

ABSTRACT

LEVY, REBECCA ABAGAIL. A Control-Group Evaluation of the HELPS-Small Group Reading Fluency Program. (Under the direction of Dr. John Begeny).

A large percentage of elementary-aged students in the United States are struggling with their reading skills—including oral reading fluency performance. Research is needed to examine the effectiveness of small-group reading fluency interventions because reading is a necessary skill for succeeding in adulthood, schools in the U.S. are increasingly challenged to do more with fewer resources, and little research has examined the effectiveness of reading fluency intervention programs that can be implemented with small groups (rather than one-on-one). This current study ultimately sought to examine the effectiveness of the HELPS-SG program as implemented by intervention teachers with their elementary-aged students ($n=18$) as compared to a wait-list control group ($n=24$). Participants' reading performance was assessed on five different outcome measures (e.g., sight word efficiency, phonemic decoding, WCPM, fluency, and comprehension) at pretest and posttest. These five different measures were then converted into pre and post factor, or composite, scores. Using these factor scores, a RM-ANOVA was conducted to compare the intervention and control groups at each time point. However, no treatment effects were found, which may have been because the program truly was not any more effective than the instruction received by the control group or the findings may have been influenced by the multiple limitations associated with sampling, adherence to the timeline and project plan, and poor implementation integrity.

© Copyright 2014 by Rebecca Abigail Levy

All Rights Reserved

A Control-Group Evaluation of the HELPS-Small Group
Reading Fluency Program

by
Rebecca Abigail Levy

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Psychology

Raleigh, North Carolina

2015

APPROVED BY:

John Begeny
Committee Chair

Scott Stage

Kristin Conradi

BIOGRAPHY

Rebecca Abigail Levy was born on November 18, 1990 in Mineola, NY. She graduated from Sanford H. Calhoun High School in 2008, and continued her higher education at Syracuse University, where she worked towards her Bachelor's degree. She graduated from Syracuse University in 2012 with a Bachelor of Arts in Psychology. Following graduation, Rebecca moved to Raleigh, North Carolina to commence her graduate training in School Psychology. After obtaining her Master of Science degree in Psychology, Rebecca will continue her doctoral training at North Carolina State University.

TABLE OF CONTENTS

LIST OF TABLES	vi
A Control-Group Evaluation of the HELPS-Small Group Reading Fluency Program	1
Oral Reading Fluency.....	2
HELPS One-on-One Program.....	3
Lack of School Resources.....	5
Small-group Reading Interventions	6
Importance of Manualized Interventions	9
RTI, Manualized Interventions, and Small-group Interventions.....	11
Ecological Validity	11
Purpose of Study	12
Method.....	13
Research Design	13
Participants.....	13
Student participants.....	13
Assigning students to HELPS-SG.....	14
Teachers and observers.....	15
Setting.....	17
Reading Assessment Materials	18
DIBELS-Next Oral Reading Fluency (DORF).....	18
Gray Oral Reading Test, Fifth Edition (GORT-5).....	19

Test of Word Reading Efficiency, Second Edition (TOWRE-2)	20
Instructional Materials	21
HELPS Curriculum	21
HELPS-SG implementation materials	21
Procedures	22
Assessments	22
Teacher training	23
HELPS-SG procedures	24
Individual assessments (Steps 1-2)	25
HELPS-SG instruction (Steps 3-13)	26
Results	27
Statistical Analysis Strategy	27
Principal Component Factor Analysis	28
Repeated-Measures Analysis of Variance	30
Implementation Integrity	30
Inter-rater Agreement	32
Discussion	32
Limitations and Future Directions	33
Sampling limitations	33
Timeline and project adherence problems	36
Poor integrity with program implementation	37

Future directions	40
REFERENCES	44
APPENDICES	58
Appendix A: HELPS Program for Small Groups Implementation Flow Chart	59
Appendix B: Scripted Directions.....	60
Appendix C: Implementation Protocol.....	63
Appendix D: Checklist for Individual Assessment Implementation	66
Appendix E: Checklist for Small-group Implementation.....	67
Appendix F: Individual Progress Graph.....	70
Appendix G: Individual Progress Tracking Form	71
Appendix H: Group Star Chart.....	72
Appendix I: Group Progress Tracking Form.....	74
Appendix J: HELPS Placement Assessment.....	75

LIST OF TABLES

Table 1	Reading Measures Means and Standard Deviations by Group at Pretest and Posttest.....	42
Table 2	Repeated Measures ANOVA of Group by Time on Reading Performance...	43

A Control-Group Evaluation of the HELPS-Small Group Reading Fluency Program

According to the National Center for Education Statistics (NCES; 2011), when students perform at the Proficient level, their achievement “represents [a] solid academic performance” (p. 6). However, in the fourth grade assessment of U.S. students’ reading skills, only 34% scored at or above the Proficient level. Equally concerning, 33% of these students are reading *below* even a Basic level. In the area of reading fluency, a critical component of early literacy development, the most recent national data show that approximately 40% of fourth graders are “nonfluent” readers (Daane, Campbell, Grigg, Goodman, & Oranje, 2005). Furthermore, the Common Core State Standards (CCSS) support the need for fluency development beginning in the early elementary grades (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). There is an apparent discrepancy between the CCSS’s emphasis on fluency instruction and the actual fluency performance of U.S. students. The aforementioned information is important to highlight because academic performance in elementary school, specifically reading achievement, is not only important for a student’s childhood, but can influence the course of his or her life.

Academic achievement in elementary school can predict high school dropout rates (Alexander, Entwisle, & Horsey, 1997), and students who are poor readers are more likely to dropout of high school than their peers who are typical readers (Daniel, Walsh, Goldston, Arnold, Reboussin, & Wood, 2006). Students who do not complete high school tend to be less successful in the job market and participate in more criminal behavior in adulthood than their peers who graduate from high school (Chapman, Laird, Ifill, & KewalRamani, 2011).

Furthermore, students who dropout of high school often miss out on opportunities that would otherwise be available to them if they earned a high school diploma: attending college and following certain career paths. For example, in 2010, young adults who earned a bachelor's degree made 114% more (based off of median incomes) than young adults who did not graduate from high school (Aud et al., 2012). It is also important to note that the occupations that are currently increasing in the U.S. workforce have higher literacy requirements than the jobs that are decreasing (Barton, 2000); students who are poor readers may have difficulty excelling in a job market that has increased literacy qualifications. Because performing well in reading and ultimately graduating from high school is important for an individual's success in life, this study will focus on improving young struggling readers' oral reading fluency performance.

Oral Reading Fluency

Oral reading fluency (ORF) is defined as a student's ability to read with "speed, accuracy, and proper expression" (National Institute of Child Health and Human Development [NICHD], 2000, p. 3-1). The Instructional Hierarchy—developed by Haring and Eaton (1978)—explained that becoming fluent in a skill (i.e., reading) happens progressively with practice drills and reinforcement only after acquiring the skill. Indeed, effortful and labored oral reading causes students to read in a disconnected manner (Chard, Vaughn, & Tyler, 2002), which can make it difficult to comprehend the text (Daane et al., 2005; Fuchs, Fuchs, Hosp, & Jenkins, 2001). Furthermore, it takes "dysfluent" readers a longer time to read a passage than their fluent peers, and this may prevent struggling readers from wanting to practice their reading skills (Daly, Chafouleas, & Skinner, 2005).

Given the importance of reading fluency and the large number of students who do not read fluently, educators and reading researchers have developed and evaluated a range of ORF interventions and strategies. Evidence-based ORF interventions include the following: modeling, repeated readings (RR), phrase-drill (PD), progress monitoring, goal setting, cued-reading, feedback, and reinforcement (Morgan & Sideridis, 2006; Therrien, 2004). Therrien (2004) conducted a meta-analysis of RR intervention studies and found that certain instructional components added to the RR intervention increased mean fluency effect sizes. Effective add-on strategies included: (a) cueing participants to focus on speed and comprehension ($ES = .94$), (b) working with adult tutors ($ES = 1.37$), (c) receiving corrective feedback ($ES = .51$), (d) setting a performance criterion ($ES = 1.70$), and (e) charting students' progress ($ES = .57$). Interventions such as these improved fluency for both learning disabled and non-learning disabled participants (Chard et al., 2002; Therrien, 2004). Furthermore, Morgan and Sideridis (2006) conducted a meta-analysis of the effectiveness of fluency interventions from single-subject design studies and found that goal setting, feedback, and contingent reinforcement were effective in improving students' reading fluency.

HELPS One-on-One Program

The Helping Early Literacy with Practice Strategies (HELPS) Program is a one-on-one (adult-student) reading fluency intervention that integrates all eight of the instructional strategies shown to be effective in prior meta-analyses (e.g., Morgan & Sideridis, 2006; NICHD, 2000; Therrien, 2004). HELPS is a manualized reading intervention that has effectively increased ORF in elementary-aged students (e.g., Begeny, 2011; Begeny, Laugle, et al., 2010; Begeny, Mitchell, Whitehouse, Samuels, & Stage, 2011; Begeny, Ross, Greene,

Mitchell, & Whitehouse, 2012). For example, using 2nd grade participants, Begeny et al. (2010) compared the HELPS program to the Great Leaps Reading Program (Mercer & Campbell, 1998) and a control group, and found that only students receiving HELPS significantly outperformed the control group. Significant differences were found with measures of reading fluency, reading comprehension, sight word efficiency, and phonemic decoding; effect sizes ranged from 0.62-1.02. In another study, Begeny, Mitchell and colleagues (2011) used an experimental control group design to examine the effects of HELPS on struggling readers when implemented by school personnel. Students in the HELPS condition performed significantly better on all assessed measures of reading performance compared to students in the control group, with effect sizes ranging from 0.50-1.05. This study also showed that both classroom teachers and teacher assistants could implement HELPS with strong implementation integrity following relatively brief training workshops and some follow up implementation feedback sessions.

Despite the availability of HELPS and a small number of other reading fluency intervention programs (e.g., Mercer, Campbell, Miller, Mercer & Lane, 2000), these mostly one-on-one formatted interventions present feasibility challenges for many schools because they require a relatively large amount of personnel resources (i.e., one educator for every student receiving the intervention). This high need for school resources is particularly important given the budget cuts and increased time demands that schools and educators currently face. The personnel challenges are also important to consider when there is such a high percentage of students (approximately 40%; Daane et al., 2005) who would likely benefit from fluency-based interventions.

Lack of School Resources

According to Oliff, Mai, and Leachman (2012), more than two-thirds of U.S. states are providing less funding to schools than they did in the year 2008, and budget cuts are causing schools to reduce the amount of time spent in the classroom. Due to less funding, teachers may be experiencing a substantial reduction in time and resources. The Metropolitan Life Insurance Company (MetLife; 2012) annual survey results indicate that 63% of U.S. teachers reported an increase in class size from the previous year. Furthermore, an annual survey of North Carolina teachers discovered that 38.5% of teachers believe class sizes are too large to have enough time to meet all of their students' needs (Hirsch & Maddock, 2012). It can be concluded from the aforementioned reports that teachers' ability to do their jobs effectively and efficiently (e.g., improve students' reading performance) can be increasingly more difficult if budget cuts are prevalent, class sizes are increasing, and class time is decreasing.

According to Begeny, Schulte, and Johnson (2012), teachers need more effective and time efficient interventions and systems to best help their students succeed academically. With all of the difficulties that teachers are facing in the classroom, small-group interventions represent one practical solution (Foorman & Torgesen, 2001). Instruction for small-groups is considered beneficial because it provides students with (a) more individualized attention than they would otherwise receive from classroom instruction; and (b) compared to whole-class instruction, a smaller and more manageable setting for mutual communication and learning between the students and teacher (Wasik, 2008). Small-group interventions have the potential

to save teachers time and increase classroom efficiency by working with more than one struggling student at a time.

Small-group Reading Interventions

Despite the strong research evidence showing the effectiveness of one-on-one ORF interventions (Chard et al., 2002; Morgan & Sideridis, 2006; Therrien 2004), initial research suggests that small-group reading interventions may be just as successful for some students. Vaughn et al. (2003) found that small-group reading interventions—when implemented with groups of 3 students—tend to be just as effective as one-on-one reading interventions on students' reading fluency performance. Additionally, Gersten et al. (2008) recommend conducting reading interventions in a small-group format (3-4 students), and argue that small groups are “considered more practical for implementation” (p. 19).

Although there are many effective small-group reading interventions that can be found in the literature, few focus on reading fluency. In an exhaustive review of the empirical literature, only 10 studies were found that examined the effects of small-group reading fluency interventions (Begeny, Braun, Lynch, Ramsay, & Wendt, 2012; Begeny, Hawkins, Krouse, & Laugle, 2011; Begeny, Krouse, Ross, & Mitchell, 2009; Begeny & Martens, 2006; Begeny & Silber, 2006; Begeny, Yeager, & Martínez, 2012; Bonfiglio, Daly, Persampieri, & Andersen, 2006; Klubnik & Ardoin, 2010; Kuhn, 2005; Ross & Begeny, 2011). For example, Kuhn (2005) compared the influence of small-group reading fluency strategies on fluency and comprehension measures to a control group in elementary-aged students. This study found that students who received a variation of the RR intervention or a wide-reading intervention—“choral reading of a given text” (p. 135)—made distinct fluency

improvements, but only students who received the wide-reading intervention made substantial comprehension improvements. In another study, Bonfiglio et al. (2006) conducted an experimental analysis in order to determine which small-group intervention package would be the most efficient while remaining effective for the participating students. The identified intervention package—which included taped-preview, error correction, and choral reading—was the most effective and efficient for improving students’ reading fluency, as measured in word correct per minute (WCPM). Also using a single-case experimental design, Begeny and Martens (2006) used an ABAB multiple-baseline design to examine the effects of a small-group ORF intervention across groups. The intervention consisted of multiple commonly used fluency strategies: choral reading, listening passage preview (LPP; similar to modeling), RR, and PD. Clear increases in WCPM from baseline to intervention indicated that the group fluency intervention improved students’ ORF.

More recently, Begeny, Braun and colleagues (2012) conducted a one-group pretest-posttest quasi-experimental design with a double pretest in order to test if seven elementary-aged participants’ reading achievement significantly improved after receiving the only known manualized small-group reading intervention that targets fluency (i.e., the HELPS Program for Small Groups; HELPS-SG). The pretest score was students’ growth from the fall ORF benchmark to winter ORF benchmark (i.e., from pretest 1 to pretest 2). The comparative posttest score was calculated as students’ growth from the winter ORF benchmark to spring ORF benchmark (i.e., from pretest 2 to posttest, or the time during which students received HELPS-SG). A nonparametric Wilcoxon two-tailed test determined there was a significant difference between student reading growth across the assessment

periods, favoring students' growth from pretest 2 to posttest and therefore providing some preliminary evidence for HELPS-SG ($T= 0, p= .02$). The authors also determined that the posttest scores of all but one student exceeded nationally derived norms of ORF growth. Although these findings offered preliminary evidence for HELPS-SG, three major limitations were reported. The researchers had a limited sample size ($N= 7$), did not utilize a control group, and only used one outcome measure of reading (i.e., growth in ORF from a validated and commonly used ORF assessment).

Collectively, the 10 small-group intervention studies that target ORF is promising because in all published studies at least some of the students benefit from the intervention, and in some studies the researchers found that there are students who benefit as much from small group instruction as they do from comparable one-on-one instruction (e.g., Begeny, Hawkins et al., 2011; Begeny, Yeager, & Martínez, 2012; Klubnik & Ardoin, 2010; Ross & Begeny, 2011). Despite this promise of small-group ORF interventions, there are some important limitations of the existing research. First, all but two of these studies (Begeny, Braun et al., 2012; Kuhn, 2005) used a single-case design to explore the effects of reading strategies. Methodologically sound single-case designs certainly provide experimental evidence of intervention effects, but one important critique of single-case designs is that they may not allow for as much generality as methodologically sound group designs (Kazdin, 2011); at minimum, a solid research-base for an intervention approach should include a mix of both single-case and group designs. Equally important, of the two group design studies the intervention groups were relatively small (5-6 in Kuhn, 2005; 7 in Begeny, Braun et al.,

2012). Clearly, more group design studies are needed to examine small-group interventions that target ORF.

Another important limitation of this research is that only one study (Begeny, Braun et al., 2012) examined a manualized intervention program. All other studies used intervention strategies and materials that were developed or used for the sole purpose of the research; those studies do not describe available materials or clear directions for day-to-day intervention implementation. The importance of this limitation is evidenced below.

Importance of Manualized Interventions

If an intervention is insufficiently described for teachers' use, then the teacher will most likely have a difficult time implementing the intervention with integrity. Manualized interventions are important for use in schools because they (a) provide teachers with specific procedures and detailed implementation instructions (Simmons et al., 2011) that may facilitate high integrity of the intervention across teachers; (b) reduce teacher idiosyncrasies (Schoenwald, Carter, Chapman, & Sheidow, 2008); (c) promote consistency across teachers, which can minimize teacher interference of a research-supported intervention (Reeves, 2010); (d) may lessen the research to practice gap by translating supported research interventions to a descriptive manual for teachers' use; and (e) typically provide teachers with all the needed materials to implement the intervention. For the purpose of this study, a manualized intervention program will be defined as: an intervention program that supplies the teacher with clear scripted directions and explicit support materials in order to increase teacher integrity (Begeny, Ross et al., 2012; Begeny, Schulte, & Johnson, 2012, p. 30; Chambless & Ollendick, 2001)

Of these reasons for using manualized interventions in schools, there is also empirical research evidencing its importance. For example, Schoenwald and colleagues (2008) conducted a study examining the maintenance outcomes of a popular manualized therapy—Multisystemic Therapy (MST)—on youth’s behavior change. The researchers focused on therapist adherence to the MST protocols; they found that participants whose therapists followed the guidelines more strictly improved more so than other participants. Although this article did not focus on a school-based intervention, it supports the use of manualized interventions with youth.

Furthermore, two research articles (e.g., Reeves, 2010; Simmons et al., 2011) support the importance of using *scripted* and *commercial* interventions in schools, which will be referred to as *manualized* from this point forward. Through a case report of two elementary teachers, Reeves (2010) found that manualized programs minimize teacher interference of the curriculum, have the potential to ensure standardization of implementation, and are becoming increasingly popular with teachers. In addition, Simmons and colleagues (2011) compared an experimental manualized reading intervention to a typical classroom reading intervention. This study concluded that the manualized reading intervention program was more effective at improving participants’ reading performance than teacher-designed supplemental interventions. Furthermore, teachers rated the acceptability of the manualized intervention highly and perceived their interventionist abilities as improved after using the program. With support for using manualized interventions, it is also important to note that these types of interventions can be implemented in accordance with instructional problem-

solving systems such as Response to Intervention (RTI). The next section will discuss the connections between RTI, manualized interventions, and small-group interventions.

RTI, Manualized Interventions, and Small-group Interventions

RTI is a problem-solving framework that is increasingly being used by more schools (Denton, 2012; Gersten et al., 2008). According to Gersten and colleagues, RTI combines “universal screening and high-quality instruction for all students with interventions targeted at struggling students” (2008, p. 4). According to Denton (2012), there are typically three tiers within RTI systems. Tier 1 is class wide instruction, or “primary prevention” (p. 233), with universal screening. Tier 2 is the first level of intervention for struggling students, and provides struggling students with supplemental, often small group, instruction. At this Tier of support, *prescriptive* problem-solving is commonly used (see Begeny, Schulte, and Johnson, 2012, p. 18), which means that a specific research-supported (often manualized) reading intervention is provided for all students who are known to lack the same targeted skills. Lastly, Tier 3 interventions are more intensive and are only necessary when students fail to respond to Tier 1 and Tier 2 instruction, and in this Tier it is more common for individualized (rather than prescriptive) problem-solving. Creating manualized interventions that fit into the RTI framework should enhance schools’ capacity to use RTI successfully, but such interventions can also be well-suited for schools using other approaches to assisting struggling learners (Begeny, Schulte, & Johnson, 2012).

Ecological Validity

As defined by Bronfenbrenner (1977), ecological validity is “the extent to which the environment experienced by the subjects in a scientific investigation has the properties it is

supposed or assumed to have by the investigator” (p. 516). In other words, when a study is ecologically valid the study is conducted in the most appropriate setting for what behavior is being observed. Ecological validity is important because it allows researchers to interpret and apply their findings to the real world because the studies are conducted in a “true environment” (Schmuckler, 1997, p. 421). Furthermore, an ecologically valid study must contain naturalistic features across the setting, stimuli, and observed behaviors (Schmuckler, 1997). In sum, an ecologically valid study will (a) take place in a real-world environment, (b) use materials and people that are typically found in that environment, and (c) measure behaviors that coincide with that environment.

Purpose of Study

The overall rationale for this study is supported by the following: ORF is a necessary skill in students’ reading development, schools often lack resources for one-on-one interventions, there are numerous advantages to using manualized interventions, and small-group reading interventions show promise as an effective practice for improving students’ fluency. However, more studies are needed to demonstrate the effectiveness of these small-group practices—particularly well-powered, ecologically valid, control-group designs. Given this rationale, the purpose of this study is to use a non-randomized control-group design to evaluate the effects of HELPS-SG implemented by teachers with their second and third grade struggling readers.

This study will answer one main research question: will students receiving HELPS-SG outperform the control group on three different areas of reading performance: reading fluency, reading comprehension, and word reading efficiency? Based on preliminary research

suggesting that students who receive the HELPS-SG intervention tend to improve their reading fluency more than expected over time (Begeny, Braun et al., 2012), it is hypothesized that students who receive HELPS-SG will significantly improve their reading performance compared to the students who do not receive the intervention.

Method

The North Carolina State University Institutional Review Board approved all of the methodological procedures prior to commencing this study.

Research Design

A two-group pretest-posttest quasi-experimental design was used to determine if students receiving HELPS-SG would have a positive significant effect on their reading performance compared to students in a wait-list control group.

Participants

Student participants. Because ORF should be targeted in the early elementary grades, participants were in the second and third grades. To be included in the study, participants had to be considered struggling readers according to the participating school. The students that were chosen to participate scored between the 10th and 60th percentile (using national norms; Hasbrouck & Tindal, 2006) on the participating school's administered fall ORF benchmark assessment using the Dynamic Indicators of Basic Early Literacy-Next ORF (DORF; Good, Kaminski, Cummings et al., 2011). Scoring in this range suggests that some students are struggling readers or at-risk for future reading difficulty (Hasbrouck & Tindal, 2006). Teachers confirmed that the identified students were struggling readers and not just students who performed poorly on the DORF assessment because of other reasons.

Students reading below the 10th percentile were not eligible for this study because it is likely these students would benefit from a more intensive reading intervention (e.g., 20-30 minutes of daily one-on-one reading intervention). Furthermore, students were excluded from participating in this study if they received special education services or pullout English as a Second Language (ESL) services. Due to school district-wide policies of not being allowed to share which participants received specific low-intensity (i.e., *transitional*) ESL services, it is uncertain which participants in this study received these transitional ESL services within their general education classrooms. Transitional ESL services are indirect in nature and for those ESL students who are nearly proficient in English; the general education teacher makes all accommodations for ESL students after consulting with the ESL teacher.

Originally 50 students were identified as participants for this study, however 4 participants moved, 3 students received HELPS-SG regularly from their teacher who learned the intervention from a prior research study, and 1 student was later determined to be receiving special education services. This left a total sample size of 42 students; 24 second graders and 18 third graders. Of the 42 participants, 57.1% were White, 16.7% were Black, 14.3% were Hispanic, 7.1% were multiracial, 2.4% were Asian, and 2.4% were Native American. Furthermore, 59.5% ($n= 25$) of the sample was female, and the mean age was 8.18 (range= 7.08-9.67).

Assigning students to HELPS-SG. Lou et al. (1996) suggest that students in homogeneous reading ability groups achieve more than students in heterogeneous groups. Furthermore, Gersten and colleagues (2008) strongly recommend that Tier 2 interventions be implemented in homogeneous groups of 3 to 4 students. Taking the recommendations from

Lou et al. (1996) and Gersten et al. (2008) into consideration, it was expected that the participants in the HELPS-SG condition would be assigned to homogeneous groups of 3 to 4 students based on ability. Reading ability was assessed with the HELPS placement assessment (Appendix J) and the DORF results. Due to the participating school's policies, teachers were in charge of determining which participants they would provide the intervention to. Basing their decisions off of the Text Reading and Comprehension (TRC; Wireless Generation, 2007), a running record type of assessment with un-reported psychometric characteristics, the participating teachers assigned eligible participants to the control or intervention group. Ultimately, 18 students received the intervention and 24 students were assigned to the control group. Despite what was asked of the teachers (to implement the intervention in groups of 3-4), the small groups ranged from 3-6 students.

A chi-square analysis for independence was conducted to ensure that the intervention and control groups were similar in age, ethnicity, and gender. This analysis determined that the intervention and control group did not significantly differ on the measures of age ($\chi^2 (1) = 21.04, p = .46$), ethnicity ($\chi^2 (1) = 6.22, p = .29$), and gender, ($\chi^2 (1) = .206, p = .65$). Additionally, there were no significant differences found between the intervention and control groups at pretest on all of the included measures (which will be discussed below): DORF-Winter ($t (40) = -.97, p = .34$), TOWRE-SWE ($t (40) = .02, p = .99$), TOWRE-PDE ($t (39.37) = -.65, p = .52$), GORT-Fluency ($t (40) = 1.16, p = .25$), and GORT-Comprehension ($t (40) = .53, p = .60$).

Teachers and observers. Two female teachers employed in the school as “intervention teachers” for their school-wide RTI model served as interventionists during the

course of this study. Both of the participating teachers hold master's degrees in education. Teacher A, who worked with all the second grade participants, had been a teacher for 23 years and at the participating school for 16 years. Teacher B worked with all third grade participants, had been a teacher for 16 years, and taught at the participating school for 4 years. Furthermore, Teacher A worked with each of her two groups for 30 total sessions, and Teacher B worked with one of her groups for 25 sessions and her other group for 26 total sessions. In sum, students in the intervention group received an average of 25.94 HELPS-SG sessions (range= 21-30; $SD= 3.15$). An independent t-test was conducted to determine if intervention group participants in the two different grades received a different number of HELPS-SG sessions. Results indicated that second grade participants ($M= 28$, $SD= 1.95$) received significantly more sessions than the third grade participants ($M= 22.71$, $SD= 1.38$; $t(16)= 6.22$, $p < .001$).

Trained undergraduate research assistants (RAs) served as observers of teachers' implementation integrity throughout the course of the study. To be considered a trained RA, they had to (a) attend regular research team meetings for training and scheduling purposes, and (b) be proficient in all trained assessments and procedures. In order to be considered proficient in the assessment procedures, all of the RAs had to achieve 100% criterion in implementation procedures with the primary investigator during mock one-on-one assessment sessions; the RAs also had to demonstrate mastery over the assessment procedures. In order to be considered a trained observer, the RAs had to be able to (a) implement HELPS-SG core procedures with 100% integrity 3 times within 5 successive sessions, (b) implement HELPS-SG tips and reminders protocol with 90% integrity 3 times

within 5 consecutive sessions (Begeny, Upright, Easton, Ehrenbock, & Tunstall, 2013), (c) correctly identify HELPS-SG core procedural mistakes with 100% integrity for 3 observed sessions, and (d) correctly identify HELPS-SG tips and reminders procedural mistakes with 90% integrity for 3 observed sessions. Furthermore, all RAs were trained and considered proficient in teacher feedback procedures for a prior research study evaluating the effects of feedback on teacher implementation integrity of HELPS-SG.

Setting

This study took place within the participants' K-5 public elementary school located in a suburban area in southeastern United States. Approximately 910 students are enrolled in this school, and 18% of these students receive free or reduced price lunch. HELPS-SG sessions were implemented in a classroom that was free from noise, other distractions, and away from wait-list control group participants. Furthermore, all student participants maintained their participation in typical classroom instruction during this study; the only difference between groups was that students in the intervention group were pulled out of class during non-core instructional times to receive HELPS-SG. During this time, the control group participants had focused reading instruction with their classroom teachers. Focused reading instruction consisted of students either working in small groups with their classroom teacher, independently reading, or reading with a reading buddy. Besides the students who had to be excluded from the study due to their classroom teacher using HELPS-SG, the research team was not aware of other control group participants' teachers using HELPS-SG over the course of this study.

Reading Assessment Materials

DIBELS-Next Oral Reading Fluency (DORF). The DORF (Good, Kaminski, Cummings et al., 2011) is a measure of ORF and was developed according to the standards, procedures, and research associated with Curriculum-based Measurement (CBM; Deno, 1989). During each DORF assessment a student is presented with grade-level reading material and asked to read three consecutive passages, one minute per passage, while the assessor listens. During each timed reading, the assessor records any errors (e.g., substitutions, omissions, hesitations for 3 or more seconds) the student makes and ultimately determines the number of words read correctly per minute (WCPM). The median WCPM score represents the overall DORF score.

According to Good, Kaminski, Dewey et al. (2011) the DORF is a reliable measurement of ORF. All of the reported reliabilities for second and third grade are above .90. The two-week alternate-form reliability coefficients range from .96-.97, the median reliability for immediate alternate-form coefficients range from .94-.97, and the two-week test-retest reliability coefficients range from .91-.93. The inter-rater reliability for both grade levels is .99.

DORF WCPM is considered a valid measure of ORF. Criterion-related validity was calculated based on students' WCPM median score compared with the Group Reading Assessment and Diagnostic Evaluation (GRADE) Total Test. The predictive validity coefficients for the beginning of the year for the two grade levels range from .66-.69, the predictive validity coefficients for the middle of the year range from .67-.76, and the concurrent validity coefficients for end of year range from .66-.73.

Gray Oral Reading Test, Fifth Edition (GORT-5). The GORT-5 (Wiederholt & Bryant, 2012) is an individual, standardized, and norm-referenced text reading fluency measure that takes an average of 20-30 minutes to administer. This assessment examines accuracy and rate while also looking for prosodic features of a student's ORF. The GORT-5 has two equivalent forms (A and B); each form has 16 reading passages with 5 comprehension questions associated with each passage, but students may not read all 16 passages due to basal and ceiling rules. Each administration of the GORT-5 yields multiple scores: rate, accuracy, fluency (combination of rate and accuracy), comprehension, and oral reading index (ORI; combination of fluency and comprehension score). The fluency and comprehension scores will be analyzed in this study.

The GORT-5 is considered to be a reliable measurement of reading fluency and comprehension. The coefficient-alpha for both forms A and B is .93 for fluency and .94 for comprehension. The alternate forms immediate reliability between forms A and B is .94 for fluency and .87 for comprehension. Test-retest reliability, which ranged from 1-2 weeks in delayed administration of the same form, is .90 for fluency and .82 for comprehension. Alternate forms reliability between form A and B (administered with a 1-2 week delay) is .88 for fluency and .77 for comprehension. The inter-scorer reliability for fluency and comprehension is .99.

GORT-5 is also considered a valid measurement of fluency and comprehension. As reported in the technical manual (Wiederholt & Bryant, 2012), the GORT-5 average coefficients of magnitude were calculated for fluency (.73) and comprehension (.74); both

coefficients were considered to be strongly correlated with other commonly used reading measures.

Test of Word Reading Efficiency, Second Edition (TOWRE-2). The TOWRE-2 (Torgesen, Wagner, & Rashotte, 2012) assesses a student's reading efficiency by having the student read real words (sight words) and nonwords (phonemic decoding). Consisting of two subtests—sight word efficiency (SWE) and phonemic decoding efficiency (PDE)—the TOWRE-2 is administered in approximately 5 minutes. The total word reading efficiency index (TWRE) is a combination of the two subtests, but in this study SWE and PDE will be reported and analyzed as separate measures. During each assessment session, the assessor has the student (a) read the SWE practice word list, (b) read the SWE test list for 45 seconds, (c) read the PDE practice nonword list, and (d) read the PDE test list for 45 seconds. During each subtest, the student is instructed to read as many words or nonwords as he or she can in the allotted time. The student's score is the total words read correctly in the 45-second sessions.

Torgesen, Wagner, and Rashotte (2012) reported that the TOWRE-2 is a reliable measure of reading efficiency. The average alternate-forms immediate reliabilities for the TOWRE-2 are: .91 for SWE, .92 for PDE, and .95 for TWRE. The TOWRE-2 provides 4 different testing forms. Forms A and B will be used in this study because they have strong immediate alternate form reliabilities: .90 for SWE and .92 for PDE. The average test-retest same forms reliability—administered 2 weeks apart—is .91 for SWE, .90 for PDE, and .93 for TWRE. The average test-retest alternate forms reliability is .87 for both SWE and PDE, and .92 for TWRE. The inter-scorer reliabilities across all measures are .99.

The TOWRE-2 is supported as being a valid measurement of sight word efficiency and phonemic decoding efficiency. Criterion-validity was calculated by comparing multiple reading measures (e.g., DIBELS NWF, DORF, GORT-4 Fluency) to the TOWRE-2 subtests. The overall coefficient averages for reading measures were strongly correlated with SWE (.90) and PDE (.89).

Instructional Materials

HELPS curriculum. As previously discussed, the HELPS program uses eight common strategies (e.g. Therrien, 2004; Morgan & Sideridis, 2006) to improve students' reading fluency. The HELPS curriculum (Begeny, Mann, Cunningham, & Tsuen, 2009), which was developed to use with the available HELPS programs (one-on-one and small-group), includes 100 reading passages with predetermined difficulty levels. The passages become slightly more difficult as a student moves through the curriculum, with Spache (1953) readability values ranging from 1.7-4.9. Passages are either narrative (i.e., fiction) or expository (i.e., non-fiction) and do not include pictures. All the Dolch sight words (Dolch, 1936) are included throughout the curriculum. Varying in types of narrators, verb tenses, and topics, the passages are intended to be age-appropriate and interesting for elementary-aged readers while giving them repeated practice with frequently used (e.g., Dolch) and novel words.

HELPS-SG implementation materials. In addition to the HELPS curriculum, the HELPS-SG instructional materials include (a) implementation protocols, (b) an individual and small group progress tracking form, (c) an individual progress graph for each student, (d) a group star chart for each small group, and (e) examiner copies of the reading passages

found in the curriculum. Four primary protocols make up the guidelines for implementing HELPS-SG correctly: (1) the implementation flow chart (Appendix A) consists of 13 core procedures of the HELPS-SG program; (2) the scripted directions (Appendix B) provide the teacher with a procedural script of the specific directions to say at each core step; (3) the extended implementation protocol (Appendix C), which helps the teacher learn and better understand the core procedures; and (4) the checklist for implementation (Appendix E) which acts as a procedural checklist of the core implementation adherence and the implementation quality (Tips and reminders; i.e., teaching characteristics and methods that implementers should use in order to best engage the students).

Procedures

Assessments. Approximately two and a half months prior to students being assigned to experimental group, the TOWRE-2 SWE and PDE measures (form A) and the GORT-5 Fluency and Comprehension measures (form A) were administered. At posttest, which began the day following the last intervention day, all participants were administered the alternate forms of the assessments that were used at the beginning of the study. Specifically, at posttest participants received the TOWRE-2 SWE and PDE measures (form B), and the GORT-5 Fluency and Comprehension measures (form B). All assessments were administered in assessor-student dyads by the primary investigator and other trained data collectors (graduate school psychology students and RAs) and occurred in an empty classroom in the participating school that was free from noise and distractions. Furthermore, the participating school conducted the DORF fall, winter, and spring benchmarks that were provided to the primary investigator of this study. The winter benchmarks were given 2-3 weeks prior to

students receiving intervention and the spring benchmarks within 2-3 weeks of the last day of HELPS-SG intervention.

Teacher training. The primary investigator met HELPS-SG implementation criteria through training provided by the HELPS-SG developer and is considered “trained” because she is (a) able to implement HELPS-SG core procedures with 100% integrity 3 times within 5 successive sessions, and is (b) able to implement HELPS-SG tips and reminders protocol with 90% integrity 3 times within 5 consecutive sessions (Begeny, Upright et al., 2013). The participating teachers were trained during 2 separate 1.5-2 hour HELPS-SG workshops. One workshop occurred two months before the start date of the study, and the other workshop occurred one week before the start of the study. As part of research guidelines, teachers were told not to discuss the HELPS-SG procedures with any non-participating teachers.

The HELPS-SG workshops consisted of instructing teachers how to use the program (e.g., implementation protocol, scripted directions) and the associated forms (e.g., Star Chart, tracking forms, flowchart), and opportunities for practice with immediate corrective feedback. Teachers had the opportunity to practice implementing HELPS-SG with the RAs and received feedback from the principal investigator and the RAs they practiced with. At the end of the training workshop, teachers were allowed to keep all provided implementation and practice materials for reference.

Following the workshop portion of training, teachers’ implementation of HELPS-SG was observed in vivo by RAs for every session during the first three weeks of the teachers implementing HELPS-SG with their participating students. After each observation, the RAs provided the teachers with specific corrective feedback about their implementation until they

were able to implement HELPS-SG core procedures with 100% integrity 3 times within 5 successive sessions, and were able to implement HELPS-SG tips and reminders with at least 80% integrity 3 times within 5 consecutive sessions. To maintain teachers' accurate implementation of HELPS-SG, for the following three weeks of the study, the teachers were observed for one session a week. After these sessions, the RAs provided corrective feedback. These three additional sessions of feedback were provided to the teachers to ensure proper implementation of HELPS-SG due to several unexpected events in the school that seemed to have a significant impact on the normal school routine and schedule. Specifically, during the first six weeks of the study the school environment (as reported regularly by school staff and administration) was frenzied due to bad-weather days, delayed starts, early dismissals, field trips, and an atypically high number of student and staff absences due to illness. After the first six weeks of the study, the school climate reportedly returned to normal and teachers no longer received feedback from the RAs. Even though the school environment was frenzied for the first half of the study, HELPS-SG was reportedly implemented in a quiet classroom away from noise and other distractions.

HELPS-SG procedures. HELPS-SG is a manualized intervention program; each of the 13 steps (see Appendix A) and materials are explicitly stated and explained in the corresponding manual (Begeny, 2013) and implementation protocols (Appendix C). Excluding the brief (approximately 2-5 minutes per student) individual assessments procedure that precedes the small group procedures (described below), the small group procedures were implemented in approximately 15-20 minutes sessions. Due to feasibility complications and larger than recommended small group sizes, participating teachers were

unable to find time to conduct the individual assessment piece of HELPS-SG prior to each group session. As a way to provide support to the school and teachers, undergraduate volunteers who are trained in HELPS-SG procedures by the primary author conducted the brief individual assessments for each small group in the morning each day the students received the intervention from their teachers. The teachers then implemented the group portion of HELPS-SG with their small-groups later in the day.

Individual assessments (Steps 1-2). As shown in the flowchart (Appendix A), each HELPS-SG session began with brief individual assessments—the only individualized portion of HELPS-SG. During the individualized assessments (similar to the DORF assessment procedures), each student read aloud the provided passage and the volunteer then graphed the student's WCPM and words incorrect per minute (WIPM). After graphing the student's individual progress (Appendix F) and writing down the student's performance on the progress tracking form (Appendix G), the volunteer provided feedback and praise for the student's reading improvements. Each student either (a) reached his or her individual reading goal (IRG), or (b) did not reach his or her IRG (see Appendix A for the IRG table). The individual assessment provides a necessary element of student progress monitoring within the HELPS intervention that cannot be achieved in a small group. This progress monitoring assessment is important for a teacher's day-to-day instructional decisions with HELPS as well as for monitoring the potential for generalized learning outcomes (Begeny, Whitehouse, Methe, Coddling, Stage, & Neupert, 2013).

HELPS-SG instruction (Steps 3-13). All core procedures are shown in the full implementation protocol (Appendix C) and summarized in the flowchart (Appendix A). However, to ensure clarity the procedures are also summarized below.

Once individual assessments for each participant were completed, the teacher then determined which flowchart path to follow. The students would reach the group reading goal (GRG) if at least half of the students met their IRG, and not meet the GRG if less than half met the IRG. After determining whether the students meet the GRG, the teacher will review the *you/me game*. The *you/me game* consists of students (“you”) and the teacher (“me”) earning points. Students were able to earn points if they are paying attention, following along, and putting in effort throughout the HELPS-SG session. Conversely, the teacher earned points if the students were *not* paying attention, following along, or putting in effort. Before the session began (at Step 5) a predetermined amount of points needed to win the *you/me game* was established and stated by the teacher.

Each group oral reading of the day’s passage (Steps 5, 8, and 10) involved the teacher randomly calling on a student in the group to read 1 to 3 sentences of the passage, which continued until the entire story was completed. When it was not a student’s turn to read aloud, he or she was instructed to follow along silently and wait for his or her turn. When a student was called on to read, this illustrates one example of how the student group (you) can earn a point in the *you/me game*. If the student knew exactly where to begin reading, the students earned a point; if the student did not know where to start reading, the teacher earned a point.

Passage retell (Step 6) is when the students are instructed to reiterate what they remember from the passage, in sequential order. One student was randomly selected to say what happened in the beginning, another student was asked to say what happened next, and so forth until all students were called on once. During modeling (Step 7), the teacher read the entire passage aloud while sporadically pausing and calling on a random student from the group to read the next word in the passage. During the phrase-drill error correction procedure (Step 9) the students chorally practiced 3-7 word phrases that contained a word that at least one student read incorrectly during that session's oral reading of the passage (based on Steps 5 and 8). Once the teacher concluded steps 1-10, she provided praise and feedback on the group's reading performance and wrote in 1-3 stars on the group star chart (Appendix H). One star was earned if the students tried their best but did not win the you/me game or reach the GRG; two stars were earned if the students won the you/me game and tried their best; and three stars were earned if the students reached the GRG, won the you/me game, and worked their hardest. Lastly, the teacher wrote down all the information from that day's session on the group progress tracking form (Appendix I).

Results

Statistical Analysis Strategy

To ensure that this proposed study had adequate statistical power (a minimum of .80), an a priori power analysis (Faul, Erdfelder, Lang, & Buchner, 2007) was conducted with the effect size set to .35. Results of this power analysis revealed that at least 20 total participants were needed to have a minimum power of .80 when running the following analyses. A post hoc power analysis revealed that the statistical power for this study was .99,

and is considered to have sufficient power to find results, if there are any to be found, in the following analyses. Furthermore, it is important to note, that all following analyses were conducted using the Statistical Package for the Social Sciences (SPSS; IBM Corp., 2010).

Principal Component Factor Analysis

A principal component factor analysis (PCA) was conducted to examine whether all reading measures are psychometrically structured as a unified construct, and thus can be reduced down from 5 separate dependent measures to 1 unified measure of “reading performance.” The PCA included the following pretest measures: DORF-Winter, TOWRE-SWE, TOWRE-PDE, GORT-Fluency, and GORT-Comprehension. The results of this analysis showed a sufficient sample for factor analysis (Kaiser-Meyer-Olkin= .801), explaining 66.8% of the variance, and Bartlett’s test of Sphericity ($\chi^2(10) = 117.63, p < .001$) indicated correlations were significantly strong enough to be analyzed together. This analysis determined all of the above-mentioned reading measures makeup a unified construct with principal components or factor loadings ranging from .709-.929, which far exceed the minimum recommendation of .60 (Tabachnick & Fidell, 2007). The results of this PCA indicate that all five of the reading measures can be analyzed together as one unified measure of “reading performance” by using factor scores (Thompson, 2004). Factor scores can be considered a composite score for each participant (Odum, 2011), and are yielded when a PCA is run. The regression approach was used to provide factor scores for all participants, creating a new unified *pretest-reading competence* dependent variable. This method was used because all three of the factor score methods available yield the same scores when used

in a PCA (Odum, 2011; Thompson, 2004), and this method is most widely used and understood (Tabachnik & Fidell, 2007).

Because one unified measure of reading performance was created for the first time point (i.e., pretest), a second PCA was run to provide a unified measure of reading competence for the second time point (i.e., posttest). Similar to the first PCA, the second PCA included the following posttest measures: DORF-Spring, TOWRE-SWE, TOWRE-PDE, GORT-Fluency, and GORT-Comprehension. The results of this second analysis showed a sufficient sample for factor analysis (Kaiser-Meyer-Olkin= .796), explaining 64.03% of the variance, and Bartlett's test of Sphericity ($\chi^2(10) = 101.21, p < .001$) indicated correlations were significantly strong enough to be analyzed together. This analysis determined all of the above-mentioned posttest reading measures make up a unified construct with principal components or factor loadings ranging from .695-.908. For this second PCA, the regression approach was also used to create a new unified measure of *posttest-reading performance*. These two new unified variables (i.e., pretest and posttest reading performance) will be used in a repeated measures analysis of variance (RM-ANOVA) to evaluate differences between groups' reading performance at pretest and posttest (Tabachnick & Fidell, 2007; Thompson, 2004). Prior to conducting the RM-ANOVA with the two new unified variables, descriptive statistics were run. The means and standard deviations for all 5 reading measures and reading performance factor scores separated by experimental group are reported in Table 1 below.

Repeated-Measures Analysis of Variance

A RM-ANOVA was conducted to compare experimental groups on the unified measures of reading performance. In this analysis, the within-subjects factor was time (pretest vs. posttest) and the between-subjects factor was group (intervention vs. control). In this RM-ANOVA, homogeneity of variance and covariance was not violated, indicating that it is appropriate to report multivariate Wilk's Lambda main effects; results are reported in Table 2. Results indicated there is no significant effect of time or significant interaction between time and group, suggesting there are no differences between the intervention and control group on reading performance.

Implementation Integrity

Implementation integrity of each teacher was conducted in vivo during 36% of the total teachers' HELPS-SG sessions. The primary investigator and RAs used a procedural checklist (Appendix E) to maintain accurate records of the teachers' small-group implementation. The checklist for implementation was used to calculate core implementation integrity percentages and tips and reminders integrity percentages. The average overall implementation integrity of core procedures across all sessions for Teacher A was 96.4% (range= 82%-100%) and 98.2% (range= 82%-100%) for Teacher B. Across all core steps (i.e., focusing on the individual components of HELPS-SG), the average implementation integrity for Teacher A ranged from 80%-100%, and ranged from 90%-100% for Teacher B. The average overall implementation integrity of tips and reminders procedures across all sessions for Teacher A was 89.6% (range= 77.8%-100%) and 94.5% (range= 83.3%-100%) for Teacher B. Across all tips and reminders steps, the average implementation integrity for

Teacher A ranged from 25%-100%, and 70%-100% for Teacher B. An independent t-test was conducted to determine if the teachers' average implementation integrity across sessions significantly differed. Results indicated that the second grade teacher and third grade teacher did not significantly differ on implementation of core procedures ($t(38) = -1.13, p = .27$). However, the second grade teacher and third grade teacher significantly differed on implementation of tips and reminders procedures ($t(38) = -2.54, p = .015$).

Furthermore, in order to be considered providing the intervention as intended, the following criteria were set by the HELPS-SG manual for interventionists to achieve for each session: all core procedures must be completed at 100% integrity, and all tips and reminders procedures must be completed at 90% integrity (Begeny, Upright et al., 2013). To further understand teachers' implementation of HELPS-SG, a chi-square analysis for independence was conducted to determine if teachers significantly differed on meeting the aforementioned criteria across all sessions observed. This analysis determined that the teachers did not significantly differ on meeting criteria for the core procedures ($\chi^2(1) = 2.13, p = .14$) or the tips and reminders procedures ($\chi^2(1) = 1.91, p = .17$) across sessions.

Implementation integrity of each volunteer was conducted in vivo or with audio tape recordings during 27.9% of the volunteers' individual assessment sessions. The primary investigator used a procedural checklist (Appendix D) to maintain accurate records of the volunteers' core procedural implementation integrity, tips and reminders implementation integrity, and inter-scorer agreement of timed readings. The average core implementation integrity across volunteers was 98.4%, and the average tips and reminders implementation

integrity across volunteers was 92.1%. The average inter-scorer agreement of WCPM and WIPM across volunteers was 99% and ranged from 95%-100%.

Inter-rater Agreement

Inter-rater agreement across RAs was conducted via audio recordings for 41.7% of the all TOWRE assessments collected and for 29.7% of all the GORT assessments collected. Inter-rater agreement percentages were calculated with the following formula: number of agreements divided by number of agreements plus disagreements. For the observed TOWRE-2 assessments, RAs had a strong inter-rater agreement on both the SWE measure ($M= 99.0%$, range= 93.6%-100%) and the PDE measure ($M= 98.1%$, range= 83.8%-100%). For the observed GORT-5 assessments, RAs also had a strong inter-rater agreement on both the Fluency measure ($M= 99.2%$, range= 97.5%-100%) and the Comprehension measure ($M= 99.3%$, range= 80%-100%).

Discussion

The primary goal of this study was to evaluate the effectiveness of the HELPS-SG reading fluency program when implemented by interventionist teachers in the school setting. Furthermore, this study evaluated the effects of HELPS-SG on struggling second and third grade students. As stated above, the sole research question of this study was: will students receiving HELPS-SG outperform the control group on relevant measures of reading performance? When comparing reading performance of the participants who received HELPS-SG to the participants who did not receive HELPS-SG, no significant differences were found. This means that results did not support the hypothesis that participants who received HELPS-SG would indeed outperform the control group. These findings are

inconsistent with previous research that suggests a treatment effect would be found when using the HELPS small-group and one-on-one programs (e.g., Begeny, Braun et al., 2012; Begeny, Laugle et al., 2010; Begeny, Mitchell et al., 2011).

There are two key explanations for these unexpected findings. First, HELPS-SG may not be effective compared to what the participating school was already using curriculum-wise within the classrooms, or effective with the specific students who were chosen to be participants in this study. Although prior research would suggest a treatment effect, this study was unique to the small-group ORF intervention literature in that it had a slightly larger sample size and, unlike most other previous studies, included a control group. Second, the myriad of limitations in the study may have masked what would have otherwise been a treatment effect. The key limitations of this study will be discussed next.

Limitations and Future Directions

A multitude of limitations and threats to internal validity occurred across almost all aspects of this study. The overarching themes of all key limitations are grouped as such: sampling limitations, timeline and project adherence problems, and poor integrity with program implementation. These limitations will be discussed below, and then future directions in response to the limitations will be discussed.

Sampling limitations. First, students were not randomized to the intervention or the control group. The intervention teachers chose which of their students would receive HELPS-SG, which may have resulted in two unequal groups. Although pretest measures showed that the two groups did not differ significantly on the reading skills that were measured, other differences between groups that were not measured may have been possible

(e.g., student behavior, motivation, self-confidence). In addition, although attempts were made to measure several different reading constructs (e.g., fluency, phonemic decoding, comprehension), results of the factor analyses revealed that these assessments actually seemed to measure one unified construct—and as such, relevant reading skills (and possible differences between the groups on those skills) may not have been adequately assessed. These unmeasured group differences may have influenced the participants' responsiveness to HELPS-SG. This indicates there was a lack of control within this study over individual idiosyncrasies between those students placed in the intervention and control groups.

Another lack of control within this study is that there were other teachers within the school that were trained in HELPS-SG for a different study. Despite the primary investigator's efforts to stress the importance of the participating intervention teachers not discussing HELPS-SG procedures with other teachers, there was no way to prevent the other trained teachers from discussing or sharing methods of HELPS-SG implementation with other classroom teachers who had students in the control group. In fact, as part of school procedure and culture, control group teachers often interacted and consulted with the intervention teachers, and this may have increased the likelihood (regardless of intent) of possible contamination between instructional procedures used by HELPS-SG teachers and control teachers. Overall, there is a possibility that control group students could have received sessions or components of HELPS-SG unbeknownst to the primary investigator. If this indeed did occur, then the control group would no longer be considered a true control group that requires an "absence of treatment" (Shadish et al., 2002, p. 159), rather the control group would be considered null and void.

Second, a small sample size limited the types of analyses that could be run for this study. In order to have enough power to be considered a “powered” study (.80; Faul et al., 2007), the individual reading measures needed to be unified into composite pretest and posttest measures. Despite that the analysis run for this study was deemed appropriate (Faul et al., 2007; Tabachnick & Fidell, 2007; Thompson, 2004), it would have been more consistent with existing intervention research to have sufficient power to analyze the effects of HELPS-SG on the individual measures that are intended to assess reading fluency, reading comprehension, and word reading efficiency. Despite this study being a powered study by way of analyses used, a larger sample might have either detected differences between the groups at pretest on all or some measures, showing that the groups were different at the start from the group assignment process; or would have detected a possible treatment effect on all or some measures at posttest with a repeated measures multivariate analysis of variance (Tabachnick & Fidell, 2007).

Third, although participants were considered struggling students based on the participating school’s standards; all of the participants were not necessarily struggling in ORF based on national norms. Inclusion criteria of struggling readers is inconsistent across the literature, ranging from including students performing below the 25th percentile (Fuchs & Fuchs, 2006), below the 35th percentile (Begeny, Hawkins, et al., 2011), or below the 40th percentile (Vellutino, Scanlon, Sipay, Small, Pratt et al., 1996) in a respective area. As a conservative level of inclusion, the cut-point set to the 35th percentile will be explored with the sample used in the current study. Using the fall DORF scores obtained in the present study, the sample included 69% of students who performed at or below the 35th percentile

(89% of the intervention group and 54% of the control group). Because the purpose of this study (and the primary intent of HELPS-SG) is to improve *struggling* reader's ORF, the inclusion of many non-struggling readers (representing nearly half of the control group) may actually have decreased the likelihood of finding a treatment effect. However, because the sample size itself was relatively small to begin with, the 31% of participants who performed above the 35th percentile were not removed from the sample.

Timeline and project adherence problems. Regarding the timeline of this study, there were multiple gaps in both pretest assessment and intervention implementation due to multiple factors. First, the pretest standardized assessments (i.e., TOWRE-2 and GORT-5) were collected 2.5 months prior to the start of the intervention. In general, it is possible that students' reading skills—especially in the early grades when skill deficits are being targeted in the classroom—may change over the course of 2.5 months. Thus, by starting the intervention 8-12 weeks after the pretest assessments, it is an invalid assumption that participants' pretest reading scores were consistent with their reading performance just before intervention began.

Second, the study was significantly shorter than anticipated. Due to weather related events and the participating school's administrative decisions, the number of intervention sessions given by the participating teachers was, on average, 33% less than what was planned. Equally important, of the intervention sessions delivered, sessions were provided sporadically (e.g., 1-3 days per week), rather than the research team's request for consistent implementation (i.e., 3 days per week for the majority of the weeks). Although 13 weeks of the intervention was provided, which falls within the suggested amount of time to provide a

Tier 2 small-group intervention (Fuchs & Fuchs, 2009), it was common for the teachers and students to meet once or twice a week, rather than three times a week as recommended by the HELPS-SG manual (Begeny, 2013). Ultimately, if the participants received HELPS-SG as often as intended they would have received 39 sessions, however the participants only received an average of 26 sessions. These timeline problems may have influenced the effectiveness of the intervention and the participating teachers' implementation integrity.

Poor integrity with program implementation. As noted previously, the teachers irregularly implemented HELPS-SG with their student participants throughout this study. Ultimately, the third grade students had fewer opportunities to receive the intervention and received significantly less intervention sessions than the second graders. Bad weather days (resulting in early student dismissal from school) were the main culprit of preventing regularity in intervention delivery, and this affected third graders' intervention time more than the second graders' due to the intervention teachers' implementation schedule. Also, third graders received a week less of intervention than second graders because of school-wide testing. If third graders received as much intervention as the second graders, student performance on the posttest measures may have been impacted.

Another major limitation of implementation integrity is that the quality of instruction provided by the two participating teachers varied. Qualitative observations revealed that the two teachers had very different demeanors towards their students, which could have influenced the effects of the intervention. As shown above in the implementation integrity data results, the two teachers were comparable in regards to implementing the core HELPS-SG steps with moderate to strong integrity. However, when looking at the tips and reminders,

or the steps of HELPS-SG that focus on quality of instruction and how much the teacher engaged the students, the participating teachers varied significantly. The teacher that worked with the third graders appeared to understand the tips and reminders better and reportedly was more likely to provide praise and motivate the participants than the teacher that worked with the second graders. Because the second grade teacher's implementation integrity of tips and reminders was highly variable compared to that of the third grade teacher, second grade participants ultimately did not receive the intervention with the same quality as the third graders, which could also have limited the impact of the intervention due to poor implementation integrity. Furthermore, both teachers did not reach 90% integrity for tips and reminders steps until much later in the study than expected. This is another indication that the teachers were having a difficult time learning, remembering, and implementing HELPS-SG tips and reminders steps correctly.

Another limitation from not following what is prescribed in the HELPS-SG manual is that the teachers did not conduct the individual assessments with their students (i.e., the first 2 steps of the HELPS-SG implementation protocol). Because the teachers did not complete these steps every day they conducted HELPS-SG sessions, the teachers had no way to gauge how their students were progressing besides the data provided on the students' performance tracking charts and how the students performed in the group setting. In this particular case, veering away from the intended implementation of HELPS-SG could have influenced teachers' behavior towards students. More specifically, teachers would not have been able to provide specific praise statements nor feedback regarding students' individual progress during the small-group sessions, which is intended to happen in all HELPS sessions (Begeny,

2013). Although HELPS-SG procedures allow interventionists to recruit some assistance with the individual assessments, it is recommended that interventionists conduct the assessments with the child at least once per week. In the present study, interventionists never completed steps 1 and 2.

Lastly, HELPS-SG intervention sessions were much longer than prescribed, groups included a range in reading ability levels, and group sizes were varied. All intervention sessions ranged from approximately 15-20 minutes, however the HELPS-SG manual suggests that the sessions should only last approximately 12 minutes. Needing more time for HELPS-SG implementation could be attributed to teachers' lack of ease in using HELPS-SG or working with larger than recommended sized groups. The groups also consisted of a range of reading ability as measured by the pretest DORF. For example, one of the second grade group's DORF scores ranged from 35-81 across participants. This large variation across scores indicates that the groups may not have been homogenous in reading ability; implementing HELPS-SG with homogenous reading ability groups is a suggestion for implementation within the HELPS-SG manual. Next, limitations regarding group sizes will be discussed.

Because the intervention groups ranged from 3-6 students and not one group was the same size as another group, each group presumably received a different amount of individual attention from the teachers. Furthermore, the second graders were in groups that were too big to be manageable (i.e., groups of 5 and 6), which goes against the recommendations set forth by Gersten and colleagues (2008) and the HELPS-SG manual. However, the third graders received HELPS-SG in groups of 3 and 4, which is what is recommended for Tier 2, small-

group interventions (Gersten et al., 2008). Furthermore, Begeny, Levy, and Field (2014) recently conducted a small-group ORF intervention literature review and found that the average group size was approximately 4 students. Out of the 12 studies Begeny and colleagues (2014) evaluated in their review, 8 of the studies used groups of 3-4 participants and only 4 of the studies used groups of 5 or more during the small-group ORF interventions. This indicates that the majority of the small-group ORF intervention literature used groups of 3-4 participants, supporting the smaller group size recommendation. In sum, HELPS-SG was not implemented exactly as intended over the course of this study.

Future directions. Given the multitude of limitations discussed above, it is recommended for future research to measure the effects of HELPS-SG with a larger homogenous sample of struggling readers, who are all within the same grade, and are randomly assigned to groups of 3-4 participants. Making the sample more homogenous will decrease variability across scores and allow for a more focused understanding of HELPS-SG treatment effects. It is also recommended that HELPS-SG be implemented consistently three times a week for at least 13 weeks. These aforementioned recommendations will provide an important contribution to the extremely limited small-group ORF intervention literature base. Out of all the small group ORF intervention studies conducted thus far, only one of these studies utilized a control group (with only 5-6 students in the experimental treatment groups), and the remaining studies used sample sizes no larger than 12 (Begeny et al., 2014). It is important to conduct future studies on small-group ORF interventions because they are efficient, effective, and allow for allocation of resources (Foorman & Torgesen, 2001; Wasik, 2008). Furthermore, more randomized, control group studies are recommended to determine

if matched controls will improve as much as those who receive the researched small-group ORF intervention (Shadish, et al., 2002).

Additionally, it is recommend that future studies evaluating HELPS-SG use a classroom quality measure, such as The Early Language and Literacy Classroom Observation Tool (ELLCO; Smith, Brady, & Clark-Chiarelli, 2008), to gain a truer understanding of how teacher quality influences implementation and treatment effects of HELPS-SG. This current study relied solely upon implementation integrity checklists as a quality measure, however using other measures, such as the ELLCO, may allow for additional insight about how teachers' quality may influence HELPS-SG implementation and student outcomes.

Because the current study did not have the teachers collect the individual assessments required for HELPS-SG, despite it being prescribed by the intervention manual, it is recommended that future HELPS-SG research focus on measuring the differences of student outcomes between groups with varying degrees of teachers collecting individual assessments. For example, one experimental group could have a teacher collect all the individual assessments, another group could have a teacher collect the individual assessments one day out of the week, and the last group could have the individual assessments be collected by other adults (i.e., what happened in this current study). In other words, it is important to examine the effects of implementing HELPS-SG as intended, compared to deviations from the manual. Gaining an understanding of these comparisons has many practical implications for the future development and research of HELPS-SG.

Table 1

Reading Measures Means and Standard Deviations by Group at Pretest and Posttest

Measure	HELPS-SG		Wait-list Control	
	Pretest	Posttest	Pretest	Posttest
DORF				
<i>M</i>	56.33	82.00	62.25	79.58
<i>SD</i>	17.29	18.97	21.11	27.49
TOWRE-2 SWE				
<i>M</i>	87.28	88.39	87.21	87.04
<i>SD</i>	11.27	9.27	13.55	14.06
TOWRE-2 PDE				
<i>M</i>	83.89	80.06	85.71	82.92
<i>SD</i>	7.19	11.09	11.00	11.67
GORT-5 Fluency				
<i>M</i>	7.50	7.61	6.88	6.88
<i>SD</i>	1.51	1.38	1.87	1.87
GORT-5 Comprehension				
<i>M</i>	7.44	7.39	7.17	7.04
<i>SD</i>	1.50	1.38	1.79	1.43
Reading Performance				
<i>M</i>	.018	.096	-.013	-.071
<i>SD</i>	.798	.855	1.15	1.11

Table 2

Repeated Measures ANOVA of Group by Time on Reading Performance

<i>Main Effect</i>	<i>Wilk's Lambda</i>	<i>F</i>	<i>df</i>	<i>p</i>	η^2
Time	1.000	.016	1, 40	.900	.000
Time x Group	.981	.786	1, 40	.381	.019

Note. Group= between subject groups, wait-list control and HELPS-SG. Time= within subjects factor pre-test to posttest.

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.
- Aud, S., Hussar, W., Johnson, F., Kena, G., Roth, E., Manning, E., Wang, X., & Zhang, J. (2012). *The condition of education 2012* (NCES 2012-045). U.S. Department of Education, National Center for Education Statistics. Washington, DC. Retrieved from <http://nced.ed.gov/pubsearch>.
- Barton, P.E. (2000). *What jobs require: Literacy, education, and training, 1940-2006*. Washington, DC: Educational Testing Service.
- Begeny, J. C. (2009). Helping early literacy with practice strategies (HELPS): A one-on-one program designed to improve students' reading fluency. Raleigh, NC: The HELPS Education Fund. Retrieved from <http://helpsprogram.org>.
- Begeny, J. C. (2011). Effects of the helping early literacy with practice strategies (HELPS) reading fluency program when implemented at different frequencies. *School Psychology Review*, 40(1), 149–157.

- Begeny, J. C. (2013). Helping Early Literacy with Practice Strategies for Small Groups (HELPS-SG): A program designed to improve students' reading fluency. Raleigh, NC: The Helps Education Fund. *Manual in preparation*.
- Begeny, J. C., Braun, L. M., Lynch, H. L., Ramsay, A. C., & Wendt, J. M. (2012). Initial evidence for using the HELPS reading fluency program with small instructional groups. *School Psychology Forum: Research in Practice*, 6(3), 50–63.
- Begeny, J. C., Hawkins, A. L., Krouse, H. E., & Laugle, K. M. (2011). Altering instructional delivery options to improve intervention outcomes: Does increased instructional intensity also increase instructional effectiveness? *Psychology in the Schools*, 48(8), 769–785.
- Begeny, J. C., Krouse, H. E., Ross, S. G., & Mitchell, R. C. (2009). Increasing elementary-aged students' reading fluency with small-group interventions: A comparison of repeated reading, listening passage preview, and listening only strategies. *Journal of Behavioral Education*, 18(3), 211–228.
- Begeny, J. C., Laugle, K. M., Krouse, H. E., Lynn, A. E., Tayrose, M. P., & Stage, S. A. (2010). A control-group comparison of two reading fluency programs: The helping early literacy with practice strategies (HELPS) program and the great leaps K-2 reading program. *School Psychology Review*, 39(1), 137–155.

- Begeny, J.C., Levy, R.A., & Field, S.A. (2014). *Using small-group instruction to improve students' reading fluency: An evaluation of the current research and suggestions for future investigations*. Manuscript in preparation.
- Begeny, J. C., & Martens, B. K. (2006). Assisting low-performing readers with a group-based reading fluency intervention. *School Psychology Review, 35*(1), 91–107.
- Begeny, J. C., Mann, C. M., Cunningham, M. D., & Tsuen, H. Y. (2009). The helping early literacy with practice strategies (HELPS) curriculum: Instructional passages developed for use with the HELPS program. Raleigh, NC: The HELPS Education Fund. Retrieved from <http://helpsprogram.org>.
- Begeny, J. C., Mitchell, R. C., Whitehouse, M. H., Samuels, F. H., & Stage, S. A. (2011). Effects of the HELPS reading fluency program when implemented by classroom teachers with low-Performing second-grade students. *Learning Disabilities Research, 26*(3), 122–133.
- Begeny, J. C., Ross, S. G., Greene, D. J., Mitchell, R. C., & Whitehouse, M. H. (2012). Effects of the helping early literacy with practice strategies (HELPS) reading fluency program with Latino English language learners: A preliminary evaluation. *Journal of Behavioral Education, 21*(2), 134–149.

- Begeny, J. C., Schulte, A. C., & Johnson, K. (2012). *Enhancing instructional problem solving: An efficient system for assisting struggling learners* (pp. 3-15). New York, NY: The Guilford Press.
- Begeny, J. C., & Silber, J. M. (2006). An examination of group-based treatment packages for increasing elementary-aged students' reading fluency. *Psychology in the Schools, 43*(2), 183–195.
- Begeny, J.C., Upright, J.J., Easton, J.E., Ehrenbock, C.A., & Tunstall, K.R. (2013). Validity estimates and functionality of materials and procedures used to monitor the implementation integrity of a reading intervention. *Journal of Applied School Psychology, 29*(3), 284-304.
- Begeny, J. C., Whitehouse, M. H., Methe, S. A., Coddling, R. S., Stage, S. A., & Neupert, S. (2013). Do intervention-embedded assessment procedures successfully measure student growth in reading? *Manuscript submitted for publication.*
- Begeny, J. C., Yeager, A., & Martínez, R. S. (2012). Effects of small-group and one-on-one reading fluency interventions with second grade, low-performing Spanish readers. *Journal of Behavioral Education, 21*(1), 58–79.

- Bonfiglio, C. M., Daly, E. J., Persampieri, M., Andersen, M. (2006). An experimental analysis of the effects of reading interventions in a small group reading instruction context. *Journal of Behavioral Education, 15*, 93-109.
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32*(7), 513-531.
- Chambless, D. L., & Ollendick, T.H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52*, 685-716.
- Chapman, C., Laird, J., Ifill, N., and KewalRamani, A. (2011). Trends in High School Dropout and Completion Rates in the United States: 1972–2009 (NCES 2012-006). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubsearch>.
- Chard, D. J., Vaughn, S., & Tyler, B. J. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities, 35*, 386-406.
- Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). Fourth-grade students reading aloud: NAEP 2002 Special Study of Oral Reading. U.S.

- Department of Education, National Center for Education Statistics. Washington, DC: Government Printing Office.
- Daly, E. J., III, Chafouleas, S., & Skinner, C. H. (2005). *Interventions for reading problems: Designing and evaluating effective strategies* (pp. 75-76). New York: Guilford Press.
- Daniel S. S., Walsh A. K., Goldston D. B., Arnold E. M., Reboussin B. A., & Wood F. B. (2006). Suicidality, school dropout, and reading problems among adolescents. *Journal of Learning Disabilities, 39*, 507-514.
- Deno, S. L. (1989). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 1-17). New York, NY: Guilford Press.
- Denton, C. A. (2012). Response to intervention for reading difficulties in the primary grades: Some answers and lingering questions. *Journal of Learning Disabilities, 45*(3), 232–243.
- Dolch, E. W. (1936). A basic sight vocabulary. *The Elementary School Journal, 36*(6), 456-460.

Faul, F., Erdfelder, E., Lang, A-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175-191.

Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research & Practice*, 16(4), 203-212.

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

Fuchs, D., & Fuchs, L.S. (2009). Responsiveness to intervention: Multilevel assessment and instruction as early intervention and disability identification. *The Reading Teacher*, 63, 250-252.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical and historical analysis. *Scientific Studies of Reading*, 5(3), 239-256.

Gersten, R., Compton, D., Connor, C.M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W.D. (2008). Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades. A practice

- guide. (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://ies.ed.gov/ncee/wwc/publications/practiceguides/>.
- Good, R. H., Kaminski, R. A., Cummings, K., Dufour-Martel, C., Petersen, K., Powell-Smith, K., Stollar, S., & Wallin, J. (2011). *Dynamic indicators of basic early literacy next assessment manual*. Eugene, OR: Dynamic Measurement Group, Inc. Retrieval from <http://www.dibels.org>.
- Good, R. H., Kaminski, R. A., Dewey, E. N., Wallin, J., Powell-Smith, K.A., & Latimer, R. J. (2011). *Dynamic indicators of basic early literacy next technical manual (draft)*. Eugene, OR: Dynamic Measurement Group, Inc. Retrieval from <http://www.dibels.org>.
- Haring, N. G., & Eaton, M. D. (1978). Systematic instructional procedures: An instructional hierarchy. In N. Haring, T. Lovitt, M. Eaton, & C. Hansen (Eds.), *The fourth R: Research in the classroom* (pp. 23-40). Columbus, OH: Merrill.
- Hasbrouck, J., & Tindal, G. A. (2006). ORF norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, 59, 636–644

- Hirsch, E., & Maddock, A. (2012). *North Carolina teacher working conditions survey summary results*. Retrieved from <http://www.ncteachingconditions.org>.
- IBM Corp. (2010). IBM SPSS Statistics for Macintosh, Version 19.0. Armonk, NY: IBM Corp.
- Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press, Inc.
- Klubnik, C., & Ardoin, S. P. (2010). Examining immediate and maintenance effects of a reading intervention package on generalization materials: Individual verses group implementation. *Journal of Behavioral Education, 19*(1), 7-29.
- Kuhn, M. R. (2005). A comparative study of small group fluency instruction. *Reading Psychology, 26*(2), 127-146.
- Lou, Y., Abrami, P.C., Spence, J.C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research, 66*(4), 423-458.
- Metropolitan Life Insurance Company. (2012). *The MetLife survey of the American teacher: Teachers, parents and the economy*. New York, NY: Peanuts Worldwide LLC.

Retrieved from

<https://www.metlife.com/assets/cao/contributions/foundation/american-teacher/MetLife-Teacher-Survey-2011.pdf>

Mercer, C. D., & Campbell, K. U. (1998). Great Leaps reading program, Kindergarten–2nd grade. Gainesville, FL: Diarmuid.

Mercer, C. D., Campbell, K. U., Miller, M. D., Mercer, K. D., & Lane, H. B. (2000). Effects of a reading fluency intervention for middle schoolers with specific learning disabilities. *Learning Disabilities Research & Practice, 15*, 179–189.

Morgan, P. L., & Sideridis, G. D. (2006). Contrasting the effectiveness of fluency interventions for students with or at risk for learning disabilities: A multilevel random coefficient modeling meta-analysis. *Learning Disabilities Research & Practice, 21*(4), 191-210.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Authors.

National Center for Education Statistics (2011). The nation's report card: Reading 2011.

National assessment of educational progress at grades 4 and 8, NCES (2012-457).

Institute of Education Sciences, U.S. Department of Education; Washington, D.C.

National Institute of Child Health and Human Development. (2000). *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* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

Odum, M. (2011). *Factor scores, structure and communality coefficients: A primer*. Paper presented at the Annual Meeting of the Southwest Educational Research Association. San Antonio, TX.

Oliff, P., Mai, C., & Leachman, M. (2012). *New school year brings more cuts in state funding for schools* (pp. 1-14). Washington, D.C.

Reeves, J. (2010). Teacher learning by script. *Language Teaching Research, 14*(3), 241-258.

Ross, S. G., & Begeny, J. C. (2011). Improving Latino, English language learners' reading fluency: The effects of small-group and one-on-one intervention. *Psychology in the Schools, 48*(6), 604-618.

- Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods*, 1, 331–340.
- Schmuckler, M. A. (1997). What is ecological validity? A dimensional analysis. *Infancy*, 2(4), 419-436.
- Schoenwald, S. K., Carter, R. E., Chapman, J. E., & Sheidow, A. J. (2008). Therapist adherence and organizational effects on change in youth behavior problems one year after multisystemic therapy. *Administration and Policy in Mental Health*, 35, 379-394.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Simmons, D. C., Coyne, M. D., Hagan-Burke, S., Kwok, O., Simmons, L., Johnson, C., Zou, Y., et al. (2011). Effects of supplemental reading interventions in authentic contexts: A comparison of kindergarteners' response. *Exceptional Children*, 77(2), 207-228.
- Smith, M., Brady, J., & Clark-Chiarelli, N. (2008). *Early Language and Literacy Classroom Observation Tool, K-3* (research ed.). Baltimore, MD: Brookes Publishing.

- Spache, G. (1953). A new readability formula for primary-grade reading materials. *The Elementary School Journal*, 53(7), 410-413.
- Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson.
- Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education*, 25(4), 252-261.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, D.C.: American Psychological Association.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of word reading efficiency*, (2nd ed.). Austin, TX: Pro-Ed.
- Vaughn, S., Linan-Thompson, S., Kouzekanani, K., Pedrotty Bryant, D., Dickson, S., & Blozis, S. A. (2003). Reading instruction grouping for students with reading difficulties. *Remedial and Special Education*, 24(5), 301-315.
- Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and

- experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*, 88, 601-638.
- Wasik, B. (2008). When fewer is more: Small groups in early childhood classrooms. *Early Childhood Education Journal*, 35, 515-521.
- Wiederholt, J. L., & Bryant, B. R. (2012). *Gray oral reading test*, (5th ed.). Austin, TX: Pro-Ed.
- Wireless Generation (2007). mCLASS: Reading 3D (2nd edition) [materials and software]. Brooklyn, NY: Wireless Generation, Inc.

APPENDICES

Appendix A

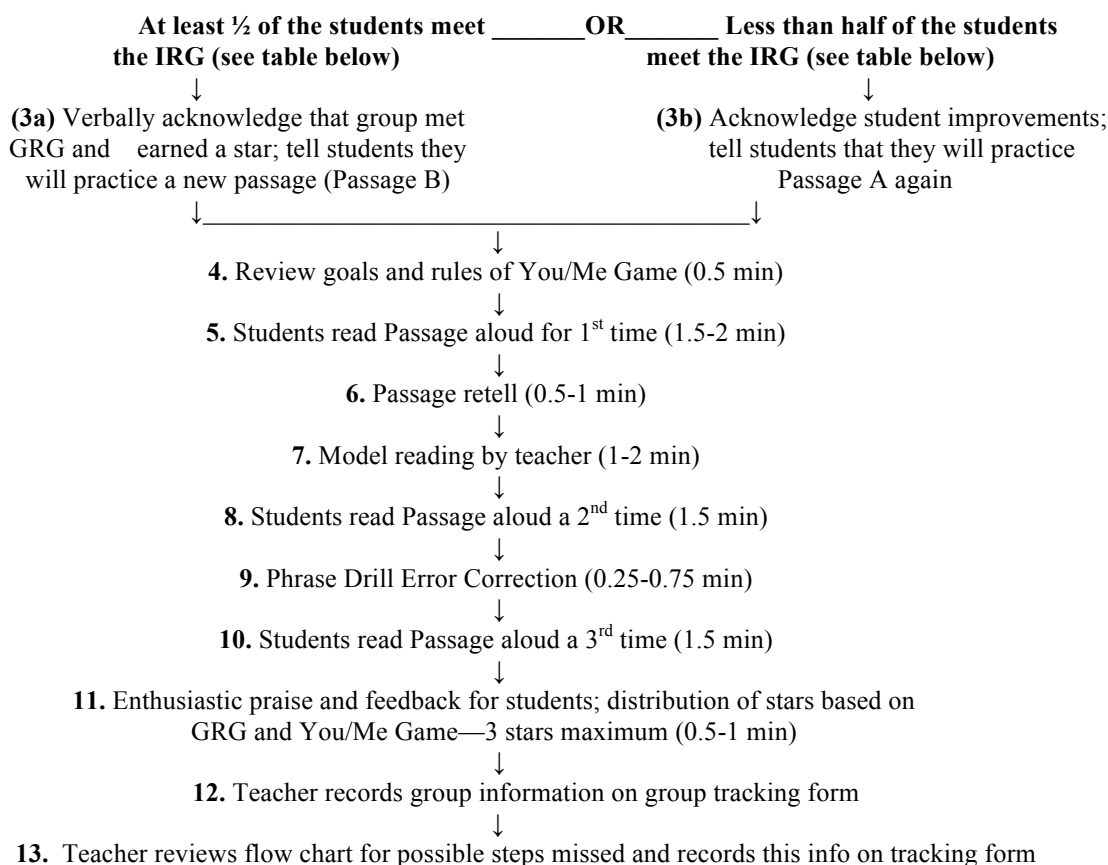
HELPS Program for Small Groups Implementation Flow Chart

1. One-on-one assessment with each student in group, facilitated by any trained educator

- a. ORF assessment with Passage A
- b. ORF assessment with Passage B (as needed)
- c. Record all data on Graph and offer praise and feedback about student's retention score
- d. Record all information on Tracking Form

2. Steps 1a – 1d are completed with each remaining student from the group.

3. Make decision about students meeting the Group Reading Goal (GRG)...



Individual Reading Goals (IRGs) According to Each Student's Grade Level

	WCPM with Passage A	WIPM with Passage A
First Grade	80 or more	3 or less
Second Grade	100 or more	3 or less
Third Grade	120 or more	3 or less
Fourth Grade	135 or more	3 or less

Appendix B

Scripted Directions

(For one-on-one assessments) Directions to administer before student reads passage:

1. Place the teacher copy of the reading passage in front of you but shielded so the student cannot see what you record. Present the student copy of the reading passage to the student.
2. Say to the student, **“Here is a story that I would like you to read. When I say ‘Begin’, start reading aloud at the top of the page and read across the page. Try to read each word. If you come to a word you don’t know, I’ll tell it to you. Do you have any questions? Be sure to read as quickly as you can without making errors, and try to read with good expression.”**
3. Say, **“Begin!”** and start the stopwatch when the student says the first word.
4. If the student hesitates on a word for more than 3 seconds, say the word.
5. At the end of one minute, place a closed bracket after the last word.
6. *If the student reads so fast that no expression is given, remind the student that when he/she reads the next story, you want him/her to read at a comfortable rate (i.e., with good expression).*
7. Remove both copies of the reading passage and record student’s score on the Individual Progress Tracking Form.

Introductory statements and expectations (includes Verbal Cuing Procedure):

As a group, you’re going to be doing some reading with me today. As you read, I want you to do your best reading. This means I want you to read as quickly as you can without making mistakes, and try to read with good expression (like I do when I read to you). I also want you to remember what happens in the story and try to remember the difficult words that we practice.

Describing the You/Me Game:

As we read, we will play the You/Me Game. To earn points for your team and win the game, you need to work hard and follow directions throughout today’s lesson. For example, when I am reading aloud or another student is reading aloud, you need to read along silently to yourself. If I call on you and you know where we are in the story, your group will earn one point. If you do not know where we are in the story, I will earn one point. At the end of the lesson, if your team earns XX [insert # based on recommendations for the Teacher’s Manual] more points than I do, your group will earn two stars on your Group Star Chart.

Directions to administer before students read passage aloud (Group Reading Procedure):

1. Say to the group, **“Now we are going to practice reading today’s story out loud. (Insert student’s name) will start reading when I say *begin*, and then I will call on**

another student to read. Each of you will take turns reading a few sentences at a time. When you are not reading aloud, you need to read along silently while your classmate reads. If I call on you and you know where we are in the story, your group will earn one point in the You/Me Game. If you do not know where we are, I will earn one point.”

2. Say “**Begin**” and have the designated student start reading aloud. After this student reads 1 to 3 sentences, randomly select another student in the group to read the next 1 to 3 sentences.
3. Continue this procedure of randomly selecting students to read 1 to 3 sentences until the group reads for approximately two minutes or until the students read the entire passage.
4. When calling on a new student to read, if that student immediately starts reading where the previous student left off, record one point for “You” (the students) in the You/Me Game. If the student cannot start where the previous student left off, record a point for “Me” (the teacher).
5. As students read, record students’ reading errors on the teacher copy of the passage. Record errors with slashes, underlines, or circles, as instructed for Reading 1, 2, or 3.
6. At the end of the activity, briefly praise the group for their effort (as applicable).

Directions for administering Retell procedure:

1. Say to the group, “**Now I would like each of you to tell me one important part that you remember from the story. Try to tell me what happened in the correct order.**”
 - For expository text, you should say, “**Now I would like each of you to tell me one important thing that you learned from the reading. Starting with the first important thing you learned, tell me about what you learned in the correct order.**”
2. Randomly select one student to begin the Retell procedure. When that student finishes a brief (approximately 10 seconds) retell, randomly select another student to retell a portion of the passage.
3. Continue this procedure until all students have been called on once. If necessary, give prompts to help students remember key parts of the passage and to correctly sequence the events.
4. At the end of the Retell procedure, briefly praise the group for their effort (as applicable).

Directions for teacher to read passage aloud (Modeling Procedure):

1. Say to the group, “**Now I am going to read today’s story to you. Please follow along with your finger, reading the words to yourself as I read them. Sometimes I will stop reading to make sure you are following along. If I call your name, you need to tell me the next word in the story. If you read the correct word, this will show me you are reading along with me and your group will earn 1 point in the You/Me Game. If you do not show me you are reading along with me, I will earn one point.**”

2. Read the passage at a comfortable reading rate and with good expression for approximately 1.5 minutes or until you read the entire passage. Make sure the students are following along with their fingers and prompt students to do this, if necessary.
3. While reading the passage, stop 5-7 times in order to randomly call on a student to read the word that immediately follows the word you stopped at. Record a point for the group (a “You” point in the You/Me Game) when a student reads the correct word. Record a point for the teacher (a “Me” point) when a student cannot immediately read the next word.
7. At the end of the activity, briefly praise the group for their effort (as applicable).

Directions for administering Phrase-drill Error-correction Procedure:

1. Say to the students, “**Now we are going to practice some of the difficult words from the story.**” Have the students sit closely enough so they can all see the one passage you will point to.
2. On a student’s copy of the passage, point to the first error word from Group Reading 2, say the word, and then say, “**All together, read this after I do,** [read the 2-5 word phrase containing the error word and then have the students chorally read the phrase]. **Again** [students should chorally read the phrase a second time], **Again** [students should chorally read the phrase a second time].” This procedure should allow the students to chorally read the phrase three times.
 - a. To better ensure all students read the phrase at the exact same time, use a prompt (e.g., snap your finger, tap the back-end of a marker on the table) to signal for students to begin reading the phrase aloud.
3. Make sure students read the phrases rather than simply memorize them and repeat them. (Teachers want students to read, rather than recite). Also be sure that all students read aloud together.
4. Repeat the above procedure for all unique error words from Group Reading 2. If time permits (i.e., very few words were read incorrectly in Group Reading 2), complete the Phrase-drill procedure with words read incorrectly in Group Reading 1.
 - a. If students make 1 or fewer errors total on Readings 1 and 2, practice 1-3 words or phrases the students read less fluently. Use the procedures above.
5. After all phrases are practiced, assign 0-5 points for the You/Me Game, based on collective student effort. Praise students’ effort accordingly.

Appendix C

Implementation Protocol

Steps below that are denoted with an asterisk should be implemented in combination with the “HELPS Program for Small Groups: Scripted Directions.”

One-on-One Assessment Procedure:

1. One student at a time is asked to read passages aloud to the teacher (*1.5 to 2.5 minutes per student*).
 - a. *The student reads the passage practiced in the most previous session (Passage A).
 - b. *The student reads the next passage in the sequenced curriculum (Passage B).
Note: Step 1b is skipped if (a) the student read Passage B in the most previous one-on-one assessment session; and (b) the group used Passage A (not Passage B) during the most previous session, which occurs when the group fails to meet the Group Reading Goal in the previous session.
 - c. Teacher records the student’s WCPM and WIPM (for each passage) on that student’s Graph and Individual Progress Tracking Form.
 - d. Teacher provides specific feedback and enthusiastic praise for reading improvements.
2. Teacher implements Steps 1a – 1d with each remaining student from the group.

Group-based Instructional Strategies:

3. Teacher determines whether at least half of the students met the Individual Reading Goal (IRG). The IRG is determined by each student’s grade level, as shown in the Table below.

Individual Reading Goals (IRGs) According to Each Student’s Grade Level

	WCPM with Passage A	WIPM with Passage A
First Grade	80 or more	3 or less
Second Grade	100 or more	3 or less
Third Grade	120 or more	3 or less
Fourth Grade	135 or more	3 or less

The IRGs above are suggested Goals for students at the given grade level. As described in the Teacher’s Manual, it is sometimes appropriate for the teacher to modify a student’s IRG if necessary.

Group Reading Goal (GRG): The Criterion and Instructional Decision Rules

The GRG criterion is met when at least half of the group’s students meet his/her IRG. (*0.25 minutes*)

- a. If the GRG criterion is met, teacher praises the students and tells them they will (a) earn 1 star on the Group Star Chart, and (b) practice a new story during that session (Passage B).

- b. If a student is called on, he/she must immediately say the next word in the passage. If the student does this, the team earns a point in the You/Me Game. If the student does not, the teacher earns a point in the You/Me Game.
8. Group Reading 2: Students read the session's passage aloud a second time (*1.5 minutes*)
 - a. Teacher implements Steps 5a – 5f.
9. *Teacher implements phrase-drill error correction on all incorrectly read words from Group Reading 2 (*0.25 to 0.75 minutes*)
 - a. All students chorally respond to phrases the teacher asks them to practice.
 - b. After all phrases are practiced, the teacher assigns 0 – 5 points for the You/Me Game based on collective student effort.
 - c. If time permits (i.e., very few words were read incorrectly in Group Reading 2), complete the phrase-drill procedure with words read incorrectly in Group Reading 1 and/or practice otherwise difficult words or phrases from the passage.
10. Group Reading 3: Students read the session's passage aloud a third time (*1.5 minutes*)
 - a. Teacher implements Steps 5a – 5f.
11. Teacher provides enthusiastic praise and feedback on the students' reading and praises effort as applicable (*0.5 to 1 minute*)
 - a. If students met the GRG criterion, teacher adds 1 star to the Group Star Chart and states the reason the students earned that star.
 - b. Based on the outcome of the You/Me Game, teacher tells students they earned an additional 1 or 2 stars on the Group Star Chart.
 - c. Students select from bonus bag and/or prize box, as applicable.
12. Teacher records group information on the Group Progress Tracking Form.
13. Teacher reviews implementation steps from flow chart and records any steps missed on the Group Progress Tracking Form.

Total implementation time: 7.5 to 10.5 minutes (excluding one-on-one assessment procedure)

Appendix D

Checklist for Individual Assessment Implementation

Tape Recorder: _____

Volunteer observed: _____ Observer: _____ Group: _____ Date: _____

Implementation of Individual Assessment

List of Steps

Student	1a	1b	1c	1d
1				
2				
3				
4				
5				
6				

Vol. Errors counted / Total words read / Obs errors counted

S1, A: ___/___/___ ISRA%:___ B: ___/___/___ ISRA%:___
 S2, A: ___/___/___ ISRA%:___ B: ___/___/___ ISRA%:___
 S3, A: ___/___/___ ISRA%:___ B: ___/___/___ ISRA%:___
 S4, A: ___/___/___ ISRA%:___ B: ___/___/___ ISRA%:___
 S5, A: ___/___/___ ISRA%:___ B: ___/___/___ ISRA%:___
 S6, A: ___/___/___ ISRA%:___ B: ___/___/___ ISRA%:___

- 1) Steps 1a-1d above were completed correctly with at least one student
- 2) Steps 1a-1d above were completed correctly with ALL students in the group

Timed Reading, Feedback, and Graphing Procedures during Individual Assessments

- After each student’s oral reading, teacher always correctly indicated on the examiner passage (with a bracket) the number of words read in one minute.
- Teacher correctly scored and calculated each student’s total WCPM and WIPM per passage read.
- While graphing, teacher gave verbal feedback and praise regarding the student’s WCPM and WIPM scores.
- Teacher connected lines between WCPM (and WIPM) scores *only* for scores of the same passage.
- Teacher circled the data point and session number when the student began a new passage. **(Applicable?)**
- Teacher completed each student’s Individual Progress Tracking Form with 100% accuracy.

Appendix E

Checklist for Small-group Implementation

Teacher observed: _____ Observer: _____
 Group: _____ Date: _____

Implementation of Individual Assessment

(Place “✓” in the box for completed steps; Place “X” in the box for skipped steps or those implemented with major errors; use arrows “→” to specify if and how a step was implemented out of order).

3. Teacher made correct decision about students meeting the Group Reading Goal

Implementation of Core Group Procedures

List of Core Group Steps: 4 ; 5 ; 6 ; 7 ; 8 ; 9 ; 10 ; 11 ; 12 ; 13

Percentage of steps completed = total steps completed / total steps possible:

_____/11 x 100: _____%

Implementation of Tips and Reminders (steps missed below are considered *minor errors*)

(Place “✓” in the box for completed steps, Place circle “O” around the box for non-applicable [NA] steps)

General Implementation Procedure

- Teacher had the following materials available and organized before starting the session: stop watch, examiner passage, student passage, dry-erase marker, pencil, student graph, Progress Tracking Form (for the group and each individual), Star Chart, Bonus Bag, Implementation Flow Chart, and Scripted Directions. Also, the prize box was reasonably accessible.
- Teacher used Scripted Directions or Abbreviated Directions as advised at top of Abbreviated Directions.
- Teacher specifically told students they met or did not meet the Group Reading Goal before starting Step 5.

Group-based Repeated Readings

- Teacher randomly selected students in the group to read (i.e., no consistent order was used to select students to read).
- Teacher selected a new student to read only when the prior student completed a sentence.
- Each student read 1-3 sentences when it was his or her turn to read.
- Teacher referred to points that the students earned as part of the You/Me Game.

Retell Check Procedure

- Before prompting student to begin the Retell Check, teacher made sure students could not review the passage during the Retell Check.
- Teacher provided the correct directions for Retell Check depending on whether the students read a narrative or expository passage.
- Teacher called on each student in the group to provide a response.

- Teacher used broad follow-up questions to solicit a student's retell *only* if the student was unable to retell anything for approximately 10 seconds. **(Applicable?)**

Modeling Procedure

- Teacher read aloud at a pace just a little faster than the group's reading ability.
- Teacher read with good expression.
- Teacher read at a volume the group could clearly hear.
- Teacher paused 2-4 times per student in order for the student to read the next word in the passage.
- Teacher randomly selected which student would read the next word in the passage (i.e., no consistent order was used to select students to read).
- Teacher referred to points that the students earned as part of the You/Me Game.

Phrase-Drill Error Correction Procedure

- Teacher always stated the word that was read incorrectly before having the group read the phrase that contained the word.
- Teacher asked students to practice "logical" phrases.
- Teacher told students to "READ" the phrases, and did not ask them to "SAY" or "REPEAT" phrases.
- Teacher had students practice all incorrectly read words from group reading #2 (Step 8), up to 5 words or until time permitted)
- Teacher pointed (or had the students point) to each word practiced.
- Teacher ensured that all students read the phrases at the same time and (if needed) used a prompt to (e.g., snap of finger, tap of marker on the table) to signal for students to begin reading the phrase aloud.
- If students made 1 or fewer errors, the teacher told the student(s) to practice 1-3 words or phrases that were read less fluently or were reading incorrectly in a previous student reading. **(Applicable?)**
- If students practiced words that were read *correctly* but less fluently (see above step), teacher explained to the student(s) that they read the words correctly, but will practice them because they are difficult. **(Applicable?)**
- Teacher referred to points that the students earned as part of the You/Me Game.

Motivational (Reward) Procedure and Data Recording

- Throughout the session, teacher provided a minimum of three different praise statements regarding group's reading behavior.
- When awarding stars on the Group Star Chart, teacher accurately told students why they earned each star (Group Reading Goal and/or You/Me Game, up to 3 stars total)
- With enthusiasm, teacher praised *specific* reading behaviors (e.g., nice job reading *accurately* and with *good expression*; I like how you *corrected words you missed previously*) and praised group for specific reading behaviors or improvements at the end of the session.
- If the group landed on OR passed a shaded square on Group Star Chart, they were allowed to select a ticket from the bonus bag and teacher correctly recorded the bonus stars written on the ticket. **(Applicable?)**
- Teacher conveyed that improved reading skills, rather than the opportunity to earn stars/prizes, is the primary reason the group should put forth effort during each HELPS session. **(Applicable?)**

Teacher completed the Group Progress Tracking form with 100% accuracy.

Total steps applicable = 32 total check boxes – number of boxes circled as NA ____ = ____

Total steps completed = number of boxes with a check mark = ____

**Percentage of items completed = total items completed / total items applicable ____ / ____ x
100**

Percentage of items completed: ____%

Time of Session (TOS): _____

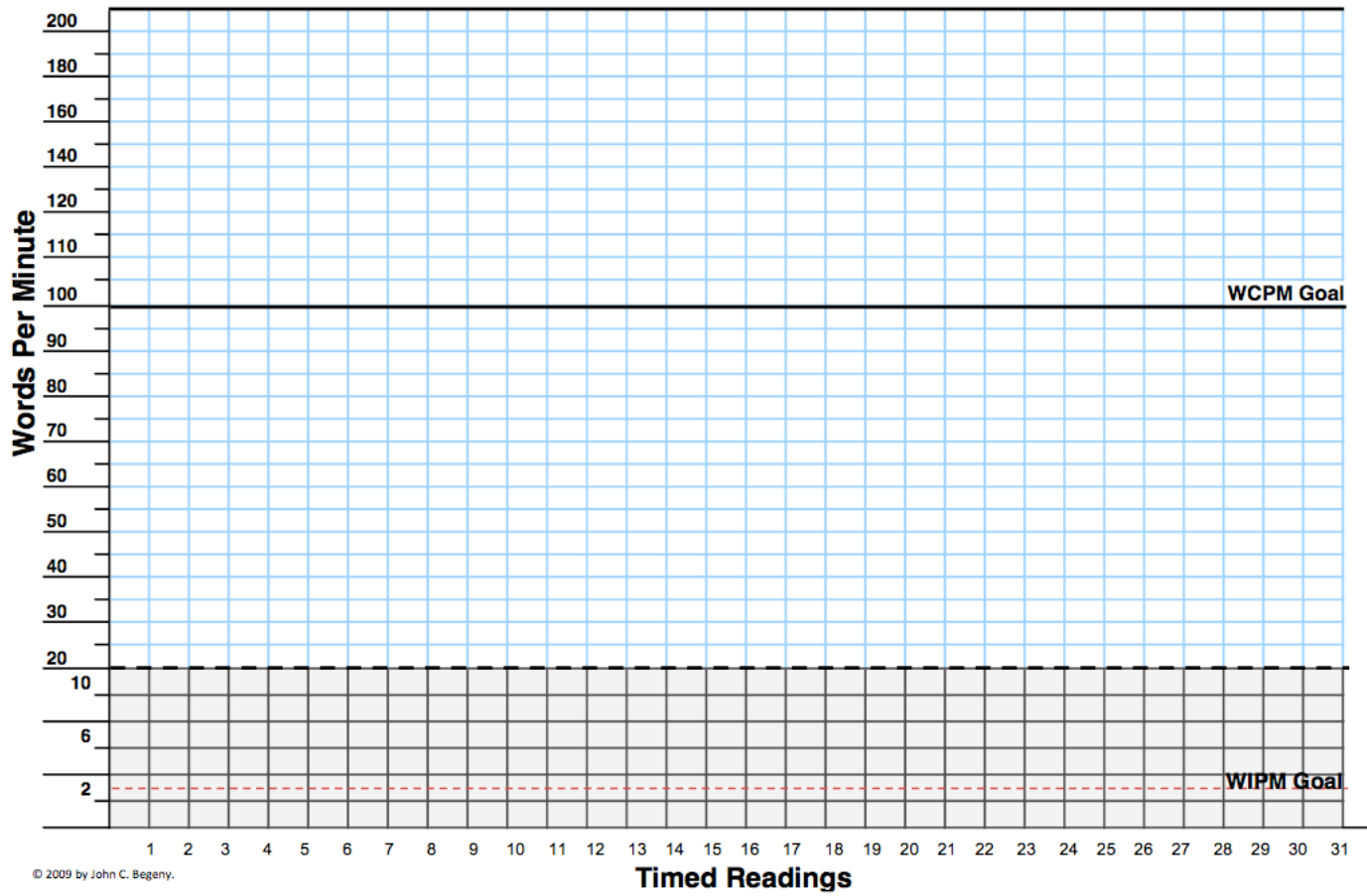
Appendix F

Individual Progress Graph

NAME: _____

HELPS Program: Student Graph - Grade 2

Notes: Circle Timed Reading # and first WCPM score each time student begins a new passage. Write date below first Timed Reading of the day.
(Mark WCPM with dot; Mark WIPM with X) Only connect dots and Xs for readings of the same passage.



Appendix H
Group Star Chart

Student Names: _____

			■						■					
		■				■								
				■								■		
							■							
		■							■					
		■				■								
				■								■		
							■							
			■						■					
		■				■								
	■							■						

Star Chart Rules:

- When the group earns a star in the last square of each row (a total of 15 stars), each student earns one prize from the special prize box.
- When the group earns a star in a shaded square, one student gets to select a ticket from the Bonus Bag. The number of stars written on the selected ticket should be immediately added to the Group Star Chart.

Appendix J

HELPS Placement Assessment

HELPS Program Placement Assessment Recording Form and Decision-Making Tables

Student's Name: _____

Student's Grade: _____

Scores from Placement Assessment

	Level 1		Level 2		Level 3		Level 4		Level 5	
	WCPM	WIPM	WCPM	WIPM	WCPM	WIPM	WCPM	WIPM	WCPM	WIPM
Assessment Passage 1	4	4	23	23	45	45	64	64	79	79
Assessment Passage 2	8	8	27	27	52	52	67	67	80	80
Average Level Score										

Based on the above scores and the tables below, this student will begin the HELPS Program at passage number: _____

Starting Point Criteria Table

Student's Grade	Target WCPM	Target WIPM
First	50-60 WCPM	5 or less
Second	70-80 WCPM	3 or less
Third	90-100 WCPM	3 or less
Fourth	105-115 WCPM	3 or less

Decision Table for Placement Assessment

Passages Administered for:	Student's score is <u>below</u> target criteria	Student's score is <u>within</u> the target criteria	Student's score <u>exceeds</u> the target criteria
Level 1 (passages 4 & 8)	Start student at passage 1	Start student at passage 5	Administer Level 2 passages
Level 2 (passages 23 & 27)	Start student at passage 5	Start student at passage 25	Administer Level 3 passages
Level 3 (passages 45 & 52)	Start student at passage 25	Start student at passage 50	Administer Level 4 passages
Level 4 (passages 64 & 67)	Start student at passage 50	Start student at passage 65	Administer Level 5 passages
Level 5 (passages 79 & 80)	Start student at passage 65	Start student at passage 75	Start student at passage 75*

* If a student exceeds the target criteria with Level 5 passages, the instructor may choose to re-evaluate whether the student is likely to benefit from the HELPS Program. In some cases the student may still benefit from the program, but in other cases the student may benefit from a reading program that specifically targets a skill other than reading fluency.