	65	3.5	1.8	59	4.6	2.2	3.02	**	0.54
DRA2	55	10.7	3.3	51	10.1	3.4	-0.97		-0.19
STAR Early Literacy	37	5.7	23.9	31	5.1	28.9	-0.09		-0.02
SPED Only									
Fountas & Pinnell Benchmarks	4	2.0	1.6	12	3.3	1.8	1.32		0.72
DRA2	4	9.0	5.9	8	7.8	6.1	-0.34		-0.19
STAR Early Literacy	0			4	23.2	22.5			

*$p < .05$. **$p < .01$.

Grade Two Gains

As with Grade One students, no between-group differences were observed for the aggregate of Grade Two students with respect to either DRA2 or STAR gain scores (see Table 12). However, at a liberal level of significance (that is, $p < .10$), one difference emerged among Grade Two females such that the DRA2 gains of the treatment group appeared to exceed those of their controls ($t(22) = 2.05$, $p < .10$, $g = 0.82$), with a substantively important effect size.

With respect to Grade Two benchmark gains, only one statistically significant difference was observed at $p < .05$, with that difference favoring treatment group students classified as Hispanic ($t(48) = 2.05$, $p < .05$, $g = 0.57$) over their control group counterparts, with a substantively important effect size. However, at the aforementioned liberal level of significance (that is, $p < .10$), additional Grade Two differences in benchmark gains surfaced and appeared to favor treatment group males over their controls ($t(38) = 2.01$, $p < .10$, $g = 0.62$) and treatment group students not classified as receiving special education services over their controls ($t(61) = 1.80$, $p < .10$, $g = 0.45$). It should be noted that the sample sizes for many subgroups, particularly for females, students not of Hispanic origin, students designated as receiving ELL services, and students designated as receiving special education services, were very small, limiting the statistical power of the test and the generalizability of the findings.

Table 12: Control and Treatment Group Gains for Grade Two Students across Three Measures of Literacy Achievement

Achievement Measure	Control			Treatment			t		g
	n	*M*	*SD*	*n*	*M*	*SD*			
All Students									
Fountas & Pinnell Benchmarks	33	3.1	1.6	37	3.8	1.8	1.68		0.40
DRA2	30	10.9	5.6	30	10.9	5.9	0.02		0.01
STAR Early Literacy	12	9.0	11.7	19	6.0	16.5	-0.56		0.20
Males Only									
Fountas & Pinnell Benchmarks	20	2.9	1.5	20	4.0	2.0	2.01	†	0.62
DRA2	20	11.5	6.0	16	8.4	5.7	-1.56		-0.51
STAR Early Literacy	3	14.4	11.3	10	7.8	12.9	-0.80		-0.49
Females Only									
Fountas & Pinnell Benchmarks	13	3.5	1.8	17	3.6	1.6	0.17		0.06
DRA2	10	9.8	4.8	14	13.9	4.8	2.05	†	0.82
STAR Early Literacy	9	7.2	11.9	9	3.9	20.5	-0.42		-0.19
Non-Hispanic Only									
Fountas & Pinnell Benchmarks	10	3.8	1.7	10	3.9	2.1	0.12		0.05
DRA2	7	6.9	5.6	5	10.0	2.4	1.16		0.63
STAR Early Literacy	4	8.6	6.6	6	3.2	19.8	-0.52		-0.30
Hispanic Only									
Fountas & Pinnell Benchmarks	23	2.8	1.5	27	3.8	1.7	2.05	*	0.57
DRA2	23	12.1	5.1	25	11.1	6.4	-0.60		-0.17
STAR Early Literacy	8	9.2	14.0	13	7.2	15.5	-0.30		-0.13
Non-ELL Only									
Fountas & Pinnell Benchmarks	24	3.2	1.6	29	3.6	1.8	0.90		0.24
DRA2	21	10.5	5.7	22	9.5	5.6	-0.57		-0.17
STAR Early Literacy	12	9.0	11.7	16	9.4	14.7	0.07		0.03
ELL Only									
Fountas & Pinnell Benchmarks	9	3.0	1.9	8	4.6	1.6	1.91		0.88
DRA2	9	11.8	5.6	8	14.8	5.4	1.11		0.51
STAR Early Literacy	0			3	-12.5	15.6			
Non-SPED Only									
Fountas & Pinnell Benchmarks	29	3.1	1.5	34	3.9	1.9	1.80	†	0.45
DRA2	27	10.7	5.5	28	11.4	5.1	0.46		0.12
STAR Early Literacy	10	8.1	11.9	17	3.7	15.8	-0.76		-0.29
SPED Only									
Fountas & Pinnell Benchmarks	4	3.5	2.4	3	3.3	0.6	0.12		-0.07
DRA2	3	12.7	7.6	2	5.0	15.6	-0.77		-0.51
STAR Early Literacy	2	13.7	13.4	2	25.0	11.0	0.92		0.52

†$p < .10$. *$p < .05$.

Post-Hoc Benchmark Analyses

Promising results obtained from a statistical analysis of Grade Two students' Benchmark scores in Denver prompted an investigation of the effects of increasing the power of the analysis by increasing sample size. Given the progress made, though not statistically significant, post-hoc analyses were conducted in order to examine these effects in a larger sample. The second grade sample from DPS and a prior study with schools in Tifton, GA and Middletown, NY were then combined, after statistically being equated. When an Analysis of Variance (ANOVA) was conducted on the pre-test Benchmark means obtained for the Grade Two LLI treatment and control groups in Denver, Middletown, and Tifton (see Table 1 and Figure 1), the results pointed to no difference in performance across the six groups considered individually ($F(5, 215) = 1.63$, $p = .239$) and suggested that combining student groups to

achieve greater power was a legitimate strategy. Interestingly, after adding together the three control groups (n = 103) and three treatment groups (n = 118) and re-analyzing the Benchmark pretest scores, an advantage for the combined LLI control group (M = 5.89, SD = 2.45) over the combined LLI treatment group (M = 5.24, SD = 2.26) emerged that proved upon testing to be both statistically significant as well as substantively meaningful ($t(207) = 2.19$, $p = .04$, $g = 0.28$).

Notwithstanding the advantage of the control group on the Benchmark pretest, the results of a hierarchical multiple regression (see Table 2) that controlled for students' demographic characteristics ($R^2 = .10$) as well as their scores on the pretest ($R^2 = .47$) indicated a positive difference on students' Benchmark posttest scores based on students' membership in the combined LLI treatment group (M = 9.80, SD =2.25) rather than the combined LLI control group (M = 8.82, SD = 2.64) and revealed by a statistically significant increase in the total proportion of variance explained (to $R^2 = .53$). According to the text by Lipsey and Wilson's (2001) on *Practical Meta-Analysis*, the effect size associated with the standardized coefficient for LLI group in this regression (that is, $B = 0.25$) roughly equates to $g = 0.51$.

Table 13: Means and Standard Deviations for Grade Two Treatment and Control Groups in Denver, Middletown and Tifton with respect to Pre- and Posttest Benchmark Outcomes

Control Groups

District	*n*	Pre-Benchmark		Post-Benchmark		Benchmark Gain	
		M	*SD*	*M*	*SD*	*M*	*SD*
Denver	33	5.73	2.17	8.85	2.08	3.12	1.62
Middletown	40	5.73	2.63	9.70	2.54	3.98	1.59
Tifton	30	6.30	2.51	7.97	3.06	1.67	1.47
Combined	103	5.89	2.45	8.92	2.64	3.03	1.82

Treatment Groups

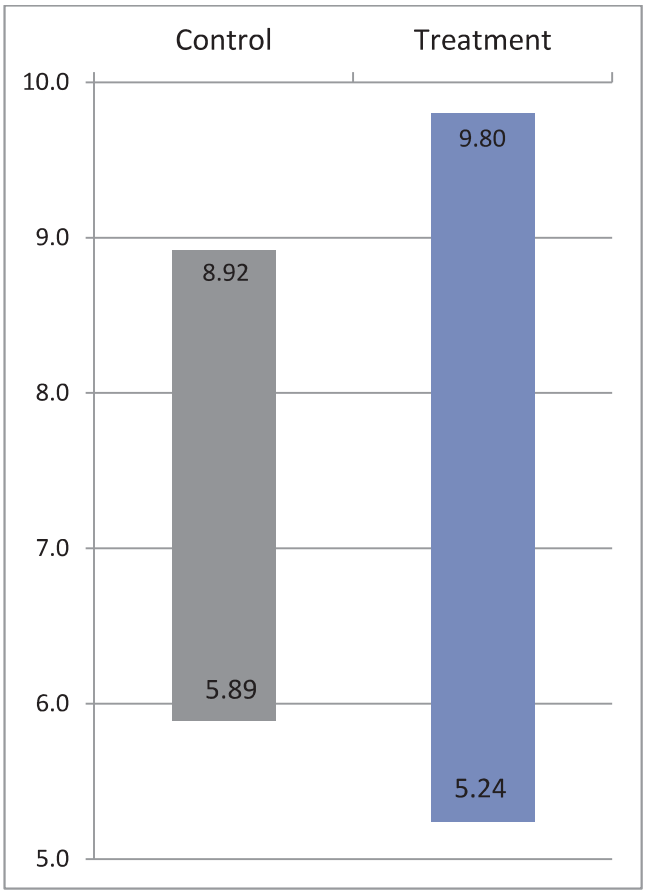
District	*n*	Pre-Benchmark		Post-Benchmark		Benchmark Gain	
		M	*SD*	*M*	*SD*	*M*	*SD*
Denver	37	5.51	2.17	9.32	2.08	3.81	1.62
Middletown	44	4.93	2.33	10.77	2.24	5.84	1.84
Tifton	37	5.32	1.92	9.11	2.50	3.78	1.46
Combined	118	5.24	2.26	9.80	2.52	4.56	1.97

Table 14: Hierarchical Regression Summary of Grade Two Student Demographics, Benchmark Pretest Scores, and Treatment/Control Group Status on Benchmark Posttest Scores

Predictor	*B*	*SEB*	*B*	*t*	*p* =
Model Fit: F (4, 216) = 6.07, $p < .001$, $R^2 = .10$					
ELL (0 = No 1 = Yes)	-0.84	0.48	-0.12	-1.75	0.081
SPED (0 = No, 1 = Yes)	-2.67	0.58	-0.30	-4.63	0.000
Gender (0 = M, 1 = F)	0.32	0.34	0.06	0.92	0.356
Hispanic (0 = No, 1 = Yes)	0.36	0.36	0.07	1.00	0.321
Model Fit: F (5, 215) =38.74, $p < .001$ F Change (1, 215) = 152.41, $p < 001$, $R^2 = .47$					
ELL (0 = No 1 = Yes)	-0.43	0.37	-0.06	-1.17	0.244
SPED (0 = No, 1 = Yes)	-1.54	0.45	-0.17	-3.40	0.001
Gender (0 = M, 1 = F)	0.05	0.26	0.01	0.18	0.861
Hispanic (0 = No, 1 = Yes)	0.11	0.28	0.02	0.41	0.682
Prestest Benchmark	0.69	0.06	0.63	12.35	0.000
Model Fit: F (6, 214) = 40.46, $p < .001$ F Change (1, 214) = 26.28, $p < 001$, $R^2 = .53$					
ELL (0 = No 1 = Yes)	-0.28	0.35	-0.04	-0.81	0.418
SPED (0 = No, 1 = Yes)	-1.23	0.43	-0.14	-2.84	0.005
Gender (0 = M, 1 = F)	-0.03	0.25	-0.01	-0.11	0.915
Hispanic (0 = No, 1 = Yes)	0.00	0.26	0.00	0.00	0.998
Prestest Benchmark	0.74	0.05	0.67	13.75	0.000
LLI Group (0 = C, 1 = T)	1.28	0.25	0.25	5.13	0.000

As depicted in Figure 2, computation of the effect size linked to the pre- to posttest Benchmark gains seen for the treatment (M = 4.56, SD = 1.97) and control groups (M = 3.03, SD = 1.82) was found to be $g = 0.80$. The treatment group gained about 4.5 levels, finishing close to Level J, while the control group only gained around 3 levels, finishing close to Level I. These results are similar to those seen when looking only at the Denver second grade students; however, with the increased combined sample size, statistical significance was also attained. In other words, the control group started with a statistically and practically significant advantage on the benchmark pretest benchmark; however, the treatment group finished with a higher posttest mean and a greater pretest to posttest gain.

Figure 2: Benchmark Assessment: 2nd grade Combined Sample Pretest to Posttest Gains for Control and Treatment Group



Other Analyses

As shown in Table 15, fidelity of implementation appeared to be unrelated to students' gains on the STAR, but did seem to influence their benchmark gains in both Kindergarten ($F(2, 107) = 10.95$, $p < .001$, $g = 1.04$) and Grade One ($F(2, 137) = 4.51$, $p < .01$, $g = 0.61$), as well as their DRA2 gains in Kindergarten ($F(2, 104) = 3.90$, $p < .05$, $g = 0.57$). Among Kindergarten students, follow-up testing indicated that students subject to "high" fidelity of LLI implementation obtained Bonferroni-corrected statistically significantly higher mean gains on both benchmark ($M = 2.7$, $SD = 1.6$) and DRA2 measures ($M = 2.2$, $SD = 1.7$), with substantively important effect sizes, when compared to Kindergarten students under conditions of "no implementation" on the same two measures. Similarly, among Grade One students, a single Bonferroni-corrected statistically significant difference was observed, with the benchmark gain made by Grade One students under "high" fidelity of implementation ($M = 4.6$, $SD = 2.2$) exceeding Grade One student performance under the "no implementation" condition ($M = 3.4$, $SD = 1.8$), with a substantively important effect size. At neither of these two grade levels, however, did either benchmark or DRA2 gains made by students in "low" fidelity of LLI implementation contexts differ statistically from those observed for students under conditions of "no implementation" or from those observed for students under conditions of "high" fidelity of implementation.

Table 15: Analysis of Variance Results for Differences in Test Score Gains by Fidelity of Implementation

Achievement Measure	Control Group (0)			Low Implementation (1)			High Implementation (2)			F		Grps
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*			
Kindergarten												
Fountas & Pinnell Benchmarks	55	1.3	1.2	30	1.9	1.2	25	2.7	1.6	10.9	***	0 < 2[1]
DRA2	52	1.3	1.5	30	1.9	1.0	25	2.2	1.7	3.9	*	0 < 2[2]
STAR Early Literacy	20	-3.3	24.2	9	1.4	33.2	8	-1.7	14.4	0.1		
Grade One												
Fountas & Pinnell Benchmarks	69	3.4	1.8	38	4.2	2.2	33	4.6	2.2	4.5	**	0 < 2[3]
DRA2	59	10.6	3.4	30	9.2	4.3	29	10.4	3.4	1.5		
STAR Early Literacy	37	5.7	23.9	18	10.6	32.6	17	3.6	24.1	0.3		
Grade Two												
Fountas & Pinnell Benchmarks	33	3.1	1.6	18	4.1	1.9	19	3.6	1.7	1.8		
DRA2	30	10.9	5.6	16	12.1	6.2	14	9.6	5.4	0.7		
STAR Early Literacy	12	9.0	11.7	8	10.6	18.8	11	2.6	14.7	0.8		

**$p < .01$. ***$p < .001$.
[1]$g = 1.04$; [2]$g = 0.57$; [3]$g = 0.61$.

Although LLI fidelity of implementation was observed to have some effect on students' literacy achievement at Kindergarten and Grade One, no such effect was observed at any of the three grade levels with respect to the percentage of days attended by LLI students and either benchmark or DRA2 outcomes (see Table 16). Interestingly, however, two statistically significant correlations emerged regarding LLI student attendance and gains made on the STAR. Although a positive correlation was observed between Grade One STAR gains and LLI student attendance at a more liberal significance level ($r(35) = .326$, $p < .10$), a negative correlation was observed between Kindergarten STAR gains and LLI student attendance ($r(17) = -.486$, $p < .05$). While the latter outcome would seem counter-intuitively to imply that more time in the program results in lower gains, the correlation is very likely spurious because far fewer Kindergarten students had paired STAR scores ($n = 17$) than had either paired benchmark scores ($n = 55$) or paired DRA2 scores ($n = 55$); and, among those few students with pairs of STAR scores, more than a few evidenced "gains" in NCEs that were actually in the negative direction.

Table 16: Correlations between Percentage of Days Attended by LLI Students and Gains Made on Three Measures of Literacy Achievement

Achievement Measure	Grade Level		
	Kindergarten	*Grade One*	*Grade Two*
Fountas & Pinnell Benchmarks	0.207 ($n = 55$)	0.137 ($n = 71$)	0.244 ($n = 37$)
DRA2	0.115 ($n = 55$)	0.107 ($n = 59$)	0.272 ($n = 30$)
STAR Early Literacy	-0.486* ($n = 17$)	0.326† ($n = 35$)	0.014 ($n = 19$)

†$p < .10$. *$p < .05$.

Summary of Treatment and Control Group Achievement Results

Overall, it appears that the LLI intervention had a more systematic impact on benchmark performance as compared to performance on the DRA2 and STAR. Effects also appeared to be stronger in Kindergarten as opposed to Grade One and Grade Two (based on both statistical significance and effect sizes), with the least impact on Grade Two. The trend for differential impact on benchmark outcomes also appeared when looking at the influence of fidelity of implementation, with high fidelity implementation having a statistically significant impact on benchmark gains compared to the control group for both Kindergarten and Grade One, although some effect of high fidelity implementation was also seen for Kindergarten DRA2 gains. Finally, program attendance did not appear to have a reliable impact on literacy achievement gains.

Comparison of Benchmark and DRA2 Outcomes

Given the grade level equivalence chart provided for different literacy assessments in the 2009 LLI Program Guide (see Appendix J), students' Fountas & Pinnell Benchmark scores and their DRA2 scores were examined to determine the degree of alignment in students' level of literacy achievement as measured by the two tests and to compare the percentages of students the two tests identified as being proficient or not proficient. Grade level equivalence information was not available for the STAR Early Literacy Assessment, which also did not have a sufficient sample size to support such an analysis; therefore, alignment was only examined between the Fountas & Pinnell Benchmarks and the DRA2. For treatment and control group students at all three grades levels, literacy achievement was defined as the students' posttest benchmark level. Because the Kindergarten posttest benchmark level was measured in Spring 2012, the comparable DRA2 achievement level score used was also for Spring 2012. However, because the posttest benchmark level for first and second grade students was collected in Winter 2011-2012, the comparable DRA2 achievement level used was for Winter 2011-2012. With all but three Kindergarten students having both sets of scores, the results of the analyses for Kindergarteners presented in Table 17 and Table 18 may be regarded as largely conclusive. However, when interpreting the results for the two upper grades, the high percentage of missing DRA2 scores observed for the Grade One control (71.0%) and treatment (52.1%) group students and the Grade Two control (57.6%) and treatment (43.2%) group students warrants a great deal of caution (see Table 17). Missing DRA2 scores were primarily due to the fact the DPS does not administer the DRA2 to all first and second graders at the Winter 2011-2012 timeframe.

In Kindergarten, analyses of the alignment of students' literacy achievement levels suggested perfect match between benchmark and DRA2 metrics for nearly one-fifth of students in the control group (18.2%) and nearly one-quarter of students in the treatment group (23.6%). According to the grade level goals published in the 2009 LLI Program Guide, 29.1% of control group students and 60.0% of treatment group students can be designated as "proficient" with respect to the benchmark criteria, having achieved a level of "C" or higher (see Table 17).[3] However, as measured in DRA2 levels comparable to these benchmark criteria, literacy proficiency was achieved by roughly three-quarters of control group students (76.4%) and nearly all treatment group students (96.4%), having attained a DRA2 level of "3" or higher on the Spring 2012 administration of that assessment (see Table 17). Cross-tabulation of the percentages of students regarded as either proficient or non-proficient on the benchmark and DRA2 measures reveals similar cross-test agreement to have occurred for slightly less

[3] Because the information regarding grade level equivalence was obtained from the 2009 LLI Program Guide, the grade level goals for the Fountas & Pinnell Benchmarks reflect the 2009 rather than the revised 2012 Fountas & Pinnell Instructional Level Expectations.

than half of the Kindergarten students in the control group (specifically, 19.2% "non-proficient" on both tests and 28.8% "proficient" on both tests) but for over 60.0% of the students in the treatment group (specifically, 3.6% "non-proficient" on both tests and 60.0% "proficient" on both tests; see Table 18).

As previously stated, the high percentage of missing values on the DRA2 makes drawing firm conclusions about the alignment of test results for students in Grades One and Two difficult at best. However, despite the inability to draw definitive conclusions about the "match" between test levels and differences in the gross percentages of students that the two tests identify as proficient, analysis of the cross-tabulated outcomes suggests that the two tests do an effective job of reaching the same judgment about students' proficiency or non-proficiency as readers. Albeit based on only a small number of cases, the results obtained for Grade One students indicate that the two tests agree about the proficiency status of students in 85.0% of the control group cases analyzed (60.0% agreement on non-proficiency, 25.0% agreement on proficiency) and 94.1% of the treatment group cases analyzed (50.0% agreement on non-proficiency, 44.1% agreement on proficiency; see Table 18). For Grade Two students, similar percentages were observed, with both tests rendering the same judgment on 92.9% of the control group cases analyzed (78.6% agreement on non-proficiency, 14.3% agreement on proficiency)and 90.5% of the treatment group cases analyzed (81.0% agreement on non-proficiency, 9.5% agreement on proficiency; see Table 18).

Table 17: Percentage of Aligned Benchmark/DRA2 Levels and Percentage Proficient on Each Assessment

Group	Match Benchmark/DRA2		Proficiency Benchmark		Proficiency DRA2	
	n	%	*n*	%	*n*	%
Kindergarten Control (*n* = 55)						
Yes	10	18.2	16	29.1	42	76.4
No	42	76.4	39	70.9	10	18.2
Missing DRA2 Score	3	5.5	0	0.0	3	5.5
Kindergarten Treatment (*n* = 55)						
Yes	13	23.6	33	60.0	53	96.4
No	42	76.4	22	40.0	2	3.6
Missing DRA2 Score	0	0.0	0	0.0	0	0.0
Grade One Control (*n* = 69)						
Yes	5	7.2	21	30.4	7	10.1
No	15	21.7	48	69.6	13	18.8
Missing DRA2 Score	49	71.0	0	0.0	49	71.0
Grade One Treatment (*n* = 71)						
Yes	11	15.5	30	42.3	15	21.1
No	23	32.4	41	57.7	19	26.8
Missing DRA2 Score	37	52.1	0	0.0	37	52.1
Grade Two Control (*n* = 33)						
Yes	6	18.2	8	24.2	2	6.1
No	8	24.2	25	75.8	12	36.4
Missing DRA2 Score	19	57.6	0	0.0	19	57.6
Grade Two Treatment (*n* = 37)						
Yes	3	8.1	11	29.7	2	5.4
No	18	48.6	26	70.3	19	51.4
Missing DRA2 Score	16	43.2	0	0.0	16	43.2

Table 18: Proficiency and Non-Proficiency Agreement between Fountas & Pinnell Benchmarks and DRA2

Benchmark Status	Control		Treatment	
	DRA2 Status			
	Not Proficient	Proficient	Not Proficient	Proficient
Kindergarten (Control *n* = 52, Treatment *n* = 55)				
Not Proficient	10 19.2%	27 51.9%	2 3.6%	20 36.4%
Proficient	0 0.0%	15 28.8%	0 0.0%	33 60.0%
Grade One (Control *n* = 20, Treatment *n* = 34)				
Not Proficient	12 60.0%	2 10.0%	17 50.0%	0 0.0%
Proficient	1 5.0%	5 25.0%	2 5.9%	15 44.1%
Grade Two (Control *n* = 14, Treatment *n* = 21)				
Not Proficient	11 78.6%	0 0.0%	17 81.0%	0 0.0%
Proficient	1 7.1%	2 14.3%	2 9.5%	2 9.5%

Comparison Group Results

Comparisons by Grade Level

Pretest and posttest scores on the DRA2 and STAR were obtained for a comparison group of LLI students in Kindergarten, Grade One, and Grade Two who were enrolled in LLI but not randomly assigned to receive it, and who could receive additional instructional time, literacy support, or interventions as needed. The gains were computed and analyzed in a manner comparable to that employed with the LLI experimental groups: Winter 2011-2012 (pretest) to Spring 2012 (posttest) for Kindergarten students; Spring 2011 (pretest) to Spring 2012 (posttest) for students in Grades One and Two on the DRA2; and Fall 2011 (pretest) to Winter 2011-2012 (posttest) for Grades One and Two on the STAR. Of the 386 students for whom data was received, a total of 334 students had complete sets of pre- and posttest scores on the DRA2 and thus were included in the analysis. Only 138 students were able to be included in the STAR analysis as a result of having complete sets of pre- and posttest scores. To facilitate analyses of STAR scores, such scores were transformed from Percentile Ranks to Normal Curve Equivalents (NCEs). It should be noted that all comparison group analyses were treated as exploratory, and therefore no corrections were made for multiple comparisons.

As may be noted in Table 19 through Table 21, gains made at all three grade levels on the DRA2 were highly statistically significant, with effect sizes that were "substantively important" (i.e., ≥ 0.25) according to What Works Clearinghouse guidelines (U.S. Department of Education, 2011), exceeding 2.0 at Kindergarten and Grade One and approaching 2.0 at Grade Two. With respect to STAR gains, only those observed at Grade Two proved to be statistically significant ($t(65) = 2.2$, $p < .05$, $g = 0.30$), with a substantively important effect size. At the same time, although the STAR gains obtained for Kindergarten were not statistically significant most likely due a small sample size ($t(7) = 1.70$, *ns*, $g = 0.80$), the effect size associated with the difference showed the NCE gains to be substantively meaningful.

Comparisons by Demographic Subgroups within Grade Level

In addition to the previously discussed analyses by grade level, analyses of DRA2 and STAR pre- and posttest scores were also conducted within grade levels for students categorized by their gender (male, female), ethnicity (not Hispanic, Hispanic), ELL status (not receiving ELL services, receiving ELL services) and special education status (not receiving SPED services, receiving SPED services) and are presented in Table 19 through Table 21. With respect to the 24 comparisons involving DRA2 scores, only two proved to be less than highly statistically significant (at $p < .001$) most likely due to very small sample sizes: specifically, the comparison involving non-Hispanic Kindergarten students ($t(2) = 5.0$, $p < .05$, $g = 1.5$) and the comparison involving special education Kindergarten students ($t(1) = 3.0$, *ns*, $g = 0.10$). It should be noted, however, that sample sizes were also very small for both gender groups as well as for both ELL groups. For the 22 DRA2 score comparisons remaining, pretest to posttest differences that were both statistically significant and substantively meaningful were observed, with all effect sizes seen to be at or above a value of 1.0 and many proving to be at or above a value of 2.0. Effect size differences in DRA2 scores that were at or above a value of 3.0 were observed for three Grade One subgroups by gender, ethnicity, and special education status: specifically, female students ($t(53) = 21.7$, $p < .001$, $g = 3.0$), non-Hispanic students ($t(35) = 17.8$, $p < .001$, $g = 3.1$), and students not classified as receiving special education services ($t(97) = 22.9$, $p < .001$, $g = 3.1$).

In contrast to the aforementioned results of comparing pre- and posttest DRA2 scores, similar comparisons involving STAR scores evidenced effect sizes that in most instances were slight and in some instances approached zero. Of these 24 STAR score comparisons, only one proved to be statistically significant and involved the NCE gain made by Grade Two males ($t(35) = 2.3$, $p < .05$, $g = 0.40$), with a substantively important effect size. It should be noted that the sample sizes for all Kindergarten subgroups were extremely small, limiting the statistical power of the test and the generalizability of the findings. However, at a liberal level of significance (that is, $p < .10$), additional gains on the STAR surfaced for Kindergarten students not classified as receiving special education services ($t(6) = 2.0$, $p < .10$, $g = 1.0$), Grade Two students classified as Hispanic ($t(48) = 2.0$, $p < .10$, $g = 0.3$), Grade Two students not classified as receiving ELL services ($t(39) = 1.9$, $p < .10$, $g = 0.3$), and Grade Two students not classified as receiving special education services ($t(56) = 2.0$, $p < .10$, $g = 0.3$), all with substantively important effect sizes. Further, while not statistically significant, the effect sizes in Kindergarten for females ($g = 0.50$), Hispanic students ($g = 0.80$), and students not classified as receiving ELL services ($g = 0.60$) were all substantively important. Finally, while not statistically significant, the effect sizes in Grade Two for non-Hispanic students ($g = 0.30$) and students classified as receiving special education services ($g = 0.40$) were substantively important.

Table 19: Results of Pretest and Posttest DRA2 and STAR Gain Analyses for a Comparison Group of Kindergarten Students

Achievement Measure	Pretest			Posttest			t		g
	n	*M*	*SD*	*n*	*M*	*SD*			
All Students									
DRA2	25	1.4	0.8	25	3.2	0.9	14.7	***	2.1
STAR Early Literacy	8	53.4	10.1	8	63.6	14.6	1.7		0.8
Males Only									
DRA2	10	1.6	0.7	10	3.5	0.7	10.6	***	2.7
STAR Early Literacy	2	46.3	0.8	2	71.4	25.8	1.4		0.0
Females Only									
DRA2	15	1.3	0.9	15	2.9	0.9	10.5	***	1.9
STAR Early Literacy	6	55.8	10.7	6	61.0	11.5	1.1		0.5
Non-Hispanic Only									
DRA2	3	1.3	1.2	3	3.0	1.0	5.0	*	1.5
STAR Early Literacy	0			0					
Hispanic Only									
DRA2	22	1.4	0.8	22	3.2	0.9	13.6	***	2.1
STAR Early Literacy	8	53.4	10.1	8	63.6	14.6	1.7		0.8
Non-ELL Only									
DRA2	10	1.4	1.0	10	3.2	0.8	9.0	***	2.0
STAR Early Literacy	6	56.2	10.3	6	61.9	10.3	1.2		0.6
ELL Only									
DRA2	15	1.4	0.7	15	3.1	0.9	11.3	***	2.0
STAR Early Literacy	2	45.2	2.3	2	68.5	29.9	1.2		0.0
Non-SPED Only									
DRA2	23	1.4	0.8	23	3.2	0.9	14.3	***	2.2
STAR Early Literacy	7	52.7	10.6	7	65.2	14.9	2.0	†	1.0
SPED Only									
DRA2	2	1.0	1.4	2	2.5	0.7	3.0		0.1
STAR Early Literacy	1	58.7	0.0	1	52.1	0.0			

†$p < .10$. *$p < .05$. ***$p < .001$.

Table 20: Results of Pretest and Posttest DRA2 and STAR Gain Analyses for a Comparison Group of Grade One Students

Achievement Measure	Pretest			Posttest			t		g
	n	*M*	*SD*	*n*	*M*	*SD*			
All Students									
DRA2	112	3.1	1.8	112	13.3	4.6	23.6	***	2.8
STAR Early Literacy	64	45.6	19.6	64	46.4	13.6	0.9		0.0
Males Only									
DRA2	58	3.3	3.3	58	13.1	5.1	14.1	***	2.5
STAR Early Literacy	29	44.2	21.6	29	45.0	16.2	0.6		0.0
Females Only									
DRA2	54	2.9	1.4	54	13.5	4.1	21.7	***	3.0
STAR Early Literacy	35	46.7	17.9	35	47.6	11.1	0.7		0.0
Non-Hispanic Only									
DRA2	36	3.4	2.1	36	14.9	4.3	17.8	***	3.1
STAR Early Literacy	29	43.7	20.9	29	45.7	12.8	1.3		0.0
Hispanic Only									
DRA2	76	2.9	1.7	76	12.6	4.6	17.5	***	2.7
STAR Early Literacy	35	47.2	18.6	35	47.0	14.4	-0.2		0.0
Non-ELL Only									
DRA2	69	3.2	2.0	69	13.1	4.7	16.5	***	2.7
STAR Early Literacy	48	44.0	19.4	48	45.7	13.0	1.6		0.0
ELL Only									
DRA2	43	2.9	1.6	43	13.6	4.5	18.5	***	2.5
STAR Early Literacy	16	50.3	20.0	16	48.7	15.6	-1.1		-0.1
Non-SPED Only									
DRA2	98	3.1	1.7	98	13.5	4.3	22.9	***	3.1
STAR Early Literacy	64	45.6	19.6	64	46.4	13.6	0.9		0.0
SPED Only									
DRA2	14	3.4	2.7	14	11.9	6.3	6.8	***	1.3
STAR Early Literacy	0			0					

***$p < .001$

Table 21: Results of Pretest and Posttest DRA2 and STAR Gain Analyses for a Comparison Group of Grade Two Students

Achievement Measure	Pretest			Posttest			t		g
	n	*M*	*SD*	*n*	*M*	*SD*			
All Students									
DRA2	197	9.8	4.5	197	21.2	7.1	27.3	***	1.8
STAR Early Literacy	66	33.9	13.5	66	38.0	15.9	2.2	*	0.3
Males Only									
DRA2	100	9.7	4.0	100	21.4	6.8	19.2	***	2.0
STAR Early Literacy	36	33.3	13.1	36	39.0	13.6	2.3	*	0.4
Females Only									
DRA2	97	9.9	5.0	97	21.0	7.4	19.4	***	1.6
STAR Early Literacy	30	34.8	14.3	30	36.9	18.4	0.8		0.1
Non-Hispanic Only									
DRA2	44	10.4	4.2	44	20.7	6.9	11.6	***	1.7
STAR Early Literacy	17	36.5	13.1	17	40.5	15.2	1.0		0.3
Hispanic Only									
DRA2	153	9.6	4.6	153	21.3	7.2	24.8	***	1.8
STAR Early Literacy	49	33.0	13.7	49	37.2	16.2	2.0	†	0.3
Non-ELL Only									
DRA2	104	11.1	4.0	104	21.8	7.1	19.5	***	1.7
STAR Early Literacy	40	34.4	12.5	40	39.2	16.2	1.9	†	0.3
ELL Only									
DRA2	93	8.4	4.6	93	20.5	7.1	19.3	***	1.9
STAR Early Literacy	26	33.3	15.3	26	36.2	15.5	1.1		0.2
Non-SPED Only									
DRA2	161	10.3	4.3	161	22.4	6.1	27.1	***	2.2
STAR Early Literacy	57	35.9	13.0	57	39.8	15.4	2.0	†	0.3
SPED Only									
DRA2	36	7.5	4.7	36	15.9	8.6	8.5	***	1.0
STAR Early Literacy	9	21.8	10.3	9	26.8	15.2	1.0		0.4

†$p < .10$. *$p < .05$. ***$p < .001$.

Intervention Fidelity

Leveled Literacy Intervention Observation Tool (LLIOT)

Descriptive Results

The Leveled Literacy Intervention Observation Tool (LLIOT) involved a targeted, 30-minute observation of LLI implementation and instructional strategies (*n* = 164 observations). Table 22 illustrates the frequencies for each item on the LLIOT, as observed during the visits. The results from the LLIOT revealed that 4 of the 10 LLI lesson components were rated "Acceptable" or "Excellent" over 90.0% of the time, and 8 of the 10 components were rated "Acceptable" or "Excellent" at least 85.0% of the time, indicating a high level of implementation fidelity.[4] The highest rated lesson components (i.e., demonstrating the highest degree of implementation fidelity) included reading a new book, introducing new text, rereading, and phonics/word work, which were rated "Acceptable" or "Excellent" 93.3%, 92.7%, 91.5%, and 91.5% of the time, respectively. The lowest rated component (i.e., demonstrating the lowest degree of implementation fidelity) was classroom connection, which was not seen in over half of the observed lessons (55.5%). Teachers were also rated highly (i.e., "Acceptable" or "Excellent") on their use of literacy instructional strategies, such as emphasizing comprehension (91.5%), assisting students in problem-solving (91.5%), engaging students in conversation about the text (90.8%), and introducing vocabulary words (90.3%). Further, in the majority of observed lessons, the lesson was well-organized, students were engaged and attentive, and the teacher continually assessed student learning (rated "Acceptable" or "Excellent" 95.1%, 93.9%, and 93.2% of the time, respectively). Overall, observers perceived that the lesson was delivered as designed 95.1% of the time, and the average rating across all subscales of the LLIOT was 2.5 (i.e., between "Acceptable" [2] and "Excellent" [3]). All items can be found in Table 22 below.

Table 22: LLIOT Response Frequencies (*n* = 164)

Item	Percent Responded			
	Excellent	Acceptable	Needs Improvement	Not Observed
Quality of LLI Implementation				
Rereading (shared or independent)	72.6	18.9	3.0	5.5
Assessment using Reading Record*	36.6	11.6	0.6	50.6
Phonics/word work (e.g., parts of words; single letter sounds; letter, word, or picture cards; magnetic letters)	67.1	24.4	6.1	2.4
Writing about reading (interactive, dictated, or independent)*	45.7	6.7	4.3	43.3
New book - Introducing new text	73.8	18.9	4.3	3.0
New book - Reading (shared or independent)	72.6	20.7	3.0	3.7
New book - Monitoring and supporting students as needed	75.6	12.8	7.3	4.3
New book - Discussing and revisiting the story	52.4	30.5	7.9	9.1
Classroom connection	18.9	25.0	0.6	55.5
Home connection	50.6	38.4	2.4	8.5

[4] These ratings include adjusted percentages for the items "Assessment using Reading Record" and "Writing about reading (interactive, dictated, or independent)" because these activities are only implemented during even-numbered lessons. Writing was rated "Acceptable" or "Excellent" in 86.5% of observed even-numbered lessons, while assessment was rated "Acceptable" or "Excellent" in 85.4% of observed even-numbered lessons. Table 22 shows the overall percentages for these items across both even- and odd-numbered lessons.

Table 22, continued

Item	Percent Responded			
	Excellent	Acceptable	Needs Improvement	Not Observed
Literacy Instructional Strategies				
Teacher models, encourages, and provides opportunities for fluent oral reading.	57.9	24.4	4.3	13.4
Teacher introduces vocabulary words (e.g., high frequency, story-specific words).	54.9	35.4	2.4	7.3
Teacher emphasizes understanding/comprehension of what is read.	67.1	24.4	6.7	1.8
Teacher models and encourages students to use appropriate reading strategies (e.g., phonemic awareness).	72.6	16.5	7.3	3.7
Teacher engages students in conversation about the text.	70.1	20.7	7.9	1.2
Teacher assists students in problem-solving.	75.6	15.9	7.3	1.2
Learning Environment				
Lesson is well organized.	78.0	17.1	4.9	0.0
Teacher appropriately paces lesson components.	59.8	30.5	9.8	0.0
Teacher engages in ongoing assessment of student learning (e.g., questioning, providing feedback/corrective instruction, checking responses).	84.1	9.1	6.7	0.0
Students are actively engaged.	79.3	14.6	6.1	0.0
Instructional modifications are observed when needed.	74.4	13.4	5.5	6.7
Instructional materials needed to implement lesson are readily available.	80.5	12.2	6.7	0.6
The lesson is delivered as designed.	64.0	31.1	4.9	0.0

Note: Item percentages may not total 100% due to missing input from some participants.
*These items were each observed approximately half of the time because they are only implemented during even-numbered lessons.

The LLIOT also included items designed to describe the groups observed, which are summarized in Table 23. Results from these items indicated that observers most frequently saw first grade groups (47.6%), followed by second grade (29.9%) and kindergarten (26.8%). The majority of observed groups took place in a designated intervention area (79.9%) and lasted approximately 30 minutes (72.0%), which was consistent with LLI's design. Further, nearly two-thirds of the observed groups had the recommended number of three students (59.8%), with most of the remaining groups having two students (31.1%). Finally, a relatively equal number of even- and odd-numbered lessons were observed (54.3% and 45.7%, respectively). All items can be found in Table 23 below.

Table 23: LLIOT Summary Items (n = 164)

Item	Percent Responded
Grade Level	
K	26.8
1	47.6
2	29.9
Location of Group	
Intervention Area	79.9
Classroom	4.3
Other	15.9
Number of Students in Group	
1	6.7
2	31.1
3	59.8
4 or more	2.4
Total Instructional Minutes	
Less than 25	8.5
25 – 35	72.0
More than 35	19.5
LLI Lesson Number	
Even	54.3
Odd	45.7

Note: Item percentages may not total 100% due to missing input or multiple responses from some participants.

Observers conducting the LLIOT also recorded open-ended comments summarizing the instructional materials used during the lesson and their perceptions of the quality of instruction, level of student participation, and overall success of the lesson. Observers' comments were summarized using a structured, multi-step process. First, the original comments were assigned codes representing their basic content. Next, these codes were grouped into categories, which were then organized into overarching themes. Final analysis produced frequency percentages for each theme. Because it was possible for some comments to contain multiple content codes, the percentages reported reflect the total number of codes within each theme and not necessarily the total number of comments received from observers. Observers' responses are summarized below.

Overall, 42.6% of the comments were related to the quality of literacy instruction, 28.6% were related to the lesson resources and materials, 21.2% were related to student participation and engagement in the lesson, and 7.7% were related to the success of the lesson. The majority of these comments (77.7%) were positive, while only 13.5% indicated areas that need to be addressed; further, 8.8% of the comments were neutral or descriptive in nature. In general, observers reported that the lessons were appropriately paced, well-organized, and delivered according to LLI guidelines; the teachers made effective use of such strategies as monitoring, questioning, prompts, and reinforcement; the students were actively engaged, motivated to learn, and enthusiastic; and a wide variety of instructional materials were readily available. However, in some lessons, observers noted poor time management, lack of enthusiasm on the part of the teacher, ineffective delivery of lesson components, off-task and unfocused student behavior, or inadequate use of lesson resources.

Table 24 provides a summary of the themes that arose from observers' comments and the percentage of responses for each. Sample comments from the observers are provided below.

"Lesson flowed very smoothly. Teacher used lots of materials (magnetic letters, boards, etc.) to reinforce instruction. Teacher was very positive and encouraging and did a great job with pacing and transitioning. Students were engaged and seemed to be learning and enjoying the lesson."

"Overall this was a very good lesson. The teacher has many strengths. She knows the students' strengths and weaknesses and is able to support them as needed. The students (3 girls) were enthusiastic and engaged the entire time. I thought the introduction to the new book and vocabulary was a bit brief. Perhaps this was because it was an independent book and/or because of the short attention span of kindergarten students. The teacher gave a general statement about the book and discussed one new vocabulary word. Materials used: white board, program guide, reading record, leveled books, writing books, word cards, take-home books and fold sheets."

"The lesson went well. The lesson observed was [an] odd lesson. Teacher introduced the book and did some letter name, sound and picture work. Students were actively engaged. Teacher asked questions to check for understanding. The time of the lesson was 21 minutes, a little shorter than advised for LLI. Teacher was well organized and paced the lesson. Materials used were small books, letter cards and pictures, and materials to go home (a worksheet and a book). I feel the lesson was a success."

Table 24: Thematic Summary of LLIOT Comments

Theme	Percent of Comments (Per Question)
Please provide a brief summary of your observation (include the instructional materials used and your perceptions of the quality of instruction, participation of the students, and success of the lesson):	
Quality of instruction	42.6
Lesson resources and materials	28.6
Student participation and engagement	21.2
Overall success of the lesson	7.7

Consistency of LLI Implementation

The LLIOT was conducted at both the beginning and the end of LLI for each LLI group containing at least one treatment group student in order to measure changes in implementation over time. For first and second grade, pretest observations were conducted in November 2011, and posttest observations were conducted in January 2012. For kindergarten, pretest observations were conducted in March 2012, and posttest observations were conducted in April and May 2012. Results are summarized by grade level below. Overall, it appears that LLI instruction remained consistent throughout the program for students at all three grade levels, with a slight improvement in literacy instructional strategies demonstrated in first grade.

Kindergarten

For each of the two occasions on which 22 kindergarten LLI groups were observed, the means and standard deviations presented in Table 25 were computed on the three subscales of the LLIOT. Subsequently, three independent *t*-tests that contrasted teacher behaviors at times one and two were conducted on the pairs of means obtained on the LLIOT's ten-item "Quality of LLI Implementation" scale ($t(42) = 0.67$, $p = 0.51$, $g = 0.20$), its six-item "Literacy Instructional Strategies" scale ($t(42) = 1.07$, $p = 0.29$, $g = 0.32$), and its seven-item "Learning Environment" scale ($t(42) = 1.14$, $p = 0.26$, $g = 0.34$), with no

statistically significant differences observed for any of the three comparisons. The average rating was between "Acceptable" (2.00) and "Excellent" (3.00) for each subscale at both time points.

First Grade

The descriptive statistics and independent *t*-test results for each of the three LLIOT subscales for the 39 observed first grade groups are presented in Table 25. There were no significant differences between the pretest and the posttest observations for two of the subscales: "Quality of LLI Implementation" ($t(76) = 0.19$, $p = 0.85$, $g = 0.04$) and "Learning Environment" ($t(76) = -0.72$, $p = 0.48$, $g = -0.16$). However, at a liberal level of statistical significance (that is, $p < .10$), scores on the "Literacy Instructional Strategies" subscale did appear to somewhat improve from pretest to posttest ($t(76) = -1.94$, $p = 0.06$, $g = -0.44$). For all subscales, the average rating was between "Acceptable" (2.00) and "Excellent" (3.00) at both time points.

Second Grade

Similar to the other two grades, the results of the three independent *t*-tests for the 20 second grade groups observed at the pretest and the 22 second grade groups observed at the posttest revealed that there were no significant differences between the two sets of observations for the subscales "Quality of LLI Implementation" ($t(40) = 1.11$, $p = 0.27$, $g = 0.34$), "Literacy Instructional Strategies" ($t(40) = -0.84$, $p = 0.41$, $g = -0.25$), and "Learning Environment" ($t(40) = 0.04$, $p = 0.97$, $g = 0.01$). The average rating was between "Acceptable" (2.00) and "Excellent" (3.00) for each subscale at both time points. Descriptive statistics and independent *t*-test results are summarized in Table 25 below.

Table 25: Independent *T*-Test Results for LLIOT Subscales by Grade Level

LLIOT Subscale	Pre-Test		Post-Test				
	M	*SD*	*M*	*SD*	*t*	*p*	*g*
Kindergarten (*n* = 22)							
Quality of LLI Implementation	2.59	0.33	2.51	0.48	0.67	0.51	0.20
Literacy Instructional Strategies	2.77	0.32	2.65	0.39	1.07	0.29	0.32
Learning Environment	2.87	0.24	2.75	0.41	1.14	0.26	0.34
1st Grade (*n* = 39)							
Quality of LLI Implementation	2.36	0.43	2.34	0.51	0.19	0.85	0.04
Literacy Instructional Strategies	2.41	0.50	2.63	0.53	-1.94	0.06†	-0.44
Learning Environment	2.61	0.48	2.69	0.51	-0.72	0.48	-0.16
2nd Grade (pretest *n* =20, posttest *n* = 22)							
Quality of LLI Implementation	2.27	0.63	2.07	0.55	1.11	0.27	0.34
Literacy Instructional Strategies	2.18	0.64	2.36	0.75	-0.84	0.41	-0.25
Learning Environment	2.51	0.44	2.51	0.69	0.04	0.97	0.01

†$p < .10$

Quality of Core Literacy Instruction

Literacy Observation Tool (LOT)

Descriptive Results

The Literacy Observation Tool (LOT) involved seven to nine 10-minute observations of K-2 core literacy instruction at each school (*n* = 26 LOTs, with each LOT including seven to nine individual classroom observations). Table 26 illustrates the frequencies for each item on the LOT, as observed during the visits. In all of the observed K-2 classrooms, the students were actively engaged, the teacher managed the classroom effectively, the alphabet was visibly displayed, and there was a classroom library of books (each "Frequently" or "Extensively" observed 100.0% of the time). The instructional orientation most commonly observed was a whole-class orientation, followed closely by small-group instruction ("Frequently" or "Extensively" observed 76.9% and 69.2% of the time, respectively). Additionally, in the majority of observed classrooms, the on-site researchers "Frequently" or "Extensively" noted a "word wall" of high-frequency or sight words (92.3%), active monitoring by the teacher (84.6%), an environment conducive to cooperative interactions (80.8%), and the use of fiction books (73.1%). In terms of instruction, the most commonly observed activities included students independently reading self-selected materials, explicit comprehension strategy instruction, and teacher-guided interactive discussion ("Frequently" or "Extensively" observed 50.0%, 38.5%, and 38.5% of the time, respectively). The least frequently observed areas included the use of basal texts, portfolios, or periodicals (each "Rarely or Not Observed" 96.2% of the time), assessment with the Informal Reading Inventory (IRI) or running records ("Rarely or Not Observed" 88.5% of the time), and all items related to the teacher's role in writing instruction – letter formation/handwriting, the writing process, language mechanics lessons, conferencing with students, and providing for students sharing (all "Rarely or Not Observed" between 76.9% and 88.5% of the time). All items can be found in Table 26 below.

Table 26: LOT Response Frequencies (*n* = 26)

Item	Percent Responded		
	Frequently or Extensively	Occasionally	Rarely or Not Observed
Instructional Orientation			
Small group	69.2	11.5	19.2
Whole class	76.9	15.4	7.7
Learning centers	50.0	26.9	23.1
Cooperative/collaborative learning	19.2	19.2	61.5
Concepts of Print			
Book/print conventions	23.1	7.7	69.2
Alphabetics			
Letter naming/knowledge	11.5	19.2	69.2
Phonemic awareness instruction	15.4	15.4	69.2
Rhyming	3.8	19.2	76.9
Explicit phonics instruction	26.9	3.8	69.2
Fluency			
Models fluent oral reading	26.9	23.1	50.0
Has student(s) read/reread orally (together)	19.2	7.7	73.1
Vocabulary			
Introduces/reviews key vocabulary	26.9	23.1	50.0
Explicit vocabulary instruction	11.5	15.4	73.1

Table 26, continued

Item	Percent Responded		
	Frequently or Extensively	Occasionally	Rarely or Not Observed
Text Comprehension			
Explicit comprehension strategy instruction	38.5	30.8	30.8
Makes connection to prior knowledge	15.4	38.5	46.2
Asks students for predictions	23.1	7.7	69.2
Uses higher level questioning	26.9	26.9	46.2
Guides visual imaging	15.4	15.4	69.2
Guides interactive discussion	38.5	30.8	30.8
Independent Reading – The Student:			
Reads self-selected materials	50.0	30.8	19.2
Writing – The Teacher:			
Letter formation/handwriting	3.8	7.7	88.5
Writing process	7.7	7.7	84.6
Language mechanics lessons	7.7	11.5	80.8
Conference with students	11.5	11.5	76.9
Provides for students sharing	11.5	11.5	76.9
Writing – The Student:			
Writes independently	30.8	23.1	46.2
Response writing	30.8	19.2	50.0
Assessment			
Formal testing	11.5	7.7	80.8
Portfolios	0.0	3.8	96.2
Informal Reading Inventory (IRI), running records	11.5	0.0	88.5
Learning Environment			
Conducive to cooperative interactions	80.8	11.5	7.7
Students actively engaged	100.0	0.0	0.0
Effective classroom management	100.0	0.0	0.0
Teacher actively monitors	84.6	11.5	3.8
Visible Print Environment			
Alphabet	100.0	0.0	0.0
Word wall	92.3	3.8	3.8
Labeling (names, objects, areas)	50.0	11.5	38.5
Classroom library	100.0	0.0	0.0
Evidence of student writing/work products	61.5	26.9	11.5
Materials Used			
Basal texts	3.8	0.0	96.2
Big books	11.5	15.4	73.1
Books on tape	3.8	15.4	80.8
Computers	23.1	38.5	38.5
Fiction books	73.1	7.7	19.2
Non-fiction books	42.3	19.2	38.5
Poetry	11.5	15.4	73.1
Newspaper/magazines	0.0	3.8	96.2
Word/vocabulary materials	23.1	30.8	46.2
Worksheets/workbooks	11.5	30.8	57.7
Other materials used	42.3	26.9	30.8

Note: Item percentages may not total 100% due to missing input from some participants.

Observers conducting the LOT were also asked to share comments related to the strengths and areas for improvement they noted across the observed classrooms during their school visit, as well as notes on the students' progress and recommendations for next steps. On-site researchers' open-ended responses were analyzed using the same structured, multi-step process employed for the LLIOT comments and are summarized by question below.

When asked to describe the strengths of the classroom literacy programs observed, on-site researchers most frequently listed instructional strategies as the prevailing strength in the classrooms (41.6% of comments). Of these, teacher preparation and small group instruction were most often cited (12.9% each), followed by effective instruction and learning centers (11.3% each). Another theme that emerged, accounting for 23.5% of the comments related to strengths, involved the classroom environment. More than half (51.4%) of these comments noted positive classroom environments. On-site researchers also observed that classroom materials (14.3%) and print-rich environments (8.6%) helped contribute to the quality of classrooms. In addition, observers listed student engagement and compliance as another fundamental theme when discussing strengths (21.5% of comments). They reported that students understood and followed the classroom routines and that students were engaged in the activities (46.9% each). Finally, on-site researchers noted the use of literacy activities as a strength in 13.4% of comments. Of these, reading (30.0%) and writing (25.0%) were most commonly mentioned. Overall, on-site researchers generally indicated that the observed schools had strong literacy programs.

When asked to discuss concerns they had regarding the classroom literacy lessons that they observed, 44.1% of on-site researchers' responses reflected the theme of general instructional strategies. Of these, observers most frequently noted cooperative learning and student discussion as lacking (14.6% of comments in this theme). On-site researchers also stated that they had seen little or no small group instruction (12.2%). Another theme regarding on-site researchers' concerns was that of literacy activities, which accounted for 37.6% of comments. Of these, the most frequently cited concern was in the area of writing (28.6%); other categories included reading (14.3%), a lack of explicit vocabulary lessons, little or no guided reading, and minimal or no labeling of items in the classrooms (each accounting for 11.4% of comments within the theme of literacy activities). Further, classroom environment was another concern for on-site researchers (12.9% of comments). Within these comments, the concern that on-site researchers most frequently referenced was that of a lack of student work on display, which made up 41.7% of comments in this theme. Finally, while accounting for the smallest percentage of the comments compared to other themes, assessment was also mentioned as a concern for on-site researchers (5.4% of comments).

When on-site researchers were asked to discuss student progress and next steps regarding the classrooms that they observed, the most frequent theme was instructional strategies (31.7% of comments), within which the most commonly listed category was critical thinking (20.0% of these comments). Literacy was another area where on-site researchers saw potential for progress, as this theme accounted for 25.0% of responses for this question. On-site researchers pointed out the need for explicit vocabulary lessons (21.1% of comments) and an emphasis on writing (15.8% of comments). Further, the observers directly stated that students generally had made progress in 20.3% of comments. Three of these comments specified reading and writing as areas where students had made advancements, while two of these comments specified that teachers are pleased with the progress made by students in the LLI program. Finally, observers listed classroom environment (e.g., displaying student work, expanding classroom libraries; 11.8% of responses), assessment (e.g., regular progress

monitoring; 7.9% of responses), and student engagement (5.1% of responses) as future areas needing to be addressed for the observed classrooms.

Table 27 provides a summary of the themes that arose from observers' comments and the percentage of responses for each. Sample comments from the observers are provided below.

Strengths:

"Word walls, alphabet, classroom libraries [were] in every room. Most students utilized book bags with self-selected books and were engaged in independent reading. Most centers/stations were focused on literacy learning: reading, writing, [and] spelling. Most classrooms displayed effective classroom management and a positive learning environment. Teachers used a variety of instructional approaches: whole group, small group, and independent."

"Little teacher monitoring was needed as far as on task behavior. In centers there was reading, writing, word work, oral language and reading instruction taking place. Learning environments were very pleasant and active."

Concerns:

"Students' creative/critical thinking response to the learning environment is minimal. There is little talk among students, no projects (student made charts/diagrams) or published books/writing, and little sharing of ideas or discussion of big ideas. [There was] little evidence of deep thinking."

"[I] never saw small group instruction in a couple of classrooms."

Progress:

"The children I observed seemed to be exposed to the reading components and were actively involved in their reading class."

"Students were reading and writing at a higher level."

Next Steps:

"A reading program that progresses from Kindergarten to 6th grade is needed. Each teacher is teaching what he/she has been determined for that grade level, as far as I could tell. Having a specific program would help insure all aspects of reading per grade level would be covered."

"Monitoring all students through informal testing could give the teacher a focus for planning lessons to meet needs of students and increase learning opportunities."

Table 27: Thematic Summary of LOT Comments

Theme	Percent of Comments (Per Question)
Please list strengths.	
Instructional strategies	41.6
Classroom environment	23.5
Student engagement and compliance	21.5
Literacy activities	13.4
Please list concerns.	
Instructional strategies	44.1
Literacy activities	37.6
Classroom environment	12.9
Assessment	5.4
How well are the child(ren) progressing? What are/should be the next steps?	
Instructional strategies	31.7
Literacy activities	24.1
General statements of progress	20.3
Classroom environment	11.4
Assessment	7.6
Student engagement and compliance	5.1

School and Home Support for Literacy

Leveled Literacy Intervention Teacher Questionnaire – Revised (LLITQ-R)

The Leveled Literacy Intervention Teacher Questionnaire – Revised (LLITQ-R) was administered online to LLI teachers at the end of the school year as a general measure of their implementation and perceptions of LLI (*n* = 21 respondents). Table 28 illustrates the frequencies of responses for each item on the LLITQ-R. Most of the respondents reported positive perceptions of LLI and its implementation in their individual schools. LLI teachers indicated that they had a good understanding of LLI; received support in implementing LLI from their district, school administration, and other school staff; and perceived a positive impact on student achievement and student attitudes towards literacy. LLI teachers also reported a positive impact of LLI on their reading instruction, particularly their understanding of the role of comprehension in successful reading, the reading process in general, and the relationship of leveled texts to successful reading. Further, LLI teachers reported implementing the LLI system with a high degree of fidelity; the majority of teachers indicated that they met with their groups daily for at least 30 minutes, followed the lesson design, and implemented both reading and writing activities.

Overall, LLI teachers were most likely to "Agree" or "Strongly Agree" that they understand the goals of LLI, receive guidance and support from their instructional and administrative staff to implement LLI, believe LLI has positively impacted participating students' literacy achievement, and find LLI to be aligned with their state and district reading and language arts standards (each 100.0%). LLI teachers were least likely to report that their schools have sufficient faculty and staff to provide LLI to all students who need it (57.2% "Disagree" or "Strongly Disagree"), that LLI helps their students with special needs (57.1% "Somewhat" or "Not At All"), that the parents of their LLI students participate in home literacy activities with their children (28.6% "Disagree" or "Strongly Disagree"), and that their teaching schedules allow time to implement LLI effectively (28.6% "Somewhat" or "Not At All"). All of the surveyed

teachers (100.0%) agreed that their school should continue LLI. All items can be found in Table 28 below.

Table 28: LLITQ-R Response Frequencies (n = 21)

Item	Percent Responded		
	Strongly Agree/Agree	Neutral	Disagree/ Strongly Disagree
I understand the goals of the Leveled Literacy Intervention (LLI) program.	100.0	0.0	0.0
I have received adequate professional development for implementing LLI.	90.5	4.8	4.8
I have a thorough understanding of how to implement LLI.	90.5	4.8	4.8
Guidance and support are provided by our instructional and administrative staff to help us implement LLI.	100.0	0.0	0.0
I believe LLI has positively impacted LLI students' literacy achievement.	100.0	0.0	0.0
LLI teachers are given sufficient planning time to implement the system.	81.0	9.5	9.5
Students who receive LLI in this school are more enthusiastic about reading, writing, and learning because of LLI.	95.2	0.0	4.8
Our school has sufficient faculty and staff to provide LLI to all students who need the intervention.	38.1	4.8	57.2
Our administration protects the time needed for daily uninterrupted LLI teaching.	71.5	4.8	19
Our students' parents participate in LLI home literacy activities with their child(ren).	42.8	28.6	28.6
Teachers in this school are generally supportive of LLI.	95.3	0.0	4.8
Ongoing communication exists between LLI teachers and classroom teachers.	66.7	9.5	19
LLI teachers are encouraged to communicate concerns, questions, and constructive ideas regarding the system to school staff or administration.	76.2	9.5	9.5
LLI allows for teachers to provide differentiated instruction to address the varying strengths and needs of students.	90.5	9.5	0.0
Instructional materials (books, assessments, and other resources) needed to implement LLI are readily available.	95.3	4.8	0.0
The faculty, staff, and administration in my school believe that all children can learn to read and write.	95.3	0.0	4.8
LLI is aligned with state and district reading and language arts standards.	100.0	0.0	0.0
LLI training has improved my reading instruction.	95.2	4.8	0.0
LLI students perform better on state assessments as a result of their participation in LLI.	66.6	33.3	0.0
Because of LLI, I have a greater understanding of…			
The reading process.	95.2	4.8	0.0
The characteristics of leveled books and their relationship to successful reading.	95.2	4.8	0.0
How to improve children's vocabulary and oral language skills.	81.0	9.5	9.5
The role of fluency in effective reading.	90.5	4.8	4.8
The role of phonics and phonemic awareness in the reading process.	90.5	0.0	9.5
The role of comprehension in successful reading.	95.3	0.0	4.8
How to improve children's writing strategies.	90.5	0.0	9.5

Table 28, continued

Item	Percent Responded		
	Extensively/ Sufficiently	Somewhat	Not At All
To what degree does your school administration support your efforts as an LLI teacher?	95.3	4.8	0.0
To what degree does the district support your efforts as an LLI teacher?	85.7	14.3	0.0
To what degree does your teaching schedule allow time to implement LLI effectively?	71.4	28.6	0.0
To what extent do you feel LLI has helped your English Language Learner students?	80.9	19.0	0.0
To what extent do you feel LLI has helped your students with special needs?	42.9	57.1	0.0

Item	Percent Responded		
	Regularly (Every day)/ Frequently (3-4 days per week)	Occasionally (1-2 days per week)	Rarely (Less than 1 day per week)/Not At All (Never)
How often did your LLI group lessons last 30 minutes or more?	81.0	14.3	4.8
Were you able to meet every day with your LLI group(s)?	95.2	0.0	4.8
How often did you follow the LLI lessons exactly as instructed in the Lesson Guide?	95.2	4.8	0.0
How often were you able to implement LLI reading activities, such as phonics/word work and guided reading?	100.0	0.0	0.0
How often were you able to implement LLI writing activities, such as interactive writing?	95.2	4.8	0.0

Item	Percent Responded
Do you think your school should continue the Leveled Literacy Intervention program?	
Yes	100.0
No	0.0

Note: Item percentages may not total 100% due to missing input from some participants.

The LLITQ also invited LLI teachers to share open-ended comments regarding the strengths and areas for improvement of LLI and the reasons that their schools should continue or not continue using the LLI system. Participants' responses to these items were summarized using a structured, multi-step process. First, the original comments were assigned codes representing their basic content. Next, these codes were grouped into categories, which were then organized into overarching themes. Final analysis produced frequency percentages for each theme. Because it was possible for some comments to contain multiple content codes, the percentages reported reflect the total number of codes within each theme and not necessarily the total number of comments received from participants. Participating LLI teachers' responses are summarized by question below.

When participating LLI teachers were asked about the strengths of LLI, their responses most frequently focused on the design and organization of the LLI system (42.9%). Of these comments, 22.2% referenced the fact that LLI contains all of the components of literacy (e.g., the "Big Five"), 22.2% mentioned the consistency and structure provided by the lesson routine, and 11.1% highlighted the pattern of alternating easier "independent level" books with harder "instructional level" books. The next most common theme regarding strengths involved the LLI system's materials and resources (30.2% of overall comments) – particularly the books, which were mentioned in nearly half of these responses (47.4%). Other resources mentioned included the supporting materials (e.g., lesson resources,

prompting guide; 15.8%) and ESL resources (10.5%). Finally, in nearly a quarter of overall responses (23.8%), the participating teachers discussed the instructional components of LLI; these responses most frequently cited the writing component, the instructional approach in general, and the word work (26.7%, 20.0%, and 13.3%, respectively).

Regarding areas for improvement of the LLI system, participating LLI teachers discussed the program design and implementation in approximately one-third of comments (32.4%). Within this theme, teachers most frequently discussed the need for more time during lessons (particularly on even-numbered lesson days; 54.6%) and less paper use (18.2%); other comments included the cost of the LLI kits, the need to start the program earlier in the year, and the desire to have more freedom for teacher creativity in the lessons. Another theme that frequently arose regarding areas for improvement involved specific strategies and instructional components (29.4% of overall comments), including writing (30.0%) and word work (30.0%), which teachers felt should be more aligned with the lessons. Further, 17.7% of teachers' overall comments were related to LLI resources, specifically the online data management system (66.7%) and the need for continued training (33.3%). Finally, in 14.7% of responses, the participating LLI teachers stated that they did not have any suggestions or shared positive feelings about LLI.

Finally, the participating LLI teachers were asked why their school should continue or not continue using LLI. Responses most frequently focused on LLI's effectiveness in improving literacy skills (40.8%), with the majority of these comments (70.0%) related to the growth and progress teachers have seen in their students. One teacher stated that she has especially seen improvement in her ELL students, while others stated that LLI is effective in the long term as well as the short term and helps prevent the need for intervention later in life. Other reasons cited by LLI teachers for continuing LLI included the program design (22.5%), instructional components and materials (18.4%), and positive effects for students and teachers (14.3%). Regarding program design, teachers most frequently mentioned the targeted instruction (27.3%), small-group format (which is non-threatening and allows teachers to closely monitor students; 18.2%), and comprehensive nature of the system (18.2%). With regard to instructional components and materials, teachers most frequently discussed the available materials (which make planning easier and less time-consuming), the high-quality books, and the take-home books for the students (each 22.2%). Finally, positive effects for students and LLI teachers included increased student confidence, students' love of LLI, and improved teacher effectiveness (each 28.6%). No teachers provided a reason for their school to discontinue the use of LLI.

Table 29 provides a summary of the themes that arose in response to each open-ended question and the percentage of responses for each. Sample comments from LLI teachers are provided below; all comments are provided in Appendix K.

Strengths:

"[LLI] covers the big 5 in comprehensive literacy. Writing and reading are highly intertwined and build off one another."

"[LLI is] very structured and easy to follow. The books are excellent and the students love them."

Areas for Improvement:

"The data management software, while very useful, does not work well consistently and takes a lot of time to use. The masters of the word cards should not waste so much paper. It would be very helpful if they filled up each by repeating the words over and over. I waste a lot of paper uselessly. The Getting Started Lessons in the Orange and Green kits are virtually impossible to do in one day with students new to the program and new to intervention. I think they should be there, but use the regular odd and even format to make it more realistic to get through in one session. Lastly, I wish the kits were more affordable. We would love to get the Red kit, but the cost is way too high for us to afford."

Reasons to Continue LLI:

"Over the past two years I have seen great growth in the students that have participated [in LLI]. They have either been able to achieve grade level reading or have made at least a year's growth."

"We're seeing students' reading and writing skills improve, especially the ELL students that we serve. Students are proud of becoming readers and writers and are excited to come to LLI lessons."

Table 29: Thematic Summary of LLITQ-R Comments

Theme	Percent of Comments (Per Question)
What are the strengths of LLI?	
Program design and organization	42.9
Materials and resources	30.2
Instructional components	23.8
Blank	3.2
What areas of LLI could be improved?	
Program design and implementation	32.4
Instructional components and strategies	29.4
Resources	17.7
No suggestions/likes LLI	14.7
Blank	5.9
Why should your school continue or not continue LLI?	
Effective in improving literacy skills	40.8
Program design	22.5
Instructional components and materials	18.4
Positive effects for students and teachers	14.3
Blank	4.1

Classroom Teacher Literacy Instruction Questionnaire (CTLIQ)

The Classroom Teacher Literacy Instruction Questionnaire (CTLIQ) was administered online to regular K-2 classroom teachers at the end of the school year as a general measure of classroom teachers' literacy instructional strategies and perceptions of LLI and the core literacy program at their schools (n = 51 respondents). Table 30 illustrates the frequencies of responses for each item on the CTLIQ. Results from the CTLIQ revealed that classroom teachers of both treatment and control group students were most likely to "Agree" or "Strongly Agree" that they implement writing activities, integrate both vocabulary and comprehension into their literacy instruction, and utilize high-quality literature to read to students and engage them in interactive discussions about the text (rated

"Regularly" or "Frequently" 96.1%, 94.2%, and 94.1% of the time, respectively). Teachers also frequently reported using both whole-group instruction and individual or small-group instruction (each rated "Regularly" or "Frequently" 90.2% of the time). Teachers were least likely to report placing students in small, heterogeneous groups to discuss the books they are reading and assigning home literacy activities for students to complete with parents (each rated "Rarely" or "Not At All" 15.7% of the time).

Overall, the participating K-2 classroom teachers reported a generally positive perception of their school's core literacy program, although some areas of concern were identified. Teachers were most likely to "Agree" or "Strongly Agree" that they understand the goals of their school's core literacy program (94.1%) and that the members of their school community believe that all children can learn to read and write (92.2%). Additionally, some of the highest areas of agreement focused on LLI; the majority of classroom teachers "Agreed" or "Strongly Agreed" that LLI supports the goals of their school's core literacy program (82.4%), that they have ongoing communication with LLI teachers (82.4%), and that students who participate in LLI show increased literacy achievement (88.3%), participation in classroom literacy instruction (82.4%), and enjoyment of reading (82.3%). However, less than two-thirds of the respondents indicated that they have a clear understanding of LLI (62.8% "Agree" or "Strongly Agree"). Regarding less positive areas of their school's core literacy program, classroom teachers were least likely to agree that the program helps their students with special needs (66.7% "Somewhat" or "Not At All") and ELL students (51.0% "Somewhat" or "Not At All"), that the district supports their efforts to implement the core literacy program (41.1% "Somewhat" or "Not At All"), and – similar to the LLI teachers – that their students' parents participate in home literacy activities with their children (45.1% "Disagree" or "Strongly Disagree"). Just over two-thirds of the participating teachers agreed that their school should continue the current core literacy program (68.6%). All items can be found in Table 30 below.

Table 30: CTLIQ Response Frequencies (n = 51)

Item	Percent Responded		
	Strongly Agree/Agree	Neutral	Disagree/ Strongly Disagree
I understand the goals of our school's core literacy program.	94.1	0.0	3.9
I have received adequate professional development for implementing our school's core literacy program.	80.4	9.8	7.8
I have a thorough understanding of how to implement our school's core literacy program.	76.5	15.7	5.9
Guidance and support are provided by our instructional and administrative staff to help us implement our core literacy program.	76.5	17.6	3.9
I believe our core literacy program has positively impacted students' literacy achievement.	66.6	15.7	13.8
Teachers are given sufficient planning time to fully implement our school's core literacy program.	31.3	29.4	37.2
Students in this school are more enthusiastic about reading, writing, and learning because of our core literacy program.	47.1	27.5	21.5
Our school has sufficient faculty and staff to fully implement its core literacy program.	58.8	17.6	21.6
Our administration protects the time needed for daily uninterrupted core literacy instruction.	60.8	23.5	13.7
Our students' parents participate in home literacy activities with their child(ren).	31.4	21.6	45.1
Teachers in this school are generally supportive of our core literacy program.	74.5	13.7	9.8

Table 30, continued

Item	Percent Responded		
	Strongly Agree/Agree	Neutral	Disagree/ Strongly Disagree
Teachers are encouraged to communicate concerns, questions, and constructive ideas regarding our core literacy program to school staff or administration.	62.7	19.6	15.7
Our core literacy program allows for teachers to provide differentiated instruction to address the varying strengths and needs of students.	66.7	17.6	13.7
Instructional materials (books, assessments, and other resources) needed to implement our core literacy program are readily available.	53.0	7.8	37.2
The faculty, staff, and administration in my school believe that all children can learn to read and write.	92.2	3.9	2.0
Our core literacy program is aligned with state and district reading and language arts standards/frameworks.	80.4	11.8	5.9
Professional development for our school's core literacy program has improved my reading instruction.	62.8	13.7	21.5
Our core literacy program adequately prepares our students for state assessments.	49.0	23.5	25.5
I have a clear understanding of the Leveled Literacy Intervention (LLI) program.	62.8	19.6	15.7
LLI supports the goals of my school's core literacy program.	82.4	15.7	0.0
Ongoing communication exists between LLI teachers and classroom teachers.	82.4	9.8	5.9
Students who participate in LLI show increased enjoyment of reading and writing.	82.3	7.8	7.8
Students who participate in LLI show increased achievement in literacy.	88.3	2.0	7.8
Students who participate in LLI show increased participation in classroom literacy activities and instruction.	82.4	7.8	5.9

Item	Percent Responded		
	Extensively/ Sufficiently	Somewhat	Not at all
To what degree does your school administration support your efforts to implement your school's core literacy program?	70.6	27.5	0.0
To what degree does the district support your efforts to implement your school's core literacy program?	47.1	47.1	0.0
To what degree does your teaching schedule allow time to implement your school's core literacy program effectively?	58.8	31.4	5.9
To what extent do you feel your school's core literacy program has helped your English Language Learner students?	45.1	49.0	2.0
To what extent do you feel your school's core literacy program has helped your students with special needs?	31.4	54.9	11.8

Item	Percent Responded		
	Regularly (Every day)/ Frequently (3-4 days per week)	Occasionally (1-2 days per week)	Rarely (Less than 1 day per week)/Not At All (Never)
Students participate in whole group reading instruction.	90.2	2.0	5.9
Students participate in small group or individual reading instruction.	90.2	7.8	0.0
I provide guided reading instruction using leveled texts for groups of students with similar learning needs.	88.2	9.8	0.0
Students meet in small, heterogeneous groups to discuss the books that they are reading.	68.7	13.7	15.7
Students participate in writing activities, such as mini-lessons, independent writing, conferencing, and sharing.	96.1	2.0	0.0

Table 30, continued

Item	Percent Responded		
	Regularly (Every day)/ Frequently (3-4 days per week)	Occasionally (1-2 days per week)	Rarely (Less than 1 day per week)/Not At All (Never)
I provide opportunities to develop oral reading fluency (e.g., shared reading, partner reading).	88.3	9.8	0.0
I teach phonological awareness (sound patterns, rhymes, etc.) to my students.	84.3	9.8	3.9
I integrate both vocabulary and comprehension into my literacy instruction and activities.	94.2	3.9	0.0
I read high-quality children's literature (e.g., fiction, non-fiction, poetry) to my students and engage them in interactive discussions about the text.	94.1	2.0	2.0
I assign students home literacy activities to encourage parent participation.	60.8	21.6	15.7

Item	Percent Responded
Do you think your school should continue the current core literacy program?	
Yes	68.6
No	29.4

Note: Item percentages may not total 100% due to missing input from some participants.

The CTLIQ also invited classroom teachers to share open-ended comments regarding the strengths and areas for improvement of their school's core literacy program and the reasons that their school should continue or not continue the core literacy program. Classroom teachers' open-ended responses were analyzed using the same structured, multi-step process employed for the LLI teacher comments and are summarized by question below.

When participating K-2 classroom teachers were asked about the strengths of their school's core literacy program, nearly half of the responses (45.6%) focused on the program's instructional components, particularly small group instruction (19.2%) and guided reading (11.5%); further, two teachers specifically mentioned LLI as a strength of the program. In over a third of the total responses (36.8%), teachers positively discussed general characteristics of the program, such as its ability to provide differentiated instruction to meet individual students' needs (19.1%), the support for ELL students (19.1%), the flexibility and autonomy afforded to teachers (14.3%), and the degree of buy-in and support for the program (14.3%). Finally, in 15.8% of the overall responses, teachers described the available resources as a strength of the core literacy program. Over half of the resources mentioned were materials (e.g., books, leveled texts, activities; 55.6%), the teaching staff comprised one-third (classroom teachers, intervention team; 33.3%), and one teacher mentioned the professional development opportunities.

Regarding areas for improvement for the core literacy program, teachers most frequently commented on the organization and delivery of the curriculum (39.7% of responses). Of these comments, nearly a third (32.0%) were related to the general curriculum (e.g., some aspects are not rigorous enough, while others "are over [the students'] heads"). Other categories within this theme included a lack of consistency between grade levels or classrooms (20.0%), ineffective instructional units (16.0%), an inadequate skills block (12.0%), the need to be more comprehensive (8.0%), and problems with student scheduling (e.g., pulling students for literacy support during the literacy block; 8.0%).

Further, over a third of the overall responses involved instruction (34.9%), with over half of these comments (54.6%) related to such instructional components as shared reading or phonics. Additional categories within this theme included suggestions to increase support for struggling students (18.2%), provide specialized instruction for student subgroups (e.g., second language learners or students with special needs; 13.6%), improve teaching strategies (9.1%), and slow the pacing to allow more time for concept mastery (4.6%). Finally, classroom teachers cited resources as an area needing improvement in 23.8% of comments, including materials (e.g., books, Spanish resources; 46.7%), professional development (20.0%), and the need for increased planning time (20.0%) and support staff (13.3%).

Finally, participating K-2 classroom teachers were asked why their school should continue or not continue using the current core literacy program. Overall, 67.8% of the comments were shared by respondents who believed the program should be continued, while 32.2% were shared by respondents who felt the program should be discontinued. Of the respondents who supported the program, 75.0% of the comments involved positive perceptions of the program, such as the belief that it is effective (33.3%) and a good core program (16.7%), meets students' needs (20.0%), and contains such beneficial instructional components as small groups, interdisciplinary units, guided reading, and a "good mix of phonics and whole language" (13.3%). However, these respondents also felt that certain areas of the program need improvement, such as the need to be more comprehensive and rigorous (40.0%), provide more teacher support (20.0%), and improve resources (20.0%). Additionally, one respondent reported that more students need LLI. Of the respondents who did not want to continue the core literacy program, 100.0% of the comments involved negative perceptions of the program. These responses recommended curricular modifications (e.g., alignment with the standards, more emphasis on phonics and reading strategies, greater comprehensiveness; 36.8%), stated that the program does not meet students' needs (26.3%), and described the program as unorganized (21.1%) and the resources as inadequate (10.5%).

Table 31 provides a summary of the themes that arose in response to each open-ended question and the percentage of responses for each. Sample comments from K-2 classroom teachers are provided below; all comments are provided in Appendix L.

Strengths:

"It allows for differentiation through small groups and conferencing and supports students' creativity."

"The reading [and] the writing are linked to support each other. For example when we are teaching the genre of how to texts, the children are exposed to them during reading time and during writing they are practicing writing how to texts."

Areas for Improvement:

"The worst part of my day is when my teammate and I are creating instruction for students learning their second language. There [are] no guidelines and we are inventing the wheel every day. Also, the resources in Spanish are limited. I want to use LLI for guided reading. I know it's a tier two intervention but most of my kids need help. Even if English is their first language, many are missing academic language. I love the fact that it has most of the components of reading and everything is right there at your fingertips instead of chasing down materials."

Reasons to Continue or Not Continue the Core Literacy Program:

"The program provides a solid foundation for progression through the grades."

"It does not effectively offer paths of differentiation for our second language learners and for our students with special needs. Furthermore, it does not integrate reading skills to make them more applicable to real world reading. Our literacy block feels isolated from our skills block. I'd like to see those two merged."

Table 31: Thematic Summary of CTLIQ Comments

Theme	Percent of Comments (Per Question)
What are the strengths of your school's core literacy program?	
Instructional components	45.6
Positive program characteristics	36.8
Resources	15.8
Not sure	1.8
What areas of your school's core literacy program could be improved?	
Curriculum organization and delivery	39.7
Instruction	34.9
Resources	23.8
Not sure	1.6
Why should your school continue or not continue the current core literacy program?	
Positive perceptions of the program	50.9
Areas for improvement/reasons to discontinue	49.2

Leveled Literacy Intervention Principal Questionnaire (LLIPQ)

The Leveled Literacy Intervention Principal Questionnaire (LLIPQ) was administered online to the principals of the schools participating in the study at the end of the school year as a general measure of their perceptions of LLI, its implementation at their schools, and its fit within their schools' literacy curriculum (n = 7 respondents). Table 32 illustrates the frequencies of responses for each item on the LLIPQ. Most of the participating principals shared positive perceptions of LLI, reporting that they support LLI in their schools, that LLI has positively impacted students' literacy achievement, and that LLI has improved their school's ability to provide literacy instruction to meet the needs of low-performing students. Principals also indicated a high degree of support for literacy in their schools and felt confident in their understanding of the goals of LLI and the core literacy program; however, only slightly more than half of the participating principals "Agreed" or "Strongly Agreed" that LLI and their school's core literacy program complement one another (57.2%).

Overall, participating principals were most likely to "Agree" or "Strongly Agree" that they understand the goals of their school's literacy program (100.0%), that the members of their school community believe that all children can learn to read and write (100.0%), and that they help protect the time needed for daily uninterrupted LLI instruction (85.8%). However, the majority of respondents reported that they believe their school's literacy program needs improvement (85.7% "Agree" or "Strongly Agree"), and only approximately half "Agreed" or "Strongly Agreed" that they have sufficient resources (e.g., materials, time, space, personnel) to implement either LLI or their core literacy program effectively (42.9% and 57.2%, respectively). Over a quarter of participating principals "Disagreed" or "Strongly Disagreed" that their students' parents participate in home literacy activities with their

children and that their school's literacy program has positively impacted student achievement (each 28.6%). Further, over two-fifths of respondents felt that the district only "Somewhat" supports LLI at their schools and that LLI "Somewhat" helps their ELL students and students with special needs (each 42.9%). However, when asked if LLI should be continued at their schools, all participants chose "Yes" (100.0%). All items can be found in Table 32 below.

Table 32: LLIPQ Response Frequencies (n = 7)

Item	Percent Responded		
	Strongly Agree/Agree	Neutral	Disagree/ Strongly Disagree
I understand the goals of our school's literacy program.	100.0	0.0	0.0
Our literacy program has positively impacted student achievement.	57.2	14.3	28.6
Our students' parents participate in home literacy activities with their child(ren).	71.4	0.0	28.6
I encourage all teachers to communicate concerns, questions, and constructive ideas regarding the literacy program to school staff or administration.	71.4	14.3	14.3
Our school has sufficient resources (instructional materials, time, space, personnel) to implement our literacy program.	57.2	14.3	28.6
The faculty, staff, and administration in my school believe that all children can learn to read and write.	100.0	0.0	0.0
Our core literacy program and the Leveled Literacy Intervention (LLI) program complement each other.	57.2	28.6	14.3
Our literacy program needs improvement.	85.7	0.0	14.3
I understand the goals of the Leveled Literacy Intervention (LLI) program.	85.7	0.0	0.0
I am familiar with the instructional components of LLI.	57.2	14.3	14.3
I support the LLI program in my school.	85.7	0.0	0.0
Our teachers have received enough guidance and professional development to implement LLI.	71.4	0.0	14.3
I believe LLI has positively impacted LLI students' literacy achievement.	85.7	0.0	0.0
LLI teachers are given sufficient planning time to implement the program.	71.4	0.0	14.3
Students who receive LLI in this school are more enthusiastic about reading, writing, and learning because of LLI.	57.1	28.6	0.0
Our school has sufficient faculty and staff to provide LLI to all students who need the intervention.	42.9	0.0	42.9
I help protect the time needed for daily uninterrupted LLI teaching.	85.8	0.0	0.0
Parents at our school support the LLI program.	42.9	42.9	0.0
Teachers in this school are generally supportive of LLI.	71.5	14.3	0.0
Ongoing communication exists between LLI teachers and classroom teachers.	57.2	14.3	14.3
LLI allows for teachers to provide differentiated instruction to address the varying strengths and needs of students.	57.2	28.6	0.0
Our school has sufficient resources (instructional materials, time, space, personnel) to implement LLI.	42.9	0.0	42.9
LLI is aligned with state and district reading and language arts standards.	42.9	42.9	0.0
LLI professional development training has improved reading instruction at our school.	57.1	28.6	0.0
LLI students perform better on state assessments as a result of their participation in LLI.	28.6	42.9	14.3
LLI has improved our school's ability to provide literacy instruction to meet the needs of low-performing students.	85.7	0.0	0.0

Table 32, continued

Item	Percent Responded		
	Extensively/ Sufficiently	Somewhat	Not at all
To what degree does the district support the LLI program at your school?	57.1	42.9	0.0
To what extent do you feel LLI has helped your English Language Learner students?	57.2	42.9	0.0
To what extent do you feel LLI has helped your students with special needs?	57.1	42.9	0.0

Item	Percent Responded
Do you think your school should continue the Leveled Literacy Intervention program?	
Yes	100.0
No	0.0

Note: Item percentages may not total 100% due to missing input from some participants.

Principals who completed the LLIPQ were also asked to discuss their perceptions of the strengths and areas for improvement of LLI, reasons that their school should continue or not continue using LLI, challenges to LLI implementation, and additional resources needed for successful implementation, as well as other literacy interventions in use at their schools and LLI's efficacy in comparison with these interventions. Their open-ended responses were analyzed using the same structured, multi-step process employed for the teacher surveys and are summarized by question below.

When principals were asked about the strengths of LLI, one-quarter of responses (25.0%) were related to each of the following areas: the organization of the system (e.g., small, fluid groups), the instructional components, and other aspects of the system (e.g., professional development, benefit for Tier 3 students, and progress monitoring data). Instructional components mentioned included the writing connection, teaching points, focused instruction, and emphasis on the components of reading. Further, a smaller portion of responses (12.5%) were related to the LLI materials, including take-home books and resources.

With regard to areas for improvement, the responding principals' comments most frequently focused on the capacity of the LLI system (36.4%), such as the fact that it only serves a small number of students at a time and that higher grade levels are needed. An equal number of comments (18.2% each) were devoted to the design of the program, other areas for improvement (e.g., the need for in-school professional development and "closure"), and blank or "N/A" responses. Regarding program design, principals suggested providing one clear learning objective for each lesson and strategies for regular classroom teachers to support LLI instruction. Finally, one principal mentioned the expense of the LLI materials as an area needing improvement.

Principals were also asked about challenges they face in implementing LLI and additional resources they may need for implementation. The most common responses were again related to capacity (33.3%), including insufficient staff, the small number of students served, and the need for higher grade levels (40.0%, 40.0%, and 20.0% of these comments, respectively). One-fifth of responses each were related to time (20.0%) and cost (20.0%), including inadequate time to serve all the students who need LLI, scheduling conflicts, and funding needs for materials and teachers. Finally, an additional

20.0% of responses were blank or "N/A," while one principal mentioned that LLI should qualify for ESL pull-out time.

When asked about other literacy interventions at their schools and their efficacy in comparison with LLI, nearly half of the responses (45.5%) listed the programs or strategies used, which included Wilsons, "double dipping," small groups with guided reading, Reading Recovery, and no other interventions (each 20.0% of these comments). In all of the responses related to these programs' effectiveness in comparison with LLI (27.3% of total comments), principals stated that LLI is more effective, with one comment crediting LLI's consistency and another its writing component.

Finally, principals were asked why their school should continue or not continue using LLI. Over half of the responses (55.6%) were related to LLI's effectiveness, with most of these comments (80.0%) referencing student growth. Other reasons mentioned included LLI's promotion of teacher collaboration and the fact that it is a clear, concise, structured system. Table 33 provides a summary of the themes that arose in response to each open-ended question and the percentage of responses for each. Sample comments from principals are provided below; all comments are provided in Appendix M.

"[Our school should continue LLI] because it [has] shown growth in Literacy based on data from several sources."

"LLI provided a clear, concise, and structured program for our teachers and students. We are beginning to see positive outcomes for our students."

"We have seen academic gains as a result of the program. Teachers have shared students and it has promoted communication and collaboration."

Table 33: Thematic Summary of LLIPQ Comments

Theme	Percent of Comments (Per Question)
What are the strengths of LLI?	
Organization	25.0
Instruction	25.0
Other/miscellaneous	25.0
Materials	12.5
Blank or N/A	12.5
What areas of LLI could be improved?	
Capacity	36.4
Design	18.2
Other/miscellaneous	18.2
Blank or N/A	18.2
Cost	9.1
What challenges has your school experienced with implementing LLI? What additional resources are needed to successfully implement LLI at your school?	
Capacity	33.3
Time	20.0
Cost	20.0
Blank or N/A	20.0
Other/miscellaneous	6.7

Table 33, continued

Theme	Percent of Comments (Per Question)
What other literacy interventions, besides LLI, are in place at your school? How effective is LLI in comparison with these interventions?	
Intervention programs and strategies	45.5
Effectiveness of LLI	27.3
Blank or N/A	27.3
Why should your school continue or not continue LLI?	
Effective/successful	55.6
Other/miscellaneous	22.2
Blank or N/A	22.2

Home Literacy Support Questionnaire (HLSQ)

The Home Literacy Support Questionnaire (HLSQ) was administered to parents/guardians of treatment and control group students at the end of the school year as a general measure of their support for literacy at home and perceptions of their child's literacy instruction at school – including LLI, if applicable (n = 137 respondents). Table 34 illustrates the frequencies of responses for each item on the HLSQ. Most of the participating parents/guardians reported exceedingly positive perceptions of their child's literacy activities at home and school and the amount of home literacy support they provide. Further, of those parents/guardians who indicated that their child has participated in LLI, almost all of them shared positive perceptions of the experience.

Overall, participating parents/guardians were most likely to "Agree" or "Strongly Agree" that they have books at home for their child to read (95.6%), that they believe their child can become a good reader and writer (95.6%), and that they encourage their child to practice reading (94.9%) and writing (93.5%) at home. Additionally, two of the largest areas of agreement focused on LLI – almost all of the parents of LLI students "Agreed" or "Strongly Agreed" that their child's school should continue using LLI (98.1%) and that their child's participation in LLI improved his/her reading and writing (96.2%). Less than half of the participating parents/guardians "Agreed" or "Strongly Agreed" that they participate in reading and writing activities at their child's school (48.9%). All items can be found in Table 34 below.

Table 34: HLSQ Response Frequencies (n = 137)

Item	Percent Responded		
	Strongly Agree/Agree	Neutral	Disagree/ Strongly Disagree
My child enjoys reading and writing.	90.6	5.1	4.4
My child reads and writes at home.	89.1	7.3	3.6
I read and write with my child at home.	87.6	8.8	3.0
I have books at home for my child to read.	95.6	2.9	1.4
I read books to my child at home.	90.5	5.8	3.7
I encourage my child to practice reading at home.	94.9	2.9	2.2
I encourage my child to practice writing at home.	93.5	4.4	2.2
I believe my child can become a good reader and writer.	95.6	2.9	1.4
I am pleased with the instruction my child is receiving in reading and writing at school.	90.5	5.1	3.7
I participate in reading and writing activities at my child's school.	48.9	27.0	19.0
I know how my child is doing in reading and writing at school.	87.6	5.1	5.1

Table 34, continued

Item	Percent Responded
Has your child participated in the Leveled Literacy Intervention (LLI) program at his/her school?	
No	12.4
Yes	41.6
Not sure	40.1

Item	Percent Responded		
	Strongly Agree/Agree	Neutral	Disagree/ Strongly Disagree
I think my child's participation in LLI has improved his/her reading and writing.	96.2	3.8	0.0
I participate in LLI take-home activities with my child.	86.8	5.7	5.7
I think my child's school should continue using the LLI program.	98.1	0.0	1.9

Note: Item percentages may not total 100% due to missing input from some participants.

Parents/guardians who completed the HLSQ also responded to open-ended questions regarding their perceptions of the strengths and areas for improvement of their child's literacy instruction at school. Their open-ended responses were analyzed using the same structured, multi-step process employed for the teacher and principal surveys and are summarized by question below.

When asked their opinion of the best things about the reading and writing instruction their child receives at school, participating parents/guardians most commonly discussed the positive impact on students (34.7% of comments). Of these responses, nearly half (45.5%) were related to the fact that their child has learned and/or noticeably improved, one-quarter (25.0%) stated that their child's enthusiasm for reading and writing has increased, and slightly less than one-fifth (18.2%) asserted that the literacy instruction builds their child's confidence. In addition, nearly a third of the overall responses (30.7%) were related to such specific instructional components as the books/materials (20.5%), reading and writing strategies (e.g., sounding out words, sight words, spelling pneumonics; 18.0%), LLI (18.0%), one-on-one instruction (15.4%), and writing activities (12.8%). Further, in 21.3% of comments, parents/guardians discussed positive characteristics of the program in general, such as the clear instructions and expectations (25.9%), support provided to struggling students (18.5%), encouragement to read and write (11.1%), instruction that is targeted to individual reading levels and needs (11.1%), and practice at home (11.1%). Finally, in 9.5% of comments, respondents generally discussed liking the literacy program and feeling pleased with the instruction their child is receiving, while respondents in only 2.4% of comments stated that they were not sure or did not share any positive perceptions.

Participating parents/guardians were also asked what changes they would like to see in their child's reading and writing instruction. Their responses most frequently stated that no changes are needed or that they are pleased with their child's progress (39.1%). In one-quarter of responses (25.0%), participants recommended instructional changes, which most frequently involved a greater emphasis on writing or handwriting (39.1%); other suggestions included more work on reading (26.1%), more challenging instruction for students (13.0%), and more homework or homework options (8.7%). Further, in slightly less than a quarter of responses (23.9%), participants discussed additional support or resources needed, including individualized attention for students (27.3%), communication and progress reports for parents/guardians (27.3%), school-level resources (e.g., after-school help, a school book program, a summer reading program, more textbooks, and a website or list of books with DRA2 levels; 22.7%), and take-home activities (13.6%). Finally, parents/guardians discussed the need for increased student achievement (e.g., helping their child read on grade level) in 7.6% of comments, while

participants in 4.4% of comments stated that they were not sure or did not provide any specific suggestions.

Table 35 provides a summary of the themes that arose in response to each open-ended question and the percentage of responses for each. Sample comments from parents/guardians are provided below; all comments are provided in Appendix N.

Strengths:

"I think the small groups and one-on-one at school have helped [name removed]. The variety of writing assignments have been good for her."

"My son loves the books that he brings home from school (LLI program). Please continue to offer this program to the children. It gave his confidence a boost!"

Changes Needed:

"His books in reading are getting harder because he is going into 2nd grade. I would like to see more support from the staff. Help him with the harder words to read."

"I know teachers are overworked, but a packet of a few books at their level for each child to bring home for a week or two. Monthly update note – your [child's] "DRA" level is: X. Websites or lists of books with DRA equivalent for titles. I would also like to see [school name removed] participate in a book program – one based on comprehension as well as number of words read."

Table 35: Thematic Summary of HLSQ Comments

Theme	Percent of Comments (Per Question)
What are the best things about the reading and writing instruction your child is receiving at school?	
Positive impact on students	34.7
Specific instructional components	30.7
Positive program characteristics	21.3
Program is great/likes it	9.5
Not sure/negative perceptions	2.4
Unusable comments	1.6
What changes would you like to see in your child's reading and writing instruction?	
No changes/pleased with child's progress	39.1
Instructional changes	25.0
Support and resources	23.9
Increase achievement	7.6
Not sure/no specific suggestions	4.4

Stakeholder Feedback

District Literacy Specialist Focus Group

In order to gain a district-level perspective of DPS LLI implementation as well as feedback regarding the use of the LLI system in an urban district, a focus group was conducted with four district-level administrators and specialists who work with LLI and the broader district literacy program. Although a structured interview protocol was utilized (see Appendix G), not all questions were asked due to time constraints; however, the participants' responses to individual questions frequently addressed multiple areas. Responses are summarized by question below.

When the focus group participants were asked what led DPS to adopt LLI as an intervention for struggling readers, they responded that LLI is the only comprehensive reading intervention available (e.g., not merely skills-based or item-based). According to participants, the comprehensive nature of LLI was a critical factor in its adoption because, when the response to intervention (RTI) model began in the district, it originally had a special education focus. This focus resulted in the implementation of intervention programs that only stressed one component of literacy, causing district administrators to search for a more comprehensive program that addressed all five components of literacy. Due to DPS's high ELL population, district administrators were also impressed with LLI's use of "real," high-quality books for ELL students. When asked to elaborate on LLI's fit within the district's overall literacy program, participants stated that it fits well because they use Fountas & Pinnell books throughout the district. Additionally, because the district's guided reading structure is based on the work of Fountas & Pinnell, LLI is aligned with DPS's theory of practice and provides "a more supportive program for teachers to do guided reading in a deeper way." Finally, participants reported that LLI's use is nearly district-wide, with only a few schools lacking the system.

When asked their perceptions of the strengths of LLI, participants were extremely positive and provided a wide variety of responses. In terms of the impact on students, participants discussed improved vocabulary development for children classified as special education or deaf/hard of hearing, as well as benefits for second language learners due to LLI's emphasis on structure (a significant issue for many ELL students). Additionally, participants reported that students love LLI (e.g., reading the new books, receiving the take-home books, having their own writing books) and therefore described it as a "self-motivating program." Participants also described several program components as strengths, including the following:

- Routine – the predictability of LLI's structured routine allows students to develop a sense of competence, which reduces the motivation and behavior problems these students may exhibit in other settings (e.g., the classroom);
- Word work – unlike other programs tried by DPS, the word work in LLI transfers well to continuous text, which means it does not just become "item work" for the students;
- Instructional and independent levels – alternating easy and challenging texts in a structured, supportive format allows students to accomplish what they could not do on their own in the classroom and gives them the opportunity to practice every other day;
- Reciprocity between reading and writing – incorporating both components allows students to see the connections and apply what they are learning; and

- The books – participants described the books as "phenomenal" and even stated that regular classrooms throughout the district are purchasing the take-home books because they need high-quality but inexpensive texts.

Finally, focus group participants discussed the positive effects of LLI on teacher development, stating that the system provides "scaffolding for teacher learning." For example, a teacher does not necessarily have to stop working with students at a certain text level after running out of books in the LLI kit; instead, the teacher can continue the lessons by applying LLI strategies to other leveled texts (e.g., Fountas & Pinnell benchmark books). Additionally, participants reported that LLI's teacher prompts for the three modalities of successful reading are difficult for teachers to learn at first, but they easily transfer to all teaching situations (not just LLI) once teachers learn them. Therefore, participants stated that LLI "makes teachers smarter."

The focus group participants' feedback regarding LLI's impact on student achievement was also generally positive. Participants reported that LLI works better for most of their struggling readers than other programs and stated that students who make progress in LLI generally score on grade level on the DRA2, depending on their starting level. According to participants, students who are identified as having a specific learning disability do not necessarily achieve grade level on the DRA2, but they do begin catching up with other students. Similarly, when LLI is used with third or fourth grade students who are reading at a first or second grade level, they make progress even though they do not achieve scores commensurate with those of their grade level peers. However, participants pointed out that many of these third or fourth grade students remained at a DRA2 Level 6 (corresponding to a first grade reading level) for three years before receiving LLI and finally showing improvement.

The focus group also discussed their perceptions of the aspects of LLI that need improvement. A unanimous response among group members was, although the LLI professional development provided by Heinemann is critical for getting teachers to use the materials well and thus produce the desired effects, it is too expensive. Participants suggested a price break on professional development for districts who purchase a large number of kits and discussed ways in which DPS has made LLI training more sustainable (e.g., training district specialists to provide LLI professional development, providing follow-up support and coaching labs for LLI teachers). Further regarding professional development, participants stated that more examples are needed (e.g., video clips of teachers performing a variety of specific instructional routines or strategies) and suggested providing webinars or a website with a bank of video clips. The word bags were also described as an area for improvement by most of the participants, who stated that teachers either do not use them or quickly become overwhelmed because the small pieces of paper are difficult for young children to manipulate. Participants discussed methods the teachers have developed to make the word bags more user-friendly, such as placing the cards on sturdier paper, on metal rings, or on the wall. Finally, participants pointed out that LLI emphasizes materials that have to be photocopied, which can be a problem in schools where teachers have to stand in line for the copy machine and have limited budgets for photocopies.

Participants also discussed the necessity of having more than one effective intervention program in place – which, they clarified, is not a weakness of LLI but is simply the nature of intervention in general. They pointed out that, although LLI helps most children, it is not effective for every single child; for example, it is too fast-paced for children who may need to spend more time on certain concepts, such as children classified as special education or ELL. As a result, it becomes necessary to either adapt the program or use something else. Participants discussed the "dynamic tension between

fidelity and responsiveness," stating that fidelity to the program (as emphasized in LLI) is not always desirable when working with a diverse group of students who have different needs.

When asked about the challenges that DPS faces when implementing LLI, group members identified three areas. First, they reiterated the cost of the LLI professional development. Second, participants reported that LLI teachers are often pulled for other activities (e.g., end-of-year DRA2 testing), thus reducing the amount of instructional time they can spend with their LLI groups. Finally, group members stated that one 18-week session of LLI nearly takes up the entire school year after accounting for holidays, assessment days, and teachers being pulled from their LLI groups for other activities; thus, it may not be realistic for an LLI teacher to complete multiple sets of LLI groups within the year. However, in terms of the amount of district support for LLI, participants were quite positive; they reported that most members of the central office administration are supportive of LLI, in addition to the majority of principals and teachers throughout the district who want LLI in their schools. Group members shared that DPS has even trained kindergarten classroom teachers to use LLI as an intervention in the classroom because it has changed the way they do guided reading groups with their students.

Finally, the focus group participants were asked about other literacy interventions in place in DPS and LLI's efficacy in comparison with these interventions. Participants explained that, when the RTI model began in DPS, the district tried multiple interventions without a system in place to evaluate which ones were effective. Additionally, interventions are not used consistently throughout the district, and the schools are allowed to select the programs they use; as a result, the central office is not always aware which interventions a school might be using. Overall, however, participants stated that the district is currently focused on supporting classroom teachers, who are generally not as well-trained in guided reading as the interventionists. For example, DPS found that the effects of Reading Recovery tended to diminish by the third grade because students were not able to be "self-extending learners" in the classroom. Group members also reported that the district has used Guided Reading Plus and other interventions, but they would be more effective if the teachers had a better understanding of how to use them. The advantage of LLI in comparison with other pedagogically sound interventions, according to the group, is that the LLI materials are already in place. As a result, it is much easier for LLI teachers to plan their instruction, and they do not have to spend extra time figuring out which materials to use (e.g., if the books are at the right level, have the right sight words, etc.).

LLI Teacher Focus Group

In order to obtain feedback regarding implementation of the LLI system in DPS from current instructors, a focus group was conducted with those LLI teachers who took part in the study. A semi-structured focus group protocol (see Appendix H) was utilized with approximately 18 participating LLI teachers. Responses are summarized by question below.

Responses from LLI teachers were overwhelmingly positive with regard to their overall perceptions of LLI. Teachers especially appreciated the writing component; they credited this component for the LLI system's comprehensiveness and discussed how students not only enjoy this aspect, but also request activities involving writing. Teachers also reported feeling that the books are a primary factor in LLI's success – that students like and are engaged by them. One teacher pointed out that the sequencing of instructional and independent books is particularly helpful in teaching. Another teacher stated that struggling readers need safety and consistency, both of which are provided within the LLI system. While the general consensus was that LLI meets students' needs, one teacher pointed

out that at times, a group of three students can be challenging because "someone [is] always left behind." However, another teacher refuted this point, stating that it is easier to differentiate teaching strategies for only three students when compared to guided reading.

When focus group participants were asked to comment on common logistical issues they encountered when implementing the LLI system, many comments centered on materials, specifically items related to lessons that the LLI teacher must cut out and the high volume of copies LLI teachers are required to make (often with limited access to a photocopier and limits on the number of copies a teacher can make). LLI teachers went on to state that they felt some copies could be included in the kits and that materials they did need to copy should be better laid out as to require fewer overall copies. Participants identified sight words as being especially problematic in that they must be cut out, and one participant was unsure how to follow up on the sight words once they are given out to students. Commenters further pointed out that knowing the pieces needed by each child versus the group is sometimes difficult, as is moving the materials around. Another area eliciting numerous remarks was the home connection portion of the system. Several participants pointed out that parents sometimes do not understand the importance and described their discovery that take-home books had at times been thrown away by parents. Similarly, the classroom connection segment of the system also prompted some comments from focus group participants who stated that it is difficult for students to go back to the classroom and complete an assignment for LLI when they are already attempting to catch up from being away from the classroom. The LLI teachers discussed how important yet challenging time management can be, especially for first-year LLI teachers and as students advance through the system. Preparing for ESL students requires additional planning time, according to participants.

When LLI teachers were asked to discuss the strengths of the LLI system, responses were varied and enthusiastic. One of the most frequently mentioned components of LLI that teachers like is the books, both those used during the lessons and the take-home books. One participant pointed out that "A ton of our kids have no books at home – they remember the books." Teachers also appreciated the ability to move between the Orange, Green, and Blue Systems and that there are so many books at each level to teach different strategies. They also valued that the books followed the word work in sequence and that the books targeted students' reading level accurately. In addition, the LLI teachers liked being able to use independent and instructional leveled books together to best help each child. Also mentioned were the transcripts for Reading Records, the ability to assess as frequently as needed, and the ability to show progress easily with the data reports. Re-reading was another aspect of the LLI system that teachers especially liked. The pace, the flexibility, and the extent of activities included in the system were also features pointed out by teachers as strengths; one teacher stated "I like there's so much to do – I never run out." The games and writing were specified as activities particularly enjoyed by LLI teachers and students. Finally, the small number of students working together was identified by a teacher as being especially effective in allowing the teacher to build relationships, which helps in managing the group and switching between students as necessary.

When asked about areas of LLI that might need improvement, LLI teachers indicated several areas that were related to the home connection portion of the system. Participants felt that the letter to parents does not include all important aspects of the LLI system and pointed out that the materials are not in all languages spoken by parents. LLI teachers commented that they are unsure parents read with their children daily; they postulated it might be once a week or once per level. Providing manipulatives to send home as an activity was suggested by an LLI teacher in response to possible improvements. Participants also gave fold sheets and Getting Started lessons as specific examples of aspects that need modification. One LLI teacher gave the opinion that these lessons contain too much

and suggested that they be separated into even/odd lessons. Lastly, the LLI Online Data Management System was discussed; one participant stated that it had taken three weeks to have a message returned from technical support staff.

LLI teachers were asked how effective LLI has been in meeting individual students' needs. Participants generally stated that LLI is highly effective, given some caveats. LLI teachers pointed out that instructional differentiation is important but that it is the teacher's responsibility. This differentiation contributes to overall success, as does the ability to use the data and Reading Records to drive instruction. One participant stated that LLI makes problems with reading easier to see. LLI teachers felt that generally LLI works best with the lowest achievers, pointing out that sometimes the progress of students at higher achieving levels is slowed by lower-achieving students in the group. Although all students are at different levels, LLI teachers pointed out that sometimes it is best to move students to different groups. One LLI teacher pointed out that the group instruction feature of the LLI system is what makes the instruction work. While there was some debate, LLI was seen as effective in working with English language learners; it was also seen as challenging. One teacher stated that instruction must be "bolstered" with these students. Another hypothesized that second-language speakers show more improvement, adding that it may be that English-speaking students who struggle with reading may have more significant problems that are undiagnosed. It was pointed out LLI is not as effective with students who have phonetic processing disorders. Finally, one teacher pointed out that efficacy increases with experience; "this second year teaching was highly effective – kids sky rocketed."

LLI teachers unanimously reported that they feel their administration is supportive of the LLI system. When asked about other school faculty, participants reported that most teachers in their schools are supportive, but there are still struggles with logistics and scheduling at times.

When asked their opinions about the training they received, many of the LLI teachers reported that the training was highly effective and supportive, in particular the training provided by DPS. One group member cited the "Behind the Glass" transitional strategies presented in one training session, as it "assure[s] we get it all in 30 minutes." One LLI teacher stressed the importance of organizational strategies being included in training. Several commenters, however, pointed out that additional training for use of the online data management system would be beneficial. One LLI teacher felt the training was sometimes time-consuming and redundant.

When given the opportunity to make additional comments, one LLI teacher questioned if the students who were being tested by on-site researchers experienced increased test anxiety due to not knowing the researchers. Another group member stated that they were unsure whether the on-site researchers understood LLI and were able to accurately observe LLI lessons as a result. Lastly, one LLI teacher expressed an interest in "going deeper" into LLI to understand why certain components are important.

On-Site Researcher Focus Group

Because the on-site researchers who collected observational and student benchmark data for the study were primarily retired teachers who had experience teaching in DPS, CREP researchers utilized focus groups to solicit their feedback regarding LLI and its implementation in DPS. The on-site researchers were able to provide an objective "outsider's" perspective based on their random observations of the LLI groups. Approximately 14 on-site researchers voluntarily participated in the focus group; responses are summarized by question below.

When asked their overall perceptions of LLI, the on-site researchers' responses were very positive. Participants were particularly impressed with the readily available lesson materials, lesson flexibility, pattern of alternating instructional and independent texts, pacing (including some teachers' use of a timer to keep the lesson moving), group size, continuity of instruction, adaptability to multiple reading styles, and structure (which is helpful for new or inexperienced teachers). On-site researchers were also asked about the students' response to LLI during the observations, and nearly all participants reported that the students liked LLI and responded well. Specifically, participants stated that students knew what was expected of them and were enthusiastic, eager for instruction, and on-task. However, participants did point out that students were more engaged if the teacher was skilled and enthusiastic; thus, classroom management was an important component.

On-site researchers also shared their perceptions of the strengths of LLI. Several participants mentioned the books, which they described as "interesting for most children most of the time." Participants also liked the fact that both fiction and nonfiction books are included and that a scripted introduction is provided for new texts. Another strength frequently mentioned by group members was the students' response to LLI; participants conveyed that students were engaged and making progress. According to group members, students were also comfortable because they understood the routine and felt free to take risks (e.g., mistakes in their writing book could easily be covered with white correction tape to allow them to try again). Other strengths mentioned included the Reading Record assessments of both fiction and nonfiction books, the group size, and the fact that a skilled teacher can pace the lesson to cover a large amount of material in a relatively short timeframe. Finally, one participant pointed out that the scripted nature of the lessons provides support for less experienced teachers, which makes LLI successful for both skilled and less skilled teachers.

When asked what areas of LLI may need improvement, on-site researchers discussed the fast pace, which makes it difficult to devote extra time to a particular component (particularly phonics). Participants also identified the word work, interactive writing, and classroom connection; according to participants, the word work does not always match the story, and the interactive writing would be easier if all three students could write the same thing at the same time. However, several areas for improvement were specifically related to DPS's implementation of LLI. For example, at some schools, teachers did not have a designated classroom and had to do LLI in the hallway or library. Additionally, participants pointed out that some teachers were more skilled than others and that students would be more likely to achieve consistently with a skilled teacher. Finally, one participant asserted that the students in LLI did not seem as needy as some of the students in the classroom, and another stated that the philosophical differences between LLI and the Read to Achieve (RTA) state grant program could be confusing for LLI teachers.

Participants in the on-site researcher focus group were also asked about the quality of both LLI and classroom literacy instruction during their observations. With regard to LLI, participants praised the rigor of the instruction but stated that teachers had to be very organized and well-trained. Regarding classroom literacy instruction, however, the responses were more varied. Some participants reported that they were impressed with the classroom instruction, particularly the use of centers and paraprofessionals, and that they perceived a strong literacy environment in most classrooms. Conversely, an approximately equal number of responses were less positive. According to these participants, there was a disconnection between the classroom instruction and the high quality of the LLI instruction. In some classrooms, the on-site researchers observed teachers exercising ineffective classroom management or "just doing activities" rather than teaching to objectives. One participant stated that she was never able to observe a guided lesson. Further, participants reported that there did

not appear to be a progressive K-6 program, which meant that teachers were using their own materials and seemed to be "coming up with the program themselves."

Finally, the on-site researchers were asked their perceptions of the training they received and the instruments they used to collect data for the study (i.e., the LOT and LLIOT observations and the Fountas & Pinnell Benchmark Assessment System). Overall, most participants were positive about the training, citing the videos, observation handbooks, benchmark instructions and assessment/scoring charts, mid-year refresher training, and responsiveness to questions as helpful. However, participants stated that there were some contradictions when questions arose and that it would have been helpful to conduct a practice observation of an LLI lesson prior to data collection. Regarding the instruments, participants described the LLIOT (the LLI group observation tool) as well-organized and easy to follow but suggested providing more space for descriptive comments. Several participants reported feeling that the LOT (the classroom literacy observation) did not provide a full picture of the reading program because it was based on 10-minute observations conducted in a single day. Participants described it as a useful tool for observing the learning environment and level of classroom management, but less useful for seeing literacy strategies "in action." Finally, regarding the Fountas & Pinnell Benchmarks, participants liked the pattern of alternating fiction and nonfiction texts because they felt that the genre affected student performance (e.g., most students performed better with fiction, but ELL students tended to perform better with nonfiction). However, participants stated that the benchmark training and materials did not prepare them for some of the complicated situations that arose with administration and scoring. Further, participants reported that the testing environment for the benchmarks was not always ideal (e.g., the hallway) and that students frequently became tired before completing the assessment (particularly at the posttest).

Conclusions

1. What progress in literacy achievement, if any, do urban students who receive LLI make compared to students who receive core literacy instruction alone?

The results of the current study revealed that LLI positively impacts urban students' literacy achievement in kindergarten and first grade; however, fewer positive effects were observed for second grade students. Demographic subgroups including males, females, Hispanic students, and ELL students were also shown to benefit from LLI. Fidelity of LLI implementation (i.e., the degree to which LLI was implemented as designed) was shown to have some impact on student achievement in kindergarten and first grade; however, the amount of LLI attendance – relative to the recommended amount – appeared to have minimal effects on student achievement.

Particularly robust effects were observed for kindergarten students, who significantly outperformed their counterparts in the control group by one level on the Fountas & Pinnell Benchmarks (2.3 versus 1.3 benchmark levels gained). While both kindergarten treatment and control group students began the study below Level A (0.8 and 0.7, respectively), treatment group students progressed to Level C (3.0), and control group students ended at Level B (1.9). Thus, while both groups started below the mid-year grade level goal (i.e., Level A), only the treatment group achieved the end-of-year grade level goal (i.e., Level C).[5] Significant effects on the benchmarks were also observed for such kindergarten subgroups as males, females, Hispanic students, and ELL students, who each gained approximately one more benchmark level and finished closer to grade level than their control group counterparts. Finally, kindergarten students in the treatment group made significant gains on the DRA2 compared to the control group, gaining nearly one more level than control group students on average (2.1 versus 1.3 levels gained) and finishing closer to DPS's end-of-year grade level goal of 4 (3.8 versus 3.1). Significant subgroup effects on the DRA2 were also noted for females, and slightly significant effects were noted for ELL students.

Several positive effects were also found for first grade students, who significantly outperformed their control group counterparts by one benchmark level (4.4 versus 3.4 benchmark levels gained). Control group students began the study at Level A, slightly above treatment group students (1.0 and 0.8, respectively); however, treatment group students progressed to Level E (5.2), while control group students finished slightly above Level D (4.4). Thus, while both groups started below the beginning-of-year grade level goal (i.e., Level C), treatment group students more closely approached the mid-year grade level goal of F. Although these students did not achieve the grade level goal, their progress is quite noteworthy given the fact that they began a full year below grade level. Further, significant effects on the benchmarks were observed for such first grade subgroups as females, Hispanic students, and ELL students, who each gained a full benchmark level more and finished closer to grade level than their control group counterparts. Finally, only one significant effect emerged on the DRA2, in which male control group students appeared to outperform their treatment group counterparts. Although unexpected, this finding is somewhat offset by the fact that treatment group males outperformed control group males on the benchmarks – a finding that, while not statistically significant, is considered

[5] All grade level goals for the Fountas & Pinnell Benchmarks were determined using the 2009 Fountas & Pinnell Instructional Level Expectations, which were more appropriate for use with the current sample than the 2012 expectations; see Appendix O for a comparison of the study results with both the 2009 and the 2012 expectations.

"substantively important" according to What Works Clearinghouse guidelines concerning effect size (U.S. Department of Education, 2011).[6]

Although some effects emerged for second grade students on the Fountas & Pinnell Benchmarks, the majority of these findings were only marginally statistically significant with second graders in Denver Public Schools. As a whole, second grade treatment group students outperformed their counterparts in the control group by less than one benchmark level (3.8 versus 3.1 benchmark levels gained). While both treatment and control group students began the study between Levels E and F (5.5 and 5.7, respectively), treatment group students ended just above Level I (9.3), and control group students finished just below Level I (8.8). Thus, both second grade groups concluded the study at the beginning-of-year grade level goal (i.e., Level I/J) rather than the mid-year grade level goal (i.e., Level K), although their progress is still impressive given the fact that they started at the *first grade* mid-year grade level goal.

Given the progress made, though not statistically significant, post-hoc analyses were conducted in order to examine these effects in a larger sample. The second grade sample from DPS and a prior study with schools in Tifton, GA and Middletown, NY were then combined, after statistically being equated, and results revealed a significant gain in favor of the second grade treatment group. The treatment group gained about 4.5 levels, finishing close to Level J, while the control group only gained around 3 levels, finishing close to Level I. These results are similar to those seen when using on the Denver second grade students; however, with the increased combined sample size, statistical significance was also attained.

Slightly significant effects on the benchmarks were also observed for male students, and significant effects were detected for Hispanic students, who gained a full benchmark level more than Hispanic students in the control group. Further, one slightly significant effect was noted for second graders on the DRA2, in which females outperformed their control counterparts by four levels (13.9 versus 9.8 levels gained) and finished closer to DPS's end-of-year grade level goal of 28 (25.1 versus 22.8). Finally, although not statistically significant, ELL students' gains compared to the control group on both the benchmarks and the DRA2 were not only "substantively important" according to What Works Clearinghouse guidelines concerning effect size (U.S. Department of Education, 2011), but they also represented the largest benchmark gains in the study (4.6 versus 3.0 benchmark levels gained, a difference of one and a half benchmark levels).

Across all three grade levels, there were no significant results on the STAR Early Literacy Assessment for the treatment and control groups, most likely due to extremely small sample sizes because very few students had complete sets of pre and post STAR scores. A possible explanation for these low numbers may be that the administration of the STAR was newly required in DPS during the 2011-2012 school year, and although a certain number of administrations was mandated throughout the year, the timing or other factors related to these administrations may have varied widely.

Teachers' fidelity to the LLI model (i.e., the degree to which they implemented LLI as it was designed) was associated with students' performance on the Fountas & Pinnell Benchmarks in kindergarten and first grade because students who received "high fidelity" LLI outperformed students who did not receive LLI at all. This finding was noteworthy in that it was highly statistically significant,

[6] Effect size measures the amount of difference between two groups (e.g., treatment and control) without taking the size of the groups into account. Thus, a "substantively important" effect size suggests that LLI had an impact even if the results were not statistically significant.

despite the fact that there were over twice as many students in the "no LLI" (i.e., control) group than in the "high fidelity" group; therefore, the performance of the "high fidelity" group was extremely strong in order to outweigh the performance of the larger "no LLI" group. Similarly, in kindergarten, students who received "high fidelity" LLI outperformed students who did not receive LLI on the DRA2. However, for both kindergarten and first grade students, there was no difference on either the benchmarks or DRA2 between students who received "low fidelity" LLI and either "high fidelity" or no LLI.

Across all three grade levels, there was no impact on either benchmark or DRA2 performance of the number of LLI instructional days attended relative to the number of LLI instructional days recommended by the LLI system. However, first grade students' STAR performance did significantly improve as their percentage of recommended instructional time increased, with students' performance on all three tests at all three grade levels following the same trend. Only one exception occurred, in which kindergarten students' STAR performance appeared to decline with increased LLI instructional time – a finding that was most likely spurious due to an extremely small sample size. These results suggest that maximizing LLI instructional time is an important factor in students' success, but achieving the recommended number of instructional days may be less critical.

Finally, comparison group students who did not participate in the RCT portion of the study but did receive LLI during the 2011-2012 school year made highly significant gains on the DRA2 at all three grade levels. All demographic subgroups made significant DRA2 gains as well, including males, females, Hispanic students, ELL students, and students classified as receiving special education services. In contrast, only second grade students made significant gains on the STAR, with significant subgroup results also obtained by second grade males. However, slightly significant STAR gains were noted for second grade Hispanic students, and "substantively important" (although not statistically significant) findings were observed for kindergarten females, kindergarten Hispanic students, and second grade students classified as receiving special education services. Because the comparison group analysis did not include a control group who did not receive LLI, it is not possible to infer a causal relationship between comparison group students' LLI participation and their growth in literacy scores. For this reason, and because comparison group students received LLI under uncontrolled conditions and were not randomly assigned to receive LLI, these results should be interpreted with caution.

2. At what level of fidelity to the program model is LLI implemented by teachers participating in the study?

Overall, the observation results from the current study suggest that LLI was implemented with a high degree of fidelity to design. In most of the observations, the majority of lesson components received high fidelity ratings. Further, the on-site researchers generally concluded that the lessons they observed were delivered as designed. On-site researchers also largely described the observed LLI instruction as rigorous and of high quality, including the pacing, organization, adherence to LLI protocols, and effective use of instructional strategies and lesson resources. Additionally, the observation results revealed that LLI implementation was consistent across the school year, with strong fidelity scores received at both time points when the observations were conducted. The only change in implementation over the year was seen in first grade, where a slightly significant improvement in "Literacy Instructional Strategies" occurred from the first observation to the second.

The observation results were corroborated by self-report feedback from the participating LLI teachers, a large majority of whom reported implementing LLI as designed (e.g., meeting daily for 30 minutes, following the LLI Lesson Guide), understanding the LLI goals and procedures, and having

sufficient training to implement LLI effectively. Further, qualitative feedback from the LLI teachers suggests that the lesson component receiving the lowest fidelity rating – the classroom connection – may have been implemented less frequently because teachers perceive that it is too difficult for students to complete an LLI assignment in their regular classroom while they attempt to catch up on the work they missed while out of the classroom for LLI.

Finally, the LLI attendance records from the current study revealed that, on average, students received less than the model's recommended number of instructional days (i.e., approximately 62 days instead of 90 for first and second grade, and approximately 45 days instead of 70 for kindergarten). Although second grade students made few significant gains, kindergarten and first grade students made significant progress in their literacy achievement despite receiving less than the recommended amount of instruction – particularly kindergarten students, who received LLI during the second semester and thus encountered multiple interruptions due to inclement weather, spring break, and end-of-year assessments. This finding suggests that LLI can still be effective during a relatively shorter timeframe, which may be valuable to districts with a large number of students to serve or limited time in which to implement early literacy interventions.

3. What are stakeholders' perceptions of the LLI system and the core literacy program?

Overall, LLI teachers in the current study supported LLI and believed it had a positive impact on their students' literacy achievement and attitudes toward reading and writing. LLI teachers indicated that they had a good understanding of LLI; received support in implementing LLI from their district, school administration, and other school staff; and thought LLI is aligned with curriculum standards while allowing them to provide differentiated instruction to meet students' individual needs. Participating LLI teachers also perceived that LLI benefitted their English Language Learner students, who did make significant gains compared to the control group in kindergarten and first grade – an important finding given the fact that ELL students comprise a third of the DPS student population. Further, LLI teachers reported a positive impact of LLI on their reading instruction, particularly their understanding of the role of comprehension in successful reading, the reading process in general, and the relationship of leveled texts to successful reading. In addition, participating LLI teachers were extremely positive about the design and organization of the LLI system, such as the writing component, comprehensive approach to literacy, lesson structure and sequencing of books, small-group format, and ease of progress monitoring using regular Reading Record assessments and data management system reports. Finally, LLI teachers liked the LLI lesson resources; they appreciated the ready availability of all materials needed for the lesson and described the books as high-quality and engaging to their students who frequently do not have books at home.

However, some participating LLI teachers perceived that LLI was less effective with both higher achieving students and students with special needs and reported that they do not have sufficient teaching staff or time to effectively implement LLI with all students who need it. LLI teachers also shared feedback regarding possible improvements to the LLI system, such as the need for more time during the lessons (particularly on even-numbered and Getting Started lesson days) and less paper use. Additionally, some LLI teachers thought the writing and word work components should be more aligned with the lessons and suggested additional training on the LLI Online Data Management System, with which they experienced several technical difficulties (e.g., missing data). Finally, LLI teachers reported struggling with the home and classroom connection components of the LLI system, stating that parents frequently do not engage in home literacy activities with their children and that the LLI parent letter and take-home materials are not in a variety of primary languages spoken by parents.

With regard to the K-2 classroom teachers in the study, teachers' perceptions of their school's core literacy program were generally positive, although some areas of concern were identified. Teachers indicated that their schools are supportive of students' literacy and noted such positive aspects of the core literacy program as small group instruction, guided reading, differentiated instruction, support for ELL students, and teacher flexibility and autonomy. Additionally, participating K-2 classroom teachers were particularly positive about LLI and its impact on students' literacy achievement, participation in classroom literacy instruction, and enjoyment of reading. However, some classroom teachers suggested that the core literacy program does not meet all students' needs (e.g., struggling students, second language learners, students with special needs). Several classroom teachers also thought the curriculum is inconsistent and needs to be more comprehensive and rigorous. Finally, some teachers described district support and parental involvement as low and discussed the need for improved resources (e.g., materials in both English and Spanish, teacher and student support).

Participating K-2 classroom teachers also provided self-report data regarding the classroom literacy instruction that students in the study received as part of the core literacy program. Teachers frequently reported implementing writing activities, integrating vocabulary and comprehension, utilizing high-quality texts to read to students and engage them in interactive discussions about the text, and using both whole-group instruction and individual or small-group instruction; however, they were less likely to report having students engage in group discussions and assigning home literacy activities for students to complete with parents. Independent observations of classroom literacy instructional practices corroborated teachers' frequent use of both whole-class and small-group instruction, explicit comprehension instruction, and teacher-guided interactive discussion; however, writing activities and vocabulary instruction were infrequently observed, as well as ongoing assessment and cooperative learning. Across all the observed K-2 classrooms, observers generally noted active student engagement and teacher monitoring, effective classroom management, independent student reading, and a classroom environment conducive to literacy learning (e.g., visibly displayed alphabet, classroom library, "word wall" of high-frequency or sight words).

Parents/guardians of students in the study were also asked to share their perceptions of the support they provide for literacy at home and the quality of their child's literacy instruction at school – including LLI, if applicable. Regarding all of these areas, parents'/guardians' feedback was extremely positive. Parents/guardians reported that they have books at home for their child to read, believe their child can become a good reader and writer, and encourage their child to practice reading and writing at home, but less than half agreed that they participate in literacy activities at their child's school. Parents/guardians also perceived that the school's literacy instruction has improved their child's achievement, enthusiasm, and confidence related to reading and writing and generally reported that they are pleased with the literacy program. Although a large percentage of parents/guardians stated that no changes are needed to the literacy program, some parents/guardians suggested such improvements as a greater emphasis on writing, more individualized support for students, and regular communication and progress reports. Specifically regarding LLI, parents/guardians of LLI students were once again extremely positive, almost unanimously agreeing that their child's school should continue using LLI and that LLI has improved their child's reading and writing. Parents/guardians stated that LLI has helped improve their child's confidence and reading skills, with one parent even suggesting that LLI should become part of the regular classroom literacy instruction.

Of the 13 schools participating in the study, approximately half of the principals provided feedback on LLI and its implementation at their schools. Most of the participating principals indicated a

high degree of support for literacy in their schools and shared positive perceptions of LLI, reporting that they support LLI in their schools, that LLI has positively impacted students' literacy achievement, and that LLI has improved their school's ability to provide literacy instruction to meet the needs of low-performing students. Principals also commented positively on LLI's organization, instructional components, and materials, but complained that it only serves a small number of students and that higher grade levels are needed; further, some principals thought LLI only somewhat benefits their ELL students and students with special needs. However, participating principals chose LLI when asked to compare its effectiveness to other literacy interventions in place at their school, and all participating principals agreed that their school should continue using LLI.

Finally, district-level administrators who work with LLI were extremely positive about LLI and its implementation in DPS. They reported that LLI has positively impacted student achievement, including that of second language learners and students classified as special education or deaf/hard of hearing. They also discussed such strengths of LLI as the books, routine, word work, sequencing of instructional and independent levels, and reciprocity between reading and writing; further, they stated that the students love LLI and that it has benefits for teacher development. However, the district literacy specialists thought the LLI professional development is too expensive and that it should include more examples (e.g., video clips). Additionally, they described the word bags and the emphasis on photocopies as areas needing improvement and stated that LLI is less effective with students who may need to spend more time on certain concepts (e.g., children classified as special education or ELL). Finally, the district-level administrators stated that it is difficult for DPS to provide the amount of instructional time recommended by the LLI system due to multiple holidays, assessment days, and teachers being pulled from their LLI groups for other activities.

4. How do the results of the Fountas & Pinnell Benchmark Assessment System compare to those of the Developmental Reading Assessment, 2nd Edition (DRA2) and the STAR Early Literacy Assessment?

Grade level equivalence information was available to compare treatment and control group students' scores on the Fountas & Pinnell Benchmarks with their scores on the DRA2 but not on the STAR Early Literacy Assessment, which also did not have a sufficient sample size in the current study to support such an analysis. Further, definitive conclusions can only be drawn about the comparison between kindergarten benchmark and DRA2 scores due to the high number of unavailable DRA2 scores in first and second grades for the time period used for analysis (i.e., Winter 2011-2012, when DPS does not administer the DRA2 to all first and second grade students).

Overall, there was a low rate of agreement between kindergarten students' posttest instructional levels on the benchmarks and their posttest DRA2 scores, with less than a quarter of scores on the two assessments aligned with one another (i.e., Level C on the benchmarks and Level 3 on the DRA2). However, approximately one-quarter of these kindergarten students did not have a comparable posttest DRA2 level for their posttest benchmark level, according to the chart in Appendix J (i.e., Level B on the benchmarks); therefore, these students were automatically considered to have "no match" because no corresponding DRA2 level was available in the grade level equivalence information provided by Heinemann. Further, kindergarten students' scores on the DRA2 were more frequently categorized as "proficient" than on the benchmarks, with agreement between the benchmarks and DRA2 regarding proficiency or non-proficiency status occurring only half of the time. This suggests that the Fountas & Pinnell Benchmarks and the DRA2 differ in their approaches to assessing growth in literacy, making comparisons between the two assessments difficult.

Nearly two-thirds of the first grade students and half of the second grade students did not have an available DRA2 score; additionally, approximately 15.0% of first grade students and 20.0% of second grade students did not have a comparable posttest DRA2 level for their posttest benchmark level, according to the chart in Appendix J (i.e., Level B, K, or M on the benchmarks). In general, there was a low level of agreement between both first and second grade benchmark and DRA2 scores, but there was a trend suggesting high agreement between the two assessments on proficiency or non-proficiency status. However, due to the extremely small sample size, the first and second grade results are highly inconclusive and must be interpreted with caution.

Limitations of the Study and Recommendations

Although the current study produced significant positive findings regarding the efficacy of LLI in urban settings, several factors were encountered that may limit the generalizability of the findings and that prevented researchers from obtaining adequate power to draw definitive conclusions. These limitations are summarized below, followed by data-based recommendations for improvement.

The main limitation facing the current study involved the sample size, or the number of students that were able to participate in the study. A major factor in this limitation was DPS's need to comply with state Read to Achieve (RTA) grant stipulations, which required certain students to receive literacy services immediately, thus making them ineligible for possible randomization to the control group (i.e., delayed LLI services). Therefore, the pool of students who were eligible to participate was greatly reduced from the number that was originally anticipated in the evaluation plan, dropping from 600-800 proposed students to a total of 320 who completed the RCT. This reduction in sample size was also due to other factors, such as the fact that schools were allowed to volunteer to participate (resulting in 13 participating schools rather than 25 as proposed); further, because parents/guardians were required to provide active consent for their children to participate, the large number of students for whom consent was never returned were considered ineligible for the study. Finally, some participating schools' limited capacity to serve multiple LLI study groups resulted in a small number of students who were able to participate at those schools. The greatest impact of the reduced sample size was seen in second grade, which had approximately half the number of participating students as first grade. The small size of the second grade sample may have made it more difficult to detect significant differences between the literacy gains of treatment and control group students. However, some positive effects were detected with marginal statistical significance and/or "substantively important" effect sizes according to What Works Clearinghouse guidelines (U.S. Department of Education, 2011). Therefore, we conduced post-hoc analyses with a larger sample of second grade students, as a previous study on LLI suggested this sample might have yielded stronger results with slightly larger sample (Ransford-Kaldon et al., 2010).

Another limitation of the study was the fact that the researchers did not have control over the administration of the DRA2 and STAR literacy assessments, which are administered by the district. Therefore, it was necessary to utilize the available scores that were provided by the district, which meant that a large number of participating students did not have complete sets of pre and post scores and had to be eliminated from the analyses. Student data was particularly limited for the STAR, with some analyses comparing as few as three or four students per subgroup. As seen above, such small sample sizes make it very difficult to detect statistically significant differences. In addition, the testing conditions for both assessments may have differed across schools – again, particularly for the STAR, which was newly required in DPS during the 2011-2012 school year. Therefore, although a certain number of STAR administrations was mandated throughout the year, the timing or other factors related

to these administrations may have varied widely. Finally, because the most widely-available DRA2 data had to be used in the student achievement analyses, the first and second grade DRA2 pre and post scores (i.e., Spring 2011 and Spring 2012) do not align with the timeframe in which treatment group students received LLI (i.e., Fall to Winter 2011-2012). As a result, control group students may have also received LLI (or another literacy intervention) by the time the posttest was taken in Spring 2012, thus increasing the likelihood that they could score similarly to the treatment group students.

A third limitation of the study design was the fact that control group students were allowed to receive other supplemental literacy services while they were participating in the study, as long as they did not receive LLI until after the study was over. This was a district-level request that was necessary in order for them to agree to participate in the research study. Treatment and control group students were required to receive the same total amount of literacy instructional time (e.g., two hours of core literacy instruction plus half an hour of LLI for treatment group students and half an hour of another literacy intervention for control group students). As a result, LLI was not compared to a typical control group in which students would not have received any additional literacy support outside of the regular classroom. Additionally, the supplemental literacy services received by control group students could vary from school to school. The fact that both groups could receive additional literacy services may have resulted in a smaller difference in literacy gains between treatment and control group students. However, this limitation serves to increase the meaningfulness of the significant gains made by treatment group students in comparison to the control group, because receiving LLI helped these treatment group students outperform control group students who not only received core literacy instruction, but also supplemental literacy services.

A final limitation of the study was the fact that treatment group students did not receive the recommended amount of LLI instructional time as a result of individual absences, delays in starting LLI due to the time required to pretest and randomize students, and district-level factors (e.g., holidays, assessment days, and LLI teachers being pulled from their LLI groups for other activities). Although first and second graders received the recommended 18 weeks of instruction, they were not full weeks and equated to 62 out of the recommended 90 instructional days. Kindergarten students received 12 out of the recommended 14 weeks, or 45 out of the recommended 70 days. Although it is not clear whether receiving the recommended number of instructional days would have produced more significant results, it is possible that maximizing LLI instructional time would have resulted in greater student gains. However, similar to the above limitation regarding control group students receiving supplemental literacy services, this limitation serves to increase the meaningfulness of treatment group students' literacy gains during the shortened timeframe in which they received LLI.

Recommendations

The limitations above reflect the inherent complications in performing educational research in a large urban district, particularly with a research design as stringent as an RCT. However, despite these limitations, the current study found significant positive effects of LLI on urban students' literacy achievement when implemented with fidelity to the LLI model. Further, stakeholders in Denver Public Schools – including teachers, administrators, and parents – were supportive of LLI and perceived positive benefits of the LLI system for their students. Altogether, the results from this evaluation allow us to conclude that LLI positively impacts urban students' literacy skills, particularly in kindergarten and first grade. These results also suggest that continued implementation of LLI would be beneficial in DPS and offer an opportunity for research-based recommendations that may enhance the system, future LLI research, and ultimately student achievement. From this evaluation, CREP proposes the following recommendations with regard to LLI and its implementation in schools:

Design

1. Data from the current evaluation indicated that the Classroom Connection component was under-utilized. The developers might consider providing suggestions for involving classroom teachers in LLI, or even eliminating this component based on real-world feedback.
2. Feedback from some LLI teachers described the activities included in even-numbered lessons and "Getting Started" lessons as including too much information for 30 minutes. The developers might consider splitting "Getting Started" lessons into even- and odd-numbered lessons, as well as suggestions for how to best manage time to include all lesson components into the allotted timeframe.
3. Specific suggestions from LLI teachers and district-level specialists included greater alignment of the word work and writing components with the lessons and reorganization of the word bags so they are more manageable. Because the small word cards are difficult to manage and for children to manipulate, it may be helpful to use sturdier paper or to place the cards on rings instead of in bags.
4. If at all possible, the kits should include a set of photocopies for teachers, and any materials to be copied by teachers should be laid out in a manner that maximizes space in order to decrease the total number of copies made. Teachers frequently have a limit on the number of photocopies they are allowed to make, and they may have limited time or access to a photocopier. In addition, the amount of materials that must be cut out can be overwhelming for teachers; if possible, it is recommended that the developers reduce the number of cutouts or include pre-cut materials in the kits.
5. It is recommended that the developers include multilingual parent materials, such as progress letters and instructions for take-home materials. Additionally, the developers may wish to provide suggestions to help teachers communicate the importance of literacy to parents/ guardians, in order to increase their participation and the use of LLI take-home materials.
6. It is recommended that the developers establish national norms for the Fountas & Pinnell Benchmark Assessment System to facilitate comparisons between the system and other nationally normed literacy assessments.
7. Because analyses from the current study suggest limited comparability between the Fountas & Pinnell Benchmark Assessment System and the DRA2, it would be helpful to examine the content of the two assessments to determine the differences in their approaches to assessing growth in literacy. Further, to facilitate possible comparisons between the benchmarks and the DRA2, it is necessary to establish a comparable DRA2 level for each Fountas & Pinnell

Benchmark level. Finally, in order to allow comparisons between the Fountas & Pinnell Benchmarks and the STAR Early Literacy Assessment, comparable scores must be established between these assessments.

8. It is suggested that the developers continue to refine the LLI Online Data Management System. LLI teachers in the current study experienced some technical difficulties, such as data that was missing for certain users and the inability to set up an online data management account. They also complained of a long turnaround time in receiving assistance from technical support staff and requested the ability to transfer students between groups in the system, as students may move from one teacher/group to another based on their progress in LLI.

Implementation

1. In both the current study and a previous evaluation of LLI (Ransford-Kaldon et al., 2010), schools experienced difficulty achieving 18 weeks, or 90 days, of LLI instruction. Even though it is possible to implement the program across 18 calendar weeks, they are not full weeks of instruction due to holidays, assessments, etc., as well as individual student and teacher absences or unavailability. This is even more difficult during the second semester, when kindergarten students are typically served, because schools have multiple end-of-year assessments and activities as well as such events as spring break. As a result, in the current study, it was not possible to achieve the recommended 14 calendar weeks for kindergarten students. Finally, due to staffing limitations, schools may have difficulty implementing multiple concurrent sessions of LLI because it is necessary for teachers to complete their current LLI groups in order to have the capacity to serve additional groups. Therefore, because both the current study and Ransford-Kaldon et al. (2010) have shown that students can make significant gains in as little as half the recommended amount of LLI instructional time, the developers should consider modifying the LLI curriculum to address schools' time constraints. Alternatively, the developers might consider providing suggestions or examples of schedules that allow for full implementation of the 30-minutes-a-day, five-days-a-week instructional pattern.
2. Although the present study showed that, on average, kindergarteners who received LLI were able to attain grade level in 45 days, our recommendation is for districts to determine eligibility in their kindergarten population and provide LLI as early as possible to those who qualify for it.
3. Schools' limited capacity to serve all eligible students can result in less LLI instructional time for kindergarteners, who start LLI later in the school year than first and second graders. Therefore, the developers should consider providing recommendations to help districts either implement LLI simultaneously across all three grade levels or prioritize instruction in certain grade levels. Suggestions might include focusing LLI in kindergarten and first grade to reduce the need for it in second grade or utilizing alternative methods of implementation (e.g., offer LLI as an after-school intervention for kindergarteners). Additionally, although the ideal group size appears to be three students, districts may be able to maximize the number of students served by including four students per group. The developers should consider offering suggestions for such districts to help them effectively manage four-student groups if necessary.
4. Student achievement data from the current study showed that fidelity to the LLI model may be a key component in obtaining the most desirable outcomes. LLI teachers should be encouraged to follow the program as closely as is feasible, although teacher judgment may be needed to determine how best to individualize instruction to meet students' needs.
5. Districts should be discouraged from pulling students for LLI (or other supplemental interventions) during the literacy block.

6. Districts should prioritize LLI teachers for LLI instruction rather than pulling them for other activities to maximize instructional time for vulnerable students.
7. Because observation data from the present study noted that students were more engaged and typically performed better with skilled and enthusiastic LLI teachers, the developers should consider including recommendations for effective LLI group management.
8. Feedback from LLI teachers and district-level specialists suggested that the LLI lessons may be too fast-paced for slower learners, resulting in a lack of time to spend on specific components students need. The developers should consider providing recommendations on how best to individualize instruction to meet the needs of both higher-achieving and lower-achieving students in a group.

Professional Development (PD) Considerations

1. Feedback from administrators indicated that the expense of the LLI PD and kits may be somewhat prohibitive. The developers might consider a discounted price on PD for districts that purchase multiple kits. Additionally, the developers could provide suggestions to help districts make the PD more sustainable, such as follow-up coaching provided by the district or peer mentoring with experienced LLI teachers.
2. Feedback from LLI teachers and district-level specialists described examples of high-quality LLI instruction as particularly useful during the LLI PD. The developers might consider including more video clips of teachers performing specific LLI instructional routines or strategies during the PD sessions. Additionally, a bank of video clips or webinars on the LLI website may be helpful in providing ongoing support to LLI teachers.
3. Additional training on the LLI Online Data Management System is suggested to help teachers and administrators more easily navigate the system and address technical issues as they arise.

Future Directions

1. In order to obtain larger sample sizes and increase the likelihood of detecting statistically significant differences, further analyses combining data from multiple studies (e.g., Ransford-Kaldon et al., 2010) could be conducted. This may increase the likelihood that generalizable conclusions can be made.
2. Feedback from the current study suggested a great need for LLI in higher grade levels. It is recommended that the developers empirically evaluate the LLI Red Kit in third through fifth grades in a setting where the issues associated with the current study are minimized.
3. It is suggested that additional studies with larger samples of students also be conducted so conclusions can be more universally applied. Also, additional studies empirically evaluating LLI instruction with groups of three compared to groups of four students could provide a research-based conclusion as to whether LLI can be adapted to address the limited staff and capacity of some districts preferring to implement LLI with groups of more than three students.
4. It is recommended that the authors empirically evaluate the revised 2012 Fountas & Pinnell Instructional Level Expectations for use with districts adopting Common Core Standards.

References

Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education*, *70*(2), 87-107.

Armbruster, B. B., Lehr, F., & Osborn, J. (2001). *Put reading first: The research building blocks for teaching children to read: Kindergarten through grade 1*. Jessup, MD: National Institute for Literacy.

Beaver, J., & Carter, M. (2009). *Developmental Reading Assessment, Second Edition: K-8 technical manual.* Upper Saddle River, NJ: Pearson.

Clay, M. (1991). *Becoming literate: The construction of inner control*. Portsmouth, NH: Heinemann.

Fountas, I. C., & Pinnell, G. S. (1996). *Guided reading: Good first teaching for all children*. Portsmouth, NH: Heinemann.

Fountas, I. C., & Pinnell, G. S. (2006). *Teaching for comprehending and fluency: Thinking, talking, and writing about reading, K-8.* Portsmouth, NH: Heinemann.

Fountas, I. C., & Pinnell, G. S. (2007). *The continuum of literacy learning, grades K-2: A guide to teaching.* Portsmouth, NH: Heinemann.

Fountas, I. C., & Pinnell, G. S. (2008). *Leveled Literacy Intervention*. Portsmouth, NH: Heinemann.

Fountas, I. C., & Pinnell, G. S. (2010). *Benchmark Assessment System 1, grades K-2, levels A-N* (2nd ed.). Portsmouth, NH: Heinemann.

Harrison, L., Peterman, R., Grehan, A., Ross, S., Dexter, E., & Inan, F. (2008, March). *Evaluation of the Leveled Literacy Intervention: Year 1.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Heinemann. (2007). *Field study of reliability and validity of the Fountas & Pinnell Benchmark Assessment Systems 1 and 2.* Retrieved September 21, 2012, from http://www.heinemann.com/fountasandpinnell/research/BASFieldStudyFullReport.pdf

Heinemann. (2008). *Research base for Leveled Literacy Intervention*. Retrieved August 29, 2008, from http://www.heinemann.com/fountasandpinnell/research/LLIResearchBase.pdf

Hiebert, E. H., & Taylor, B. M. (Eds.). (1994). *Getting reading right from the start.* Boston: Allyn & Bacon.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*, 437-447.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis: Applied social research methods series #49*. Thousand Oaks, CA: Sage.

National Institute of Child Health and Human Development. (2000a). *Report of the 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: National Institutes of Health.

National Institute of Child Health and Human Development. (2000b). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups.* Washington, DC: National Institutes of Health.

Peterman, R., Grehan, A., Ross, S., Gallagher, B., & Dexter, E. (2009, April). *An evaluation of the Leveled Literacy Intervention program: A small-group intervention for students in K-2.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Ransford-Kaldon, C. R., Flynt, E. S., Ross, C. L., Franceschini, L. A., Zoblotsky, T. A., Huang, Y., & Gallagher, B. (2010). *Implementation of effective intervention: An empirical study to evaluate the efficacy of Fountas & Pinnell's Leveled Literacy Intervention system (LLI).* Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Renaissance Learning. (2009). *STAR Early Literacy: Technical manual*. Wisconsin Rapids, WI: Author.

Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy, 85*(3), 257-268.

Sterbinsky, A., & Ross, S. M. (2003). *Literacy Observation Tool reliability study*. Memphis, TN: University of Memphis, Center for Research in Educational Policy.

Tabors, P. O., Snow, C. E., & Dickinson, D. K. (2001). Homes and schools together: Supporting language and literacy development. In D. K. Dickinson & P. O. Tabors (Eds.), *Beginning literacy with language: Young children learning at home and school* (pp. 313-334). Baltimore: Brookes.

U.S. Department of Education. (2011). *What works clearinghouse procedures and standards handbook (version 2.1)*. Retrieved from http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=19

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review, 36*(4), 541-561.

ISBN 978-0-325-05938-9