A COMPARISON OF TWO SIGHT WORD INTERVENTIONS:

TRADITIONAL DRILL AND WORKSHEETS

A dissertation presented by

Christina Marie Mulé, C.A.G.S., NCSP

Submitted to

The Department of Counseling and Applied Educational Psychology

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in the field of

School Psychology

Northeastern University
Boston, Massachusetts
March 2013

Dissertation Committee:

Robert J. Volpe, Ph.D., Advisor

Amy M. Briesch, Ph.D.

Laurice M. Joseph, Ph.D.
Dedication

Dedicated to my parents, William and Lisa Sullivan, whose love, encouragement, and support allow me to reach my dreams.
Acknowledgements

I would like to thank my advisor, Dr. Robert Volpe, for his constant encouragement, his dedication and investment in my personal and professional growth, and most of all his belief in me. I am grateful to have him as my mentor, someone who was there when I could not direct the winds, but still taught me how to sail. I am thankful to have been a part of his research lab and only hope to be half as great of a teacher as he is someday. To my committee member, Dr. Amy Briesch, I will never be able to thank her enough for her mentorship throughout the years and her constant calming nature. She is one of the brightest people with whom I have had the pleasure to train under, yet is by far one of the humblest. These characteristics are ones I hope to emulate in my future career and are ones that I will never forget. I wish to thank my committee member, Dr. Laurice Joseph, whose work was the impetus of my interest in my dissertation topic. I have greatly respected her work throughout the years and feel honored to have had her mentorship and collaboration on this dissertation research. I wish to thank faculty members at Northeastern University, University of Wisconsin-Madison, Tufts Medical Center, and the May Institute who have supported and encouraged me throughout the years: Drs. Melanie DuBard, Jessica Hoffman, Louis Kruger, Laurel Leslie, Chieh Li, Karin Lifter, Takuya Minami, and Emanuel Mason. These faculty members, while not a part of my dissertation committee, have nurtured me to this point and have enriched my knowledge as a researcher and practitioner.

My sincerest appreciation goes to the graduate students of the Spring 2012 Academic Interventions course who were integral in the data collection of this dissertation work. A very special thanks to the fieldwork advisors, administrators, and teachers at their training sites that allowed them to carry out this work. Most of all, thank you to the elementary students who participated in this study. I wish them all the best in their futures! A special thank you to my
peer, colleague, and friend, Matthew DuBois, who participated in this study by helping me to review published studies and conduct reliability calculations.

Finally, I could not have completed this dissertation work without the unwavering support and encouragement from my friends and family. I would like to thank my soccer team and my fellow yogis who remind me to keep balance in my life. To my best friends, Beth Ann Barnosky, Drew Brais, Brittany Lyman, Jay Morosey, Gregory Penoyer, Jody Podpora, Pauline Shaffer, Craig Strohmeier, Emily Wilson, and Amanda Zangrillo, thank you for your friendship during this year of my dissertation work and always. You are the best friends anyone could possibly ask for. Thank you to my mom and dad, William and Lisa Sullivan, for always pushing me to be better, both personally and professionally. But most of all, thank you for your love throughout my years in graduate school and always, without it I could not have come this far. Thank you to my siblings, April Mulé and Ryan Sullivan, for encouraging me to pursue my dreams. To my in-laws, John and Karen Kitchura, thank you for your encouragement and support in this year and always. Last but not least, thank you to my husband, John Michael Kitchura, Jr., for believing in me and supporting me in everything I do. I owe you so much more than a "thanks."
VITA

2006.............................................................B.S. Psychology, Rochester Institute of Technology

2007............................................................. M.S. Psychology, Northeastern University

2009............................................................ C.A.G.S. School Psychology, Northeastern University

PUBLICATIONS


Building your capacity: Advancing research through community engagement. Leslie, L. K., Leung-Rubin, C., Dawson, E. H., Mulé, C. M., Mbawuike, V., Rios, M., & Daudelin, D. (2011). Curriculum guide on community-based participatory research. Fifteen modules were developed to teach community organizations about research and how best to collaborate with research institutions so that both parties needs are met. CTSI Grant. P.I. Laurel K. Leslie, MD, MPH.

Examination of the Rogers process for youth in the custody of the Massachusetts Department of Children and Families submitted to the Massachusetts Court Improvement Program. Leslie, L. K., Mackie, T. I., Mulé, C. M., Wade, R., Leung, C., Dawson, E. H., . . . Hyde, J. (2011). Final Report on a qualitative examination of the Rogers process in Massachusetts (i.e., process by which antipsychotic medications are approved for children in child welfare custody). One-hundred and nine interviews were conducted with stakeholder group members over the span of three months. Massachusetts Court Improvement Grant. P.I. Laurel K. Leslie, MD, MPH.


RAP Grant Progress Report submitted to the Substance Abuse and Mental Health Administration. Amaro, H (PI), Matsumoto, A., Dai, J., Garcia, J., Mulé, C., & Ubiem, C. (2008). Final Report on Reentry Assistance Program (RAP). Role on Project: Graduate Fellow Assistant. Served as a member of the evaluation team for the grant on a wrap around services project developed for incarcerated men with history of substance abuse that incorporated case management based services. Each year, 100 men were recruited at the Suffolk County Correctional Facility and followed up at 6- and 12-month points after release. Center for Substance Abuse Treatment (SAMHSA) Grant. P.I. Hortensia Amaro, Ph.D.

**FIELDS OF STUDY**

Major Field: School Psychology
# Table of Contents

Dedication...................................................................................................................................ii
Acknowledgements.......................................................................................................................iii
Vita..................................................................................................................................................v

**Article 1: Literature Review**

Title Page ..................................................................................................................................1
Abstract ......................................................................................................................................2
Literature Review .......................................................................................................................3
Methods.....................................................................................................................................11
Results.................................................................................................................................14
Discussion..............................................................................................................................27
References..............................................................................................................................35
Appendix.................................................................................................................................42

**Article 2: Empirical Research Proposal**

Title Page ..................................................................................................................................49
Abstract ......................................................................................................................................50
Literature Review .....................................................................................................................51
Methods.....................................................................................................................................56
Results.......................................................................................................................................64
Discussion...............................................................................................................................68
References...............................................................................................................................81
Appendix.....................................................................................................................................87

**Appendices:**

- Appendix A: High Frequency Words......................................................................................93
- Appendix B: TDP Intervention Script .....................................................................................97
- Appendix C: WS Intervention Script.....................................................................................100
- Appendix D: TDP Procedural Integrity Checklist.....................................................................103
- Appendix E: WS Procedural Integrity Checklist.....................................................................105
- Appendix F: Example Retention Assessment Sheet..............................................................107
- Appendix G: Example Maintenance Assessment Sheet.........................................................109
- Appendix H: Example Generalization Assessment Sheets......................................................111
Appendix I: Retention Assessment Procedural Integrity Checklist............................114
Appendix J: Maintenance Assessment Procedural Integrity Checklist........................116
Appendix K: Generalization Assessment Procedural Integrity Checklist.....................118
Literature Review:

An Evaluative Review of Flashcard Drill Interventions Targeting Sight Word Recognition

Christina Marie Mulé, C.A.G.S., NCSP

Northeastern University
Abstract

The purpose of this review is to synthesize the research on the effectiveness of sight word interventions, identify limitations of this research, offer suggestions for future research, and discuss implications for practice. Electronic databases (i.e., ERIC, PsycArticles, and PsycINFO) were searched for articles published from 2001 to 2011 on flashcard drill methods targeting sight word knowledge. Sixteen articles met the selection criteria and were subsequently reviewed with particular attention to intervention components, and treatment outcomes with regard to instructional-effectiveness and -efficiency. This review of the literature identified incremental rehearsal (IR), traditional drill and practice (TDP), interspersal training (IST), direct instruction (DI), and constant time delay (CTD) as the most investigated sight word interventions. However, effectiveness and efficiency data were somewhat inconsistent across studies. Although strong conclusions could not be drawn as to which flashcard drill method is superior with regard to effectiveness, with regard to efficiency, findings supporting the superiority of TDP over other methods was more consistent. In light of these findings, recommendations for practice and for future research are discussed.

Keywords: Reading intervention, Sight word intervention, Sight word learning, Traditional Drill, Incremental Rehearsal, Interspersal Rehearsal, and Drill Sandwich
Literature Review:

An Evaluative Review of Flashcard Drill Interventions Targeting Sight Word Recognition

The ultimate goal of reading is to progress from learning to read, to reading to learn (Chall, 1983; Cunningham & Stanovich, 1998; Joseph, 2006). However, before a student is able to use reading as an instrument to obtain information, there are several components, or foundational skills, that need to be mastered first (e.g., concept about print, phonological awareness, orthographic knowledge, alphabetic principle, and phonological recoding and decoding; Joseph, 2006). Once these foundational skills are acquired, students are expected to read words with automaticity or read words "on sight." This skill is referred to as sight word recognition, which is often used synonymously with word reading and word identification.

Sight word recognition is the ability to accurately read words without analysis or hesitation (Joseph, 2006; Lerner, 2003; Samuels, 1997). This rapid identification of words is a key development in reading as it allows students to increase their reading rate, as well as comprehension (i.e. what meaning they are able derive from text; Burns, Dean, & Foley, 2004; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Samuels, 1997; Szadokierski & Burns, 2008). Skilled readers are able to read text fluidly, devoting more attentional and processing resources to comprehension rather than decoding (Eveleigh, 2010; Samuels, 1997; Szadokierski & Burns, 2008). Conversely, unskilled readers exert more effort to decode single words; leaving fewer cognitive resources to retain the meaning of the words they have read (Eveleigh, 2010; Samuels, 1997; Szadokierski & Burns, 2008). In view of this contrast, it is critical that students increase the pool of words that they can read automatically, so that they are able to reach the ultimate goal of reading to learn (Joseph, 2006; Samuels, 1997).
Given the importance of sight word recognition, various methods have been developed to target this skill. For example, sight words have been taught with community-based instruction, which presents naturally occurring words (e.g., signs in the community) to students and eliminates the need to program in for the generalization of words learned in an artificial setting (Cuvo & Klatt, 1992; Mosley, Flynt, & Morton, 1997). Community-based instruction has been cited to be an effective method for struggling readers, as it provides contextual cues in addition to text cues (Cuvo & Klatt, 1992). However, this method of sight word instruction is time consuming and is sometimes logistically impossible as it requires educators and students to leave the school grounds (Cuvo & Klatt, 1992). Therefore, flashcard drill and practice methods, which can be conducted in the school setting, have long been used as a technique for teaching sight word recognition (i.e., estimates indicate that they have been used since at least the 19th century; Pruzan, 2005).

Flashcard drill techniques have been dubbed this term because they employ the use of unknown stimuli (e.g., sight words or math facts), which are typically presented on flashcards until the unknown stimuli become known. A noteworthy feature of these methods is that they involve an important instructional strategy of repeated practice, which increases opportunities to respond to unknown stimuli (Cates, Burns, & Joseph, 2010; Cooke & Guzaukas, 1993). These increased opportunities to respond directly translate into student engagement and learning opportunities (Cates et al., 2010). As a result of these features, flashcard drill techniques have been associated with increased learning rates and are also well accepted by educators and school psychologists (Greenwood, Delquadri, & Hall, 1984).

A variety of flashcard drill techniques have been developed throughout the years. Although not exhaustive, flashcard drill and practice techniques that have received noteworthy
attention in the education and school psychology literature include: Traditional Drill and Practice (TDP), Constant Time Delay (CTD), Interspersal Training (IST), Direct Instruction (DI), and Incremental Rehearsal (IR). A review of the procedures for each intervention is provided below.

**Flashcard Drill and Practice Procedures**

**Traditional Drill and Practice**

Traditional Drill and Practice (TDP) is a standard flashcard drill method for teaching sight words to students in which 100% of the words presented are unknown or cannot be read with automaticity by the student (e.g., 1U-2U-3U-4U-5U-1U-2U-3U-4U-5U, where U is an unknown word; Tan & Nicholson, 1997; Nist & Joseph, 2008; Volpe, Mulé, Briesch, Joseph, & Burns, 2011). In the initial trial, the interventionist presents a word to the student, models reading the word accurately, and then asks the student to read the word aloud. Once all words have been modeled, the student attempts to read the words independently and is provided with corrective feedback on any miscues (words read inaccurately). Once the interventionist has cycled through all of the unknown words he or she shuffles the flashcards, altering the order of the target words. Once flashcards have been shuffled they are presented again without modeling. The student is asked to read the words aloud and is provided with immediate feedback. This procedure typically continues until the student can automatically recognize all of the unknown words targeted in that session (or until they recognize them for a number of consecutive trials).

**Constant Time Delay**

Constant time delay (CTD) is a response prompting procedure that has been combined with flashcard drill instruction (Browder & Minarovic, 2000; Keel, Slaton, & Blackhurst, 2001). It has been successful in teaching multiplication facts, science words, spelling, and word definitions (Knight, Ross, Taylor, & Ramasamy, 2003). Therefore, its use has been extended to
teaching sight word recognition (Knight et al., 2003). Like TDP, CTD utilizes only unknown words. In the initial trial of instruction, the interventionist presents a word to the student, immediately models reading the word accurately (0 s time delay), and then asks the student to read the word aloud themselves. Once all words receive training, the student is asked to read the words independently. If the student gives a correct response, verbal praise is provided by the interventionist. If they mispronounce the word or hesitate, the interventionist provides error correction. However, error correction is offered at a fixed delay interval (i.e., the number of seconds in which the interventionist waits before modeling the correct pronunciation). Several studies have used a 4 s delay or a progressive delay schedule (e.g., modeling is first delayed by 2 s, then 4 s, 6 s, and 8 s; Browder & Minarovic, 2000; Keel et al., 2001; Knight et al., 2003). The delay procedures allow students the opportunity to respond without assistance when possible, with the overall goal of being able to consistently and accurately respond to the target word prior to error correction (Keel et al., 2001). The CTD procedures typically continue until the student is able to read each word two consecutive times without the need for modeling or error correction (Browder & Minarovic, 2000).

**Interspersal Training**

Interspersal Training (IST, sometimes referred to as "Drill Sandwich" and "Interspersal of Known Items") is another flashcard drill method that is frequently used as an alternative to the TDP method (Nist & Joseph, 2008). It is distinct from TDP because it intersperses (or "sandwiches") known words between unknown words (Neef, Iwata, & Page, 1980). This practice is rooted in reinforcement principles that suggest that students are more likely to engage in reading unfamiliar words when they experience success in reading words that are familiar to them (Joseph, 2006; Neef et al., 1980). In IST, the ratio of known to unknown words can vary
and should be determined with reference to the student’s skill level (Cooke & Richard, 1996; Joseph & Nist, 2006). Once the ratio of known to unknown words has been determined, the unknown words are modeled by the instructor, and then are practiced independently by the student, with error correction. The unknown words are then arranged with the known words (i.e., 1K-1U-2U-3U-2K-4U-5U-6U-3K, where K is a known word and U is an unknown word) and presented to the student. The student receives corrective feedback for miscues and verbal praise for correct responses.

Direct Instruction

Direct Instruction (DI) is another IST flashcard drill technique that appears in the education literature. Like IST, the ratio of known to unknown words varies based on the student’s skill level. Initial presentations of unknown words receive modeling by the interventionist. The student is then asked to repeat the word. On subsequent trials words are presented in random order, and the student is asked to read the target words independently. If they respond correctly, verbal praise is provided and the interventionist places the flashcard at the back of the deck (Ruwe, McLaughlin, Derby, & Johnson, 2011). If the student hesitates or responds incorrectly, error correction is provided, the student is asked to repeat the word, and then the card is placed two or three cards back from the front of the flashcard stack. This procedure continues until the student is able to accurately respond to the target word for three consecutive presentations (Ruwe et al., 2011).

Incremental Rehearsal

Finally, Incremental Rehearsal (IR, also referred to as the "folding-in technique") is an IST flashcard drill technique, which utilizes both known and unknown words for intervention (Shapiro, 2004). In contrast to IST and DI, IR is unique as it typically employs a 9:1 ratio of
known to unknown words (Tucker, 1988); however, this ratio can vary depending on the student's developmental level (e.g., acquisition rate). Additionally, unknown words are gradually introduced or incrementally presented to the student (e.g., 1U, 1K, 1U, 1K, 2K, 1U, 1K, 2K, 3K, etc., where 1U is the first unknown word and 1K is the first known word; MacQuarrie, Tucker, Burns, & Hartman, 2002). The unknown words are first modeled by the interventionist and the student is asked to repeat them (error correction and verbal praise are provided throughout the procedure). Once it has been established that the student is able to correctly pronounce each word, the interventionist arranges the flashcards in the order described above. Subsequently, the first unknown word is presented, followed by the first known word, the first unknown is presented again, followed by the first known and the second known, and so on. This pattern of presentation continues until the unknown word had been practiced with all of the identified known words. At this juncture, it is assumed that the student has learned the unknown word; consequently, it becomes the first known word in the next IR sequence, which increases opportunities for practice and should further support retention of the learned word (MacQuarrie et al., 2002; Nist & Joseph, 2008; Szadokierski & Burns, 2008). The last known word is then removed from the sequence, a new unknown word is added, and the procedure is repeated.

**Research to Practice Gap**

Despite substantial research regarding effective interventions for sight word development, there are no existing data that examine the applied usage of flashcard drill procedures in school settings (i.e., the frequency with which flashcard drill methods are used as interventions to target sight word recognition). Additionally, survey data suggests that school psychologists do not always use research to inform practice (or to make remediation recommendations to educators); instead, many rely on personal experiences to guide their
practice and recommendations (Bramlett, Murphy, Johnson, Wallingsford, & Hall, 2002; Bramlett, Cates, Savina, & Lauinger, 2010; Walker, 2004). For example, Bramlett et al. (2010) reported that few school psychologists use peer reviewed journals to gather intervention ideas and even fewer develop, implement, and evaluate interventions. Unfortunately, data also suggest that teachers seldom use research to inform their practice as well (Hirsch, 2003; Levy & Vaughn, 2002). An observational study revealed that teachers do not always employ reading instruction methods that are evidence-based (Levy & Vaughn, 2002). Additionally, for both school psychologists and teachers, Walker (2004) has speculated that there is close to a 20 year time lag between the development of evidence-based practices and their integration into routine practice.

Walker (2004) suggests that this research to practice gap likely occurs for several reasons. First, school psychologists and educators are under enormous pressures as class sizes increase, while staffing numbers decrease as a result of state budget cuts (Walker, 2004). Furthermore, federal accountability pressures have served to further escalate these pressures, which result in school environments that constrain school psychologists’ and educators' abilities to access peer reviewed journal articles for each learning problem either referred to them, or found within their classroom (Bramlett et al., 2010; Walker, 2004). This is troubling, particularly for reading, as it is the most common difficulty for which students are referred to school psychologists (Bahr et al., 2006; Bramlett et al., 2002). It is even more concerning because data suggest that many school psychologists do not feel entirely equipped to intervene with academic referral concerns (Bramlett et al., 2002; Walker, 2004). This problem has reached a climax in recent years with the widespread adoption of response to intervention (RtI; Skinner, 2008).

As mentioned above, because time is a critical factor for practicing school psychologists and teachers, it is essential for interventions to not only be effective, but efficient as well. That is,
interventions that result in greater amounts of gain (or learning) as a function of instructional
time should be utilized (e.g., an intervention that results in 1 learned word per minute is more
*efficient* than an intervention that results in 1 learned words per 3 minutes of instruction; Cates et
al., 2003). Although there is an advantage for school professionals to employ efficient methods
of intervention so that they are able to spend less 1:1 time with a particular student and more
time teaching groups of students, there is also a significant advantage for students as the
consumers of these interventions (Skinner, 2008). Specifically, learning problems can be
operationalized as difficulty learning at the same rate as a student's same-aged peers (Skinner,
2008); consequently, if specific interventions can be identified to result in greater learning rates,
remediation can occur more rapidly and allow struggling readers to more readily catch up to their
peers (Cates et al., 2003; Skinner, Fletcher, & Hennington, 1996).

In response to the need for school psychologists and educators to employ evidence- based
interventions that are also efficient methods, the current study was conducted to provide a critical
review and synthesis of the extant research evaluating flashcard drill interventions used to target
sight word recognition. Specifically, this review is intended to consolidate the current literature
on sight word interventions so that this knowledge is more accessible to current practitioners
who are faced with competing demands. Additionally, it is hoped that this review will also
inform future lines of research. The following research questions guided this review:

1. For what populations have flashcard drill methods targeting sight word recognition
   been implemented?

2. What research design features (e.g., number of words targeted, number of
   opportunities to respond provided, time allotted, etc) have been used in flashcard drill
   methods targeting sight word recognition?
3. Which flashcard drill method was most effective in improving sight word recognition?

4. Which flashcard drill method was most efficient in improving sight word recognition?

**Method**

**Procedures**

Three commonly used electronic databases (ERIC, PsycArticles, and PsycINFO) were utilized to identify research articles evaluating instructional effectiveness and efficiency of flashcard drill methods targeting sight word recognition. The ERIC, PsycArticles, and PsycINFO electronic databases were utilized in August 2011 to locate articles for this review. The search terms used to conduct the search were limited to: (a) sight word recognition, (b) sight word, (c) flashcards, (d) drill and practice, (e) flashcard drill and practice methods, (f) whole word reading, (g) traditional drill and practice, (h) incremental rehearsal, and (i) interspersal training. The following criteria were used to select articles for the current review:

1. The effects of flashcard sight word interventions were examined using single-case or group design.

2. The dependent variables under investigation were measures of instructional effectiveness (e.g., retention of learned words) and/or efficiency.

3. Intervention was carried out in a 1:1 setting.

4. The study involved students or children in the primary or middle school grades.

5. The study was published in a peer-reviewed journal.

6. The interventions were described with enough detail so that they could be replicated.

7. The study was published within the past decade (2001 – 2011).
8. The study was written in English.

Utilizing the nine search terms above, the database searches produced a total of 610 references, of which many were duplicates due to the use of multiple databases (sight word recognition resulted in 22 hits, sight words 129, flashcards 37, drill and practice 366, flashcard drill and practice 3, whole word reading 18, traditional drill and practice 12, incremental rehearsal 19, and interspersal training 4). Abstracts were reviewed in order to identify studies meeting the inclusionary criteria. If abstracts did not provide adequate information, the article was reviewed further.

A total of 591 articles were excluded on the basis of the inclusionary criteria. Specifically, 22 studies were excluded because their participants were not primary or middle school students (e.g., some intervention studies have been carried out with high school age and college students, as well as adults with cognitive impairment), one study was excluded because it was conducted in a group format, 52 did not employ flashcards as a means of instruction, 20 were review or theoretical articles, two studies did not describe the intervention with enough detail (i.e., it could not be discerned which flashcard drill technique was employed), and 465 did not examine sight word learning as a dependent variable (e.g., some studies examined spelling, math facts, and physics. Other studies examined the effect of sight word training on disruptive behavior, the relationship between reading proficiency and mental health, etc).

Sixteen articles were determined to have met the selection criteria listed above (Note. The sum of excluded articles and articles meeting selection criteria do not total to the 610 produced references because many articles were duplicated across search terms. Additionally, one article contained two empirical studies; therefore 17 studies are contained within this review). These 16 articles were subsequently reviewed by the primary researcher. Specifically, the primary
A researcher reviewed the article for the five content areas: (1) sample characteristics; (2) interventionist and setting characteristics; (3) study components, (4) instructional effectiveness, and (5) instructional efficiency.

Within sample characteristics, demographic information (e.g., age, grade level, ethnic background of students, disability status, and English language learners [ELL]) were coded. Within interventionist and setting characteristics, interventionist characteristics (e.g., intervention conducted by practicing school psychologist, teacher, or graduate student research team members), and setting characteristics (i.e., the environment in which the intervention was conducted in) were coded. Next, within study components the following content was coded: (1) intervention type, (2) intervention design (e.g., single-subject or group), (3) number of sessions, (4) time required to implement the intervention, (5) number of words targeted in each session, and (6) the number of opportunities to respond that were given to the students. Finally, effectiveness and efficiency data were coded as separate content areas. Each sample characteristic and study component was assigned a number and coded in an Excel spreadsheet.

**Interrater Agreement**

All studies were reviewed by the primary researcher. Over a third (38%, n = 6) were then randomly selected and coded by a research assistant (doctoral student in school psychology) to establish interrater agreement. The second researcher coded the six randomly selected articles for the same five content areas mentioned above. Agreement was calculated by dividing the total number agreements, by the total number agreements plus disagreements, and then multiplying by 100. Interrater agreement was 100% across all categories.
Results

Sample Characteristics

Characteristics of the sixteen articles meeting the selection criteria for this study are summarized in Table 1 and Table 2. A total of 288 students participated in the studies reviewed. Of these participants, 44% \((n = 127)\) were males, 35% \((n = 101)\) were females, and 21% \((n = 60)\) were not reported. Thirty-nine percent \((n = 113)\) of participants were Caucasian, 12% \((n = 35)\) African American, 1% \((n = 4)\) Hispanic, 1% \((n = 3)\) Multiracial, less than 1% \((n = 2)\) Asian, and less than 1% \((n = 1)\) Somalian (for 45% \([n = 130]\) of participants' race was not reported). Eleven studies reported participant age, which ranged from 6- to 14-years-old \((Note. A mean age was not calculated because not all studies reported the age of their participants)\). As reported in all 17 of the studies, students grade levels ranged from the first through the eighth grades. Twenty-five percent \((n = 72)\) of participants were regular education students and 27% \((n = 67)\) of students qualified for special education services (four studies did not report the educational status of their participants, leaving the educational placement status of 44% \([n = 128]\) of students unknown). Thirteen studies reported disability status of participants, and five further delineated disabilities. Sixteen students were determined to have cognitive disabilities and 17 had learning disabilities. Finally, only one student participant was an English Language Learner (ELL).

Interventionist and Setting Characteristics

Undergraduate and graduate students in school psychology or special education were most often responsible for carrying out interventions with the student participants \((n = 13)\). However, in two studies a practicing school psychologist delivered intervention, one was carried out by a teacher, and another by a university faculty researcher. Finally, in almost all of the reviewed studies \((n = 16)\), interventions were delivered in a private or semi-private space that
was free from external noises and distractions. In the remaining study, the intervention was purposely carried out in a semi-private space, where distracters could be viewed, so that the effect of drill ratios on recall and on-task behavior could be measured.

**Study Components**

**Intervention type.** Intervention studies were grouped according to flashcard drill method. The following groupings emerged: IR, TDP, IST, CT, and DI *(Note.* In some cases flashcard drill methods were compared to other intervention methods that did not meet the selection criteria for this review [e.g., phonics analysis]; these interventions are beyond the scope of this review and will only be given a brief mention to share with the reader how their effects compared to the effects of flashcard drill methods). A large number of studies (75%, \(n = 12\)) included within this review tested the effects of IR on sight word learning. Slightly fewer studies (63%, \(n = 10\)) examined TDP as an instructional condition. Finally, IST, CT, and DI were the least frequently investigated methods among the studies reviewed, with only 31% (\(n = 5\)), 6% (\(n = 1\)), and 6% (\(n = 1\)) examining effects on sight word recognition, respectively *(Note.* Percentages do not total 100% because several studies examined more than one instructional condition).

**Study design.** Of the 17 identified studies, 53% (\(n = 9\)) utilized a single-subject design (three of those utilized an alternating treatment design, two utilized a multiple-baseline design, one utilized a multi-element design, one utilized a parallel treatment design, one utilized an A-B-A-B design, and one utilized an A-B-C-D design). The remaining eight studies employed group designs that made use of a variety of statistical methods to assess treatment effects (e.g., paired-samples \(t\) tests, analysis of variance, or multivariate analysis of covariance).
**Duration of intervention and required time.** Across studies the duration of intervention varied from 1 to 20 sessions. A mean and standard deviation are not reported because reviewed studies utilized different study design methods. For example, studies employing multiple-baseline designs that evaluated only one intervention varied the number of sessions by word sets; whereas other alternating treatment design studies held the number of sessions constant across conditions.

Regarding the length of time required to carry out interventions, studies inconsistently reported on these data or gave approximations in place of exact figures; therefore, a mean and standard deviation for this variable are not reported. Additionally, several studies examined multiple intervention conditions within one intervention session and only reported the total time required for the intervention session; therefore, it is difficult to parse out how long individual or isolated intervention conditions took to implement. However, for the seven studies reporting this variable, the length of sessions ranged from 3 to 30 minutes, with the number of conditions ranging from one to four (Note. The number of target words within each condition varied across studies as well; therefore, it is difficult to draw strong conclusions about the length of time required to implement each intervention).

Regarding the length of time to administer individual instructional conditions, one study reported that DI took approximately 5 minutes to implement (Note. This study reported only an approximation; Ruwe et al., 2011). CTD took approximately 15 minutes to implement (Note. This is again, an approximation; Knight et al., 2003). IST took an average of 2.49 minutes to implement (Nist & Joseph, 2008). Nist and Joseph (2008) and Burns and Sterling-Turner (2010) respectively reported that IR took an average of 6.71 and 8.90 minutes to implement (Note. In both studies six words were targeted for instruction in the IR condition). When opportunities to
respond were held constant and only three words were targeted in an modified IR (with omission of folding-in technique), Volpe et al. (2011) found that the modified IR took an average of 2.99 minutes. Other studies providing an estimation of the time required to implement IR, reported that it took upwards to 15 minutes and took approximately 5-15 minutes longer than other conditions (Burns et al., 2004; Matchett & Burns, 2009). TDP was reported to take between 0.85 to 3.52 minutes (Burns & Sterling-Turner, 2010; Nist & Joseph, 2008; Volpe et al., 2011). A noteworthy finding from comparison studies is that the IR method took considerably longer to administer than both the IST and TDP conditions (Burns & Sterling-Turner, 2010; Joseph & Schisler, 2007; Nist & Joseph, 2008; Volpe et al., 2011).

**Number of words targeted.** The majority of studies ($n = 11$) determined a priori the number of words that would be targeted for instruction per session. They ranged from 5 to 18 ($M = 10.91; SD = 4.97$). The remaining studies allowed the number of target words to vary. The number of words targeted for instruction was selected based on participants' performance on grade level passages (i.e., miscued words were used for intervention; Burns et al., 2004; Joseph & Schisler, 2007; Martin-Chang et al., 2007). Additionally, in one study intervention was discontinued once the participant made a predetermined number of errors (Matchett & Burns, 2009). Therefore, it was possible that the interventionists only introduced two words per session because the student participant made a number of errors; conversely, the student could have performed better in another session, resulting in a greater number of target words. Finally, the Ruwe et al. (2011) study allowed for the ratio of known to unknown words to vary (therefore, allowing the number of target words to vary) across participants but did not explain why they chose to do so.
**Opportunities to respond.** The number of opportunities to respond (i.e., the interaction between a student’s response to instructional stimuli and feedback received from the interventionist; Greenwood et al., 1984), was inconsistently reported across studies. Four studies made no mention of these data (Burns et al., 2004; Joseph & Nist, 2006; Joseph & Schisler, 2007; Matchett & Burns, 2009). A number of studies \((n = 8)\) allowed for opportunities to respond to vary across instructional conditions and by student. Specifically, opportunities for practice ranged from 3 to 45. Although many studies allowed for opportunities to respond to vary, five studies elected to hold this study component constant to allow for direct comparisons across instructional conditions (Martin-Chang et al., 2007; Nist & Joseph, 2008; Schmidgall & Joseph, 2007; Volpe et al., 2011). Nist and Joseph (2008) was one study that reported having held constant the number of opportunities for practice across TDP, IST, and IR instructional conditions. However, investigators failed to consider the additional trials of practice afforded by the folding-in technique utilized in the IR condition. Consequently, opportunities to respond were not truly held constant as they were reported to be in this study. Nevertheless, five of the studies contained within this review were able to truly hold opportunities to respond constant across conditions (Joseph & Nist, 2006; Martin-Chang et al, 2007; Schmidgall & Joseph, 2007; Volpe et al., 2011). Joseph and Nist (2006) allowed for 18 opportunities for practice across three intervention conditions: High-Probability Sequencing, IST, and TDP. Martin-Chang et al. (2007) allowed for four opportunities to practice target words across both TDP and a Context Training condition. When comparing the effects of TDP, IST, and Wordboxes (a phonics analysis technique), Schmigdall and Joseph (2007) allowed for six opportunities to practice each target word. When comparing the effects of a modified IR (which eliminated the folding-in technique) to TDP, Volpe et al. (2011) held opportunities to respond constant, allowing for five practice
opportunities for each unknown word. It should be noted, however, that when IR is implemented with the use of the folding-in technique, it results in a greater number of opportunities to respond than other instructional conditions (i.e., TDP and IST; Greenwood et al., 1984; Szadokierski & Burns, 2008).

Instructional Effectiveness

A measure of instructional effectiveness was reported across all studies \((n = 17)\). Several different measures of effectiveness were reported, including: (a) retention (i.e., an assessment of initial acquisition that took place during the intervention session or on the subsequent day; \(n = 10\)); (b) maintenance (i.e., retention of learned words over a delayed period of time \((n = 12)\)
\(^1\), (c) generalization (i.e., ability to read sight words in the context of a sentence or passage; \(n = 6\)), (d) sight word recognition fluency (i.e., the rate at which students were able to read a target word list; \(n = 1\)), (e) oral reading fluency (i.e., the rate at which students were able to read passages that included target words; \(n = 2\)), and (f) reading comprehension (i.e., students' ability to accurately answer reading comprehension questions; \(n = 1\)).

Fifty-nine percent \((n = 10)\) of studies reported effect sizes to demonstrate intervention effects (i.e., two single subject design studies and all eight group design studies). The remaining studies utilized visual analysis and descriptive statistics to evaluate effectiveness of instructional conditions. It should be noted that it is challenging to summarize unique effect sizes across studies, as many of the studies contained within this review compared the effects of multiple interventions, employed different experimental designs, and rarely evaluated intervention conditions in isolation from one another (Note. Burns et al. [2004]; Matchett & Burns, [2009];

\(^1\) Delays varied across studies from 5-30 days, with a majority of studies opting for a one week delay period. Additionally, while some studies report conducting retention assessments, because assessments took place with such a large gap between the time of intervention and assessment (e.g., one week post-intervention or greater) they have been categorized as conducting maintenance assessments instead. This decision was made to avoid confusion for the reader and to provide an accurate account of each study's methods.
and Ruwe et al. [2011] did evaluate IR and DI in isolation from other sight word recognition interventions). Nevertheless, below effectiveness data are discussed in more detail.

**Incremental rehearsal.** In studies where IR was examined in isolation utilizing a multiple-baseline design (Matchett & Burns, 2009) or versus a no treatment control condition (Burns et al., 2004), it was found to be an effective method for teaching sight words, as measured by sight word recognition fluency and oral reading fluency, respectively. Burns et al. (2004) also found that IR was an effective method for increasing reading comprehension. When traditional IR procedures were enhanced by increasing opportunities to respond, it was further supported as an effective method for teaching sight word recognition to children (Burns, 2007a; Szadokierski & Burns, 2008); in both studies maintenance assessments were used as a measure of instructional effectiveness. Interestingly, when Burns and Dean (2005a) manipulated the ratio of known to unknown words in the IR procedure (i.e., 0%, 50%, 83%, and 90% known) they found that a 9:1 ratio of known to unknown words led to the greatest maintenance. Furthermore, when Szadokierski and Burns (2008) manipulated the ratio of known to unknown words (i.e., increasing and decreasing the ratio) and opportunities to respond by replacing unknown words with new unknown words, the ratio of known to unknown words had no significant effect on maintenance data. However, this study revealed a significant main effect for opportunities to respond. This finding seems to support prior research advocating for the use of a 9:1 ratio, as this ratio leads to the greatest number of opportunities to practice the unknown or target word.

When IR was compared to other sight word recognition interventions (e.g., TDP, IST, and Word Boxes) there were mixed results. Consistent with the aforementioned studies, when IR was compared to TDP and IST, it was the most effective when comparing retention (Burns & Sterling-Turner, 2010; MacQuarrie et al., 2002; Nist & Joseph, 2008), maintenance (Burns &
Boice, 2009; Burns & Dean, 2005a; Burns & Sterling-Turner, 2010; MacQuarrie et al., 2002; Nist & Joseph, 2008), and generalization (Nist & Joseph, 2008) assessment results. That is, results indicated that IR led to better retention, maintenance, and generalization of learned words and was a superior intervention to TDP and IST. It should be noted, however, that because of the folding-in technique, the IR condition most likely resulted in a higher number of opportunities for practice than the TDP and IST conditions. In contrast, in one study (i.e., Joseph & Schisler, 2007) where opportunities to respond varied across conditions and oral reading fluency was used as the dependent variable, differences between IR and TDP were not statistically significant. Interestingly, when Volpe et al. (2011) held time and opportunities to respond constant in a multi-element design, they also found minimal differences between a modified IR (omission of folding in technique) and TDP (Note. In this study retention, maintenance, and generalization were all assessed). Finally, when IR was compared to CTD (allowing opportunities for practice to vary) utilizing retention assessment data, CTD was a more effective intervention for students with cognitive impairments; however, IR was equally as effective as CTD for students with specific learning disabilities (Knight et al., 2003).

IR has been tested with difficult to remediate populations (i.e., cognitively impaired and ELL). Results from these studies demonstrate that IR is an effective intervention for children with special learning needs (Burns, 2007a; Burns & Boice, 2009; Burns & Dean, 2005a; Knight et al., 2003; Matchett & Burns, 2009). However, Joseph and Schisler (2007) and Knight et al. (2003) found that IR was no more effective for children in special education than TDP, Word Boxes (i.e., a variation of "say it move it" phonic analysis), or CTD.

**Traditional drill and practice.** Effectiveness data for TDP were mixed. Although TDP was an effective method for teaching sight words, it was not consistently as effective as other
methods. Four studies found that IR was superior to TDP when examining retention (Burns & Boice, 2009; Burns & Sterling-Turner, 2010; MacQuarrie et al., 2002; Nist & Joseph, 2008), maintenance (Burns & Sterling-Turner, 2010; MacQuarrie et al., 2002; Nist & Joseph, 2008), and generalization (Nist & Joseph, 2008). Additionally, one study found that Word Boxes was more effective across retention, maintenance, and generalization assessments (Schmidgall & Joseph, 2007). One additional study found that Context Training (intervention that teaches target words in stories, which provides additional contextual cues to the learner) was superior to TDP for initial acquisition, maintenance, and generalization (Martin-Chang et al., 2006). However, when this study was replicated, TDP resulted in better generalization than Context Training and a control condition.

A number of studies have reported no significant differences between TDP and other experimental conditions when holding opportunities to respond constant, and even when allowing them to vary (i.e., IR, IST, and high-probability sequencing [variation of IST with less challenging ratio of unknown to known words]; Burns & Boice, 2009; Joseph & Nist, 2006; Joseph & Schisler, 2007; Schmidgall & Joseph, 2007; Volpe et al., 2011). Interestingly, while Joseph and Schisler (2007) did not find significant differences between TDP, IR, and Word Boxes in terms of cumulative words learned, they found that TDP resulted in the highest reading fluency rates on a curriculum based measurement probe.

In addition to the instructional effectiveness data presented above, two interesting findings emerged from the literature. First, only three studies examined TDP with special education populations (Burns & Boice, 2009; Burns & Dean, 2005a; Joseph & Schisler, 2007). Second, findings from these three studies were mixed. Whereas, Burns and Boice (2009) and Burns and Dean (2005a) found that IR was superior to TDP in terms of effectiveness (as
measured by maintenance assessments), Joseph and Schisler (2007) found that TDP did not fare better or worse than IR and Word Boxes when the measure of instructional effectiveness was oral reading fluency.

**Interspersal training.** Instructional effectiveness findings for IST were also mixed. Instructional effectiveness data from Burns and Boice (2009), Nist and Joseph (2008) and Schmidgall and Joseph (2007) suggests that IST follows IR and Word Boxes in terms of its effectiveness (*Note*. Nist & Joseph [2008] and Schmidgall & Joseph [2007] examined retention, maintenance, and generalization; whereas Burns & Boice [2009] only examined maintenance). Interestingly, when opportunities to respond were allowed to vary and even when they were held constant, studies reported that there were no significant differences between IST and other experimental conditions (i.e., High-Probability Sequencing and TDP; Burns & Boice, 2009; Joseph & Nist, 2006; Schmidgall & Joseph, 2007). Finally, Burns and Boice (2009) were the only researchers to examine IST with a special education population. Their findings suggest that IR was more effective than IST, but that there were no significant differences between IST and TDP in increasing sight word recognition for special education populations.

**Direct Instruction.** The search terms resulted in one DI study that evaluated effects utilizing a multiple-baseline, single-subject design (Ruhe et al., 2011). Utilizing a same day retention probe (word list containing both taught and untaught words), data suggested that DI is an effective method for teaching sight words for children with cognitive impairments. Generalization data produced less clear results, however. Ruwe et al. (2011) report that students had a tendency to guess at words while reading a passage probe. However, instances of guessing for trained words were lower than on untrained words (Ruhe et al., 2011).
Constant time delay. Finally, when the effects of CTD were compared to those of IR in a parallel treatments design, utilizing next day retention probes, CTD was found to be more effective for students with cognitive impairments (Knight et al., 2003). Interestingly, IR was equally effective as CTD for students with specific learning disabilities (Knight et al., 2003).

Instructional Efficiency

Of the studies contained within this review, only 44% (n = 7) reported instructional efficiency. The remaining studies (n = 10) did not provide efficiency data and did not report the necessary variables (i.e., length of intervention by instructional condition) needed to extrapolate instructional efficiency.

Five different efficiency calculations were conducted across the seven studies, including: (1) retention efficiency, calculated by multiplying the cumulative number of words read accurately (WRA) at retention assessment by 60 s and then dividing the sum by the total instructional time in seconds (Retention WRA x 60 s/ Instructional Time [s]); (2) maintenance efficiency, calculated by multiplying the cumulative number of WRA at maintenance assessment by 60 s and then dividing the by the sum of the total instructional time in seconds (Maintenance WRA x 60 s/ Instructional Time [s]), (3) generalization efficiency, calculated by multiplying the cumulative number of WRA at generalization assessment by 60 s and then dividing the sum by the total instructional time in seconds (Generalization WRA x 60 s/ Instructional Time [s]), (4) fluency efficiency, calculated by multiplying total oral reading fluency (ORF) score by 60 s and then dividing by the sum of the total instructional time in seconds (ORF x 60 s/ Instructional Time [s]), and (5) triad efficiency (i.e., efficiency by which students were able to accurately read sets of three words presented during a single instructional session), calculated by multiplying the total number of triads read accurately (TRA) by 60 s and then dividing by the sum of the total
Instructional time in seconds (TRA x 60 s/ Instructional Time [s]). Below efficiency data are discussed in more detail.

**Incremental rehearsal.** Only five of the 12 studies that evaluated IR reported efficiency data (Burns & Sterling-Turner, 2010; Joseph & Schisler, 2007; Knight et al., 2003; Nist & Joseph, 2008; Volpe et al., 2011). Efficiency of IR was calculated using retention (Range = 0.31 - 0.69), maintenance (Range = 0.15 - 0.46), generalization (Range = 0.15 - 0.21), oral reading fluency (28.33 words read/instructional time), and triad mastery data (0.035 triads learned per minute). Despite several measures of efficiency, these five comparison studies rather consistently reported IR as the least efficient sight word intervention. That is, the rate of skill acquisition was greater for other interventions than it was for IR. However, because interspersal methods require considerably more time to administer, it is not surprising that numerous studies have found IR to be the least efficient. Interestingly, in contrast and worthy of note, when Burns and Sterling-Turner (2010) examined IR and TDP’s instructional efficiency with maintenance data (i.e., words were maintained over time), there were no statistically significant differences between the two instructional conditions. Additionally, when Volpe et al. (2011) compared instructional efficiency of a modified IR to TDP (holding both time and opportunities to respond constant) using next day retention, maintenance, and generalization data, they found only minimal differences between the two instructional conditions, with TDP only faring better than IR in the conditions where opportunities to respond were held constant. Further, when IR's efficiency was evaluated using the number triads mastered, it was less efficient than CTD for students with specific learning disabilities and cognitive impairments (Knight et al., 2003).

**Traditional drill and practice.** Only six of the eight articles that investigated TDP reported efficiency data (Burns & Sterling-Turner, 2010; Joseph & Nist, 2006; Joseph &
In all six studies, TDP produced the most efficient rates of learning at retention (Range = 0.34 - 2.79), maintenance (Range = 0.21 - 2.38), generalization (Range = 0.19 - 0.90), and oral reading fluency assessments (54.41 words read/instructional time). That is, TDP resulted in the greatest gains in students' ability to retain, maintain, and generalize learned words (Note. only in the opportunities to respond held constant condition of the Volpe et al. [2011] study did TDP fare better than IR.). TDP was also the most efficient method for increasing oral reading fluency (Joseph & Schisler, 2007). However, as mentioned above, when maintenance data were used to calculate instructional efficiency, Burns and Sterling-Turner (2010) reported no statistically significant differences between TDP and IR. Similarly, when time was held constant in the Volpe et al. (2011) study there were no significant difference between TDP and IR across retention, maintenance, and generalization efficiency calculations.

**Interspersal training.** Finally, three of the five articles that examined IST as an instructional condition reported instructional efficiency (Joseph & Nist, 2006; Nist & Joseph, 2008; Schmidgall & Joseph, 2007). Studies reported efficiency rates using retention (Range = 0.71 - 1.70) and maintenance assessment data (Range = 0.61 - 1.31). Across all three studies, IST was more efficient than IR, but less efficient than the TDP condition.

**Constant time delay.** Utilizing an assessment of triads mastered, Knight and colleagues (2003) report that CTD was more efficient than IR for students with cognitive impairments and for students with specific learning disabilities. Students learned approximately 0.0975 triads per minute.

**Direct instruction.** No instructional efficiency data were reported for the DI condition.
Discussion

The aim of this review was to consolidate and summarize the literature on flashcard drill methods targeting sight word recognition to inform future practice and research. Specifically, sample characteristics; research design features; instructional effectiveness; and instructional efficiency of flashcard drill interventions were examined. The literature review identified the following five flashcard drill methods targeting sight word recognition: IR, TDP, IST, CTD, and DI. Across methods children from various backgrounds (i.e., gender and race) with a variety of learning profiles (e.g., typically developing children, children with learning disabilities and cognitive impairments, and children who are ELL) were targeted for intervention; and a variety of research design features were employed across studies (i.e., number of target words, opportunities to respond, instructional time allotted, etc). The results from this review demonstrate variable instructional effectiveness and efficiency of flashcard drill methods across studies and methods, and are discussed in further detail below.

Although all intervention conditions helped children to make gains in their sight word knowledge, instructional effectiveness results were somewhat inconclusive with respect to which flashcard drill technique was most superior. As a result, it is difficult to suggest one instructional method over the other. Furthermore, while studies' instructional effectiveness results were mixed, it is further challenging to distinguish if one instructional method is superior to another because comparison studies have not yet been conducted with all of the instructional conditions examined in this review (i.e., CTD was only compared to IR, and DI has only been evaluated in isolation). Additionally, very few comparison studies tightly controlled for time and

---

2 Select flashcard drill methods reviewed in this study have been implemented with choice populations; that is, not all methods have been tested across all learning profiles identified in this study (e.g., IR has been evaluated with an ELL child, whereas DI has not been). Likewise, research design features varied across methods (e.g., IR consistently resulted in more opportunities to respond than any other flashcard drill method [with the exception of the Volpe et al., 2011 study]).
opportunities to respond, making it challenging to make fair comparisons between intervention methods. In fact, only Joseph and Nist (2006) and Schmidgall and Joseph (2007) were able to hold opportunities to respond constant; and Volpe et al. (2011) were able to hold time and opportunities to respond constant. In these studies, that compared three to four instructional conditions, investigators found no significant differences between conditions (i.e., TDP, IST, and IR [devoid of folding-in technique]) in terms of effectiveness. What can be said, however, is that opportunities to respond seem to be tightly linked to learning. Therefore, in studies where practice opportunities were allowed to vary across conditions, intervention methods resulting in the most opportunities to respond most often resulted in more learning (Burns, 2007a; Burns & Boice, 2009; Burns & Dean, 2005a; Burns & Sterling-Turner, 2010; MacQuarrie et al., 2002; Nist & Joseph, 2008; Szadokierski & Burns, 2008).

With respect to instructional efficiency, results were again somewhat inconclusive. Conclusions are difficult to draw, as many studies did not report on instructional efficiency. Furthermore, instructional efficiency has not yet been compared across all instructional conditions (e.g., CTD has only been compared to IR and there were no efficiency data for DI). Nevertheless, there does seem to be stronger support for TDP as being the most efficient method for teaching sight word recognition (at least when analyzing initial acquisition of words), as all six studies evaluating its efficiency consistently reported it to be the most efficient method (Burns & Sterling-Turner, 2010; Joseph & Nist, 2006; Joseph & Schisler, 2007; Nist & Joseph, 2008; Schmidgall & Joseph, 2007; Volpe et al., 2011). For example, whereas children learned up to 2.79 words per minute in the TDP condition on retention measures, they learned up to 1.70 words per minute in the IST condition, and up to .69 words in the IR condition; additionally, .10 triads per minute were learned in the CTD condition. TDP was somewhat expected to be the
most efficient method as it only targets unknown words; whereas interspersal techniques devote instructional time to rehearsing known words as well. It should be noted that the single other technique that utilizes only unknown words, CTD, has not yet been compared to TDP and the dependent variable in this study was triads learned instead of the typical convention of words learned. However, when CTD was compared to IR, it was the most efficient for teaching sight words to children with intellectual and specific learning disabilities.

In terms of interspersal methods, several studies have compared the instructional efficiency of IST to TDP and IR, and it was reported to be the second most efficient, next to TDP. Additionally, when studies compared instructional efficiency of IR to TDP and IST, data consistently demonstrated IR to be the least efficient method for initial sight word acquisition. Interestingly, when maintenance data were used to examine instructional efficiency, IR was equally as efficient as TDP (Burns & Sterling-Turner, 2010).

The findings from this review are expected to assist practicing school psychologists, as well as educators, to make informed decisions when choosing among sight word interventions. Most importantly, school psychologists and educators need to consider what is ecologically feasible to implement in their setting (Bramlett et al., 2010). If school personnel are pressured by time constraints and competing demands, it seems reasonable to employ the TDP method, as it seems to have the most support in terms of its efficiency, so as to maximize learning during a specified instructional period. Furthermore, regardless of environmental pressures, some will argue that the most efficient instructional method should be used at all times because it results in greater learning rates, allowing struggling readers to more rapidly catch up to their peers.

In accordance and because of its noted inefficiencies, Skinner (2008) suggests that IR should be the last procedure called upon when the goal is remediation. However, because
interspersal techniques have increased student preference for academic tasks (Joseph, 2006; Neef et al., 1980), an argument has been made for using IR with children who are reluctant to engage in intervention or who have severe deficits (Burns & Sterling-Turner, 2010; McCurdy, Skinner, Grantham, Watson, & Hindman, 2001). Additionally, IR seems to be the most frequently examined and most well supported method in terms of remediating children with special learning needs (e.g., cognitively impaired, behavior disorders, specific learning disabilities, and ELL). Simply put, across intervention methods, rather than selecting any one method based solely upon its instructional effectiveness and efficiency data, the selection of interventions should be made on the basis of and in accordance to a child’s specific needs, as well as to how they are responding to intervention.

Beyond the issue of intervention selection, researchers and practitioners are urged to consider the number of unknown words targeted per intervention session (Burns, 2007b). In the studies examined within this review, there was a considerable range of words targeted in each session (i.e., 3 to 25). However, previous work by Burns and Dean (2005b) suggests that acquisition rates (i.e., the amount of new information a student can successfully rehearse and later recall) should determine the amount of information introduced in each session. Specifically, previous research has suggested five to six target words as the appropriate set for children in third through fifth grade (Burns, 2004a). Furthermore, work by Burns and Dean (2005b) suggests that acquisition rates for children with specific learning disabilities are lower, ranging from 2 to 4 words. Thus, instructed material above and beyond 4 units may exceed the capacity of a child with learning disabilities and result in a failure to learn new information or reduce retention of previously learned material (Burns & Dean, 2005b).
Limitations

Although the current study provides a timely review of the literature on flashcard drill methods used to increase sight word recognition, limitations are worth noting. First, over half of the studies included within this investigation were single-subject designs; therefore, the extent to which effects can be generalized to students who are different in some way from those investigated is somewhat limited (Horner et al., 2005). Additionally, as is mentioned in the sample characteristics, few ethnically diverse children are represented in the articles reviewed for this study; thus again, limiting generalizability to the wider population. Nevertheless, external validity is somewhat improved upon, as intervention studies have been replicated (Horner et al., 2005), although with slightly differing methods.

Second, although it was possible to calculate effect sizes for studies that failed to report them, the primary researcher decided otherwise. Namely, this practice was determined to be ineffective to practitioners and the wider scientific community because research designs differed across studies, negating useful comparisons. Additionally, although effect sizes could be compared by intervention design (i.e., single-subject and group designs), they still would not be easily comparable because study components (e.g., number of target words per session, number of intervention sessions, etc) still differed significantly across studies.

Third, the current review was limited to studies targeting sight word recognition using flashcard drill interventions, and therefore, did not include other methods (e.g., phonics analysis, word lists, and taped passages) for developing this skill. Consequently, the results from this study should not be oversimplified to summarize findings across all sight word recognition interventions, as it is possible that the results from this study could vary based on the investigations included for analysis.
Finally, as Burns (2004b) has noted in his empirical review of drill ratio research, review articles that only include published works may be susceptible to bias, as studies that result in non-significant effects are presumably less likely to be published. Therefore, given that this current review only includes studies that were published, this bias cannot be ruled out with certainty.

**Future Directions**

Although the current literature has provided an impressive foundation for flashcard drill methods that target sight word recognition, there are areas for which research could be improved. First and foremost, there is a general need for additional review articles that aim to summarize empirically supported interventions for practicing school psychologists and educators. Given the debilitating effects of reading problems on academic and social development (Rathvon, 2008), as well as the fact that the demands for practicing school psychologists and educators only continue to increase, there is a need for researchers to synthesize current studies and produce literatures that are geared toward helping practitioners select interventions that can ultimately help the children receive the interventions that they need.

Second, future investigations are encouraged to include children from ethnically diverse backgrounds within their sample. Additionally, studies that include children with special needs are encouraged to report the disability status of their participants (i.e., many studies reported that children were educated in special education classrooms, but did not indicate if they were classified as cognitively impaired, learning disabled, etc). Including diverse populations within future research and reporting on the disability status of children who have already participated in investigations, will enhance external validity and our ability to generalize findings to the wider population.
Third, although Volpe et al. (2011) and Burns (2007a) have begun this work, more studies need to examine and further identify what the active ingredients to these sight word interventions may be (e.g., opportunities to respond, time, ratio of unknown to known, etc). Having a better understanding of what the active ingredients include, will allow practitioners and researchers to perhaps strengthen intervention by enhancing those components.

Fourth, future lines of research are encouraged to address the most obvious gap in this literature review; that is, the lack of comparison studies between CTD and DI which seem to be used more frequently in education research, to IR, IST, and TDP which are more frequently investigated methods in the school psychology literature. These comparison studies need to be conducted so that we can better understand the spectrum of instructional effectiveness and efficiency. More than simply conducting these comparison studies, researchers need to be mindful of holding practice opportunities and/or time constant, to accurately link effects to the intervention itself.

Finally, while there is increased pressure to use efficient methods of instruction, there is a need for researchers to report on instructional efficiency at initial acquisition, maintenance, and generalization, as there may be differential effects of flashcard drill methods on short- and long-term memory. Additionally, researchers need to explore more precise ways of calculating instructional efficiency (e.g., by holding time constant across intervention conditions). Furthermore, there is need for investigations that manipulate study components to enhance instructional efficiency. For example, researchers need to consider innovative ways to increase opportunities for practice within a given session (e.g., this might mean presenting stimuli in an electronic format as has been done with letter sound knowledge; Volpe, Burns, DuBois, & Zaslofsky, 2011). Lastly, researchers should consider ways to modify intervention methods so
that they may be delivered at a group level, instead of at the individual level. An improvement of this nature will certainly be an efficient way to provide remediation and will maximize school resources to their fullest potential.
References


Psychology in the Schools, 47, 114-125.

practices in school psychology: A national survey of roles and referral problems.
Psychology in the Schools, 45, 483-498.

employees with moderate mental retardation in competitive jobs. Education and Training
in Mental Retardation and Developmental Disabilities, 35, 78-89.

Burns, M. K. (2004a). Age as a predictor of acquisition rates as measured by curriculum-based
assessment: Evidence of consistency with cognitive research. Assessment for effective
intervention, 29, 31-38.

Burns, M. K. (2004b). Empirical analysis of drill ratio research: Refining the instructional level

Burns, M. K. (2007a). Comparison of opportunities to respond within a drill model when
rehearsing sight words with a child with mental retardation. School Psychology
Quarterly, 22, 250-263.


(Reprinted from *The Reading Teacher*, 32, by S.J. Samuels, 1979.)


Tucker, J. A. (1988). *Basic flashcard technique when vocabulary is the goal*. Unpublished teaching material. University of Tennessee at Chattanooga, Chattanooga, TN.


## Appendix

### Table 1

*Summary of Study Sample Characteristics*

<table>
<thead>
<tr>
<th>Study</th>
<th>Intervention</th>
<th>No. of boys/girls</th>
<th>Mean age, years</th>
<th>Grade</th>
<th>Race/Ethnicity</th>
<th>Ed. Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burns (2007a)</td>
<td>IR</td>
<td>1/0</td>
<td>9.00</td>
<td>2\textsuperscript{nd}-3\textsuperscript{rd}</td>
<td>1 Caucasian</td>
<td>0/1</td>
</tr>
<tr>
<td>Burns &amp; Boice (2009)</td>
<td>IST, IR, &amp; TDP</td>
<td>12/8</td>
<td>NS</td>
<td>7\textsuperscript{th}-8\textsuperscript{th}</td>
<td>15 Caucasian, 5 African American</td>
<td>0/20</td>
</tr>
<tr>
<td>Burns &amp; Dean (2005a)</td>
<td>IR (manipulating ratio of K:U) &amp; TDP</td>
<td>3/2</td>
<td>NS</td>
<td>4\textsuperscript{th}</td>
<td>5 Caucasian</td>
<td>0/5</td>
</tr>
<tr>
<td>Burns et al. (2004)</td>
<td>IR</td>
<td>17/3</td>
<td>NS</td>
<td>3\textsuperscript{rd}-4\textsuperscript{th}</td>
<td>15 Caucasian, 5 African American</td>
<td>0/20</td>
</tr>
<tr>
<td>Burns &amp; Sterling-Turner (2010)</td>
<td>IR &amp; TDP</td>
<td>12/13</td>
<td>NS</td>
<td>4\textsuperscript{th}</td>
<td>15 Caucasian, 3 African American</td>
<td>20/0</td>
</tr>
</tbody>
</table>
## SIGHT WORD INTERVENTIONS

### Table 1 continued…

<table>
<thead>
<tr>
<th>Study</th>
<th>Intervention(s)</th>
<th>Baseline</th>
<th>Condition(s)</th>
<th>Age Group</th>
<th>Ethnicity</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joseph &amp; Nist (2006)</td>
<td>HPS, IST, &amp; TDP</td>
<td>3/0</td>
<td>NS</td>
<td>5th-6th</td>
<td>3 Caucasian</td>
<td>3/0</td>
</tr>
<tr>
<td>Joseph &amp; Schisler (2007)</td>
<td>IR, TDP, &amp; WB</td>
<td>NS</td>
<td>NS</td>
<td>1st-3rd</td>
<td>NS</td>
<td>32/28</td>
</tr>
<tr>
<td>Knight et al. (2003)</td>
<td>CTD &amp; IR</td>
<td>1/3</td>
<td>8.0</td>
<td>3rd</td>
<td>NS</td>
<td>0/4</td>
</tr>
<tr>
<td>MacQuarrie et al. (2002)</td>
<td>IST, IR, &amp; TDP</td>
<td>27/24</td>
<td>NS</td>
<td>3rd-7th</td>
<td>31 Caucasian, 16 African American, 3 Hispanic, 1 Asian</td>
<td>NS</td>
</tr>
<tr>
<td>Martin-Chang et al. (2007)</td>
<td>TDP &amp; Context</td>
<td>16/12</td>
<td>7.75</td>
<td>2nd</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Martin-Chang et al. (2007)</td>
<td>Training</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martin-Chang et al. (2007)</td>
<td>TDP &amp; Context</td>
<td>10/14</td>
<td>7.5</td>
<td>2nd</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td></td>
<td>Training</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 1 continued…

<table>
<thead>
<tr>
<th>Study</th>
<th>Intervention</th>
<th>Gain</th>
<th>Mean</th>
<th>Grade</th>
<th>Ethnicity</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matchett &amp; Burns (2009)</td>
<td>IR</td>
<td>0/1</td>
<td>10.00</td>
<td>3rd</td>
<td>1 Hispanic</td>
<td>1*/0</td>
</tr>
<tr>
<td>Nist &amp; Joseph (2008)</td>
<td>TDP, IR, &amp; IST</td>
<td>2/4</td>
<td>7.20</td>
<td>1st</td>
<td>6 Caucasian</td>
<td>6/0</td>
</tr>
<tr>
<td>Ruwe et al. (2011)</td>
<td>DI</td>
<td>3/0</td>
<td>13.7</td>
<td>7th-8th</td>
<td>NS</td>
<td>0/3</td>
</tr>
<tr>
<td>Szadokierski &amp; Burns (2008)</td>
<td>IR</td>
<td>13/14</td>
<td>10.00</td>
<td>4th</td>
<td>22 Caucasian, 1 African American, 1 Asian, 3 Multiracial</td>
<td>NS</td>
</tr>
<tr>
<td>Volpe et al. (2011)</td>
<td>IR &amp; TDP</td>
<td>1/3</td>
<td>6.00</td>
<td>1st</td>
<td>4 African American</td>
<td>4/0</td>
</tr>
</tbody>
</table>
Table 1 continued…

<table>
<thead>
<tr>
<th>Study</th>
<th>Intervention</th>
</tr>
</thead>
</table>

*Participant was an English Language Learner.

1 Study 1

2 Study 2

Note. CTD = Constant Time Delay; DI = Direct Instruction; IR = Incremental Rehearsal; IST = Interspersal Training; HPS = high-probability sequencing; K:U = Known: Unknown; NS = Not Specified; PT = Phase Training; REG ED = Regular Education; SPED = Special Education; TDP = Traditional Drill and Practice; WB = Word Boxes.
## Table 2

*Summary of Intervention Components*

<table>
<thead>
<tr>
<th>Article</th>
<th>Intervention</th>
<th>Study Design</th>
<th>Target Words Per Session</th>
<th>No. of Sessions</th>
<th>Time to Conduct Intervention</th>
<th>OTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burns (2007a)</td>
<td>IR</td>
<td>Single-subject (n = 1)</td>
<td>5</td>
<td>20</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>High OTR = 9 - 45; Mod OTR = 3-15</td>
</tr>
<tr>
<td>Burns &amp; Boice (2009)</td>
<td>IST, IR, &amp; TDP</td>
<td>Group (n = 20)</td>
<td>9</td>
<td>3</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NS; Varied by Condition and Student</td>
</tr>
<tr>
<td>Burns &amp; Dean (2005a)</td>
<td>IR (manipulating ratio of K:U) &amp; TDP</td>
<td>Single-subject (n = 5)</td>
<td>10</td>
<td>4</td>
<td>NS</td>
<td>TDP = 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IR(50% K) =3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IR (83% &amp; 90% K) = Varied</td>
</tr>
</tbody>
</table>
Table 2 continued…

<table>
<thead>
<tr>
<th>Study</th>
<th>Intervention</th>
<th>Group Type</th>
<th>Sample Size</th>
<th>Duration</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burns et al. (2004)</td>
<td>IR</td>
<td>Group</td>
<td>n = 20</td>
<td>≤ 11</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>NS (EC took ≈ 5-15 min longer than control)</td>
</tr>
<tr>
<td>Burns &amp; Sterling-Turner (2010)</td>
<td>IR &amp; TDP</td>
<td>Group</td>
<td>n = 25</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TDP = 3.52 min (SD = .73)</td>
<td>NS; Varied by Condition and Student IR = 8.90 min (SD = 1.49)</td>
</tr>
<tr>
<td>Joseph &amp; Nist (2006)</td>
<td>HPS, IST, &amp; TDP</td>
<td>Single-subject</td>
<td>n = 3</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Joseph &amp; Schisler (2007)</td>
<td>IR, TDP, &amp; WB</td>
<td>Group</td>
<td>n = 60</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Varied</td>
<td></td>
<td></td>
<td>NS</td>
</tr>
</tbody>
</table>
## SIGHT WORD INTERVENTIONS

Table 2 continued…

<table>
<thead>
<tr>
<th>Study authors</th>
<th>Intervention(s)</th>
<th>Design</th>
<th>Number of participants</th>
<th>Duration</th>
<th>CTD</th>
<th>IST</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knight et al. (2003)</td>
<td>CTD &amp; IR</td>
<td>Single-subject</td>
<td>n = 4</td>
<td>≈ 15 min</td>
<td>CTD = 30</td>
<td>IST = 22</td>
<td>unknown words, 36 known words</td>
</tr>
<tr>
<td>MacQuarrie et al. (2002)</td>
<td>IST, IR, &amp; TDP</td>
<td>Group</td>
<td>n = 51</td>
<td>NS</td>
<td>NS; Varied by Condition and Student</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martin-Chang et al. (2007)&lt;sup&gt;1&lt;/sup&gt;</td>
<td>TDP &amp; Context Training</td>
<td>Group</td>
<td>n = 28</td>
<td>NS</td>
<td>4 per condition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martin-Chang et al. (2007)&lt;sup&gt;2&lt;/sup&gt;</td>
<td>TDP, Context Training, &amp; Control</td>
<td>Group</td>
<td>n = 24</td>
<td>NS</td>
<td>4 per condition</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 2 continued…

<table>
<thead>
<tr>
<th>Study</th>
<th>Intervention(s)</th>
<th>Design</th>
<th>N</th>
<th>Word Count</th>
<th>Duration</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matchett &amp; Burns (2009)</td>
<td>IR</td>
<td>Single-subject</td>
<td>(n = 1)</td>
<td>Varied by Word set</td>
<td>5-9, ≈ 15 min</td>
<td>NS</td>
</tr>
<tr>
<td>Nist &amp; Joseph (2008)</td>
<td>TDP, IST, &amp; IR</td>
<td>Single-subject</td>
<td>(n = 6)</td>
<td>18</td>
<td>12</td>
<td>TDP ≈ 1.37 min, IST ≈ 2.49 min, IR ≈ 6.71 min, TDP = 9, IST = 9, IR = Varied</td>
</tr>
<tr>
<td>Ruwe et al. (2011)</td>
<td>DI</td>
<td>Single-subject</td>
<td>(n = 3)</td>
<td>Varied by Word Set &amp; Participant</td>
<td>≤ 16, ≈ 5 min</td>
<td>Varied by Student</td>
</tr>
<tr>
<td>Schmidgall &amp; Joseph (2007)</td>
<td>WB, IST, &amp; TDP</td>
<td>Single-subject</td>
<td>(n = 6)</td>
<td>18</td>
<td>20</td>
<td>NS, 6 per condition</td>
</tr>
</tbody>
</table>
Table 2 continued…

<table>
<thead>
<tr>
<th>Study</th>
<th>Condition</th>
<th>Group</th>
<th>IR (High/Low)</th>
<th>OTR and Ratio</th>
<th>Know:Unknown Words (K:U)</th>
<th>Time constant (TC)</th>
<th>Opportunities to Respond (OTR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Szadokierski</td>
<td>IR &amp; Burns</td>
<td>Group</td>
<td>9</td>
<td>4</td>
<td>15 – 30</td>
<td>Varied by condition</td>
<td>(n = 27)</td>
</tr>
<tr>
<td>&amp; Burns (2008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volpe et al.</td>
<td>IR &amp; TDP</td>
<td>Single-subject</td>
<td>12</td>
<td>12</td>
<td>3 min in TC, 5 in OTR-C, Varied in OTR-C, Varied in TC</td>
<td>(n = 4)</td>
<td></td>
</tr>
<tr>
<td>(2011)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. CTD = Constant Time Delay; DI = Direct Instruction; EC = Experimental Condition; HPS = high-probability sequencing; IST = Interspersal Training; IR = Incremental Rehearsal; K:U = Know:Unknown Words; OTR = Opportunities to Respond; OTR-C = Opportunities to Respond Constant; PT = Phase Training; TC = Time constant; TDP = Traditional Drill and Practice; WB = Word Boxes.

1Study 1
2Study 2
A Comparison of Two Sight Word Interventions: Traditional Drill and WordSheets

Christina Marie Mulé, C.A.G.S., NCSP

Northeastern University
Abstract

Traditional drill and practice (TDP) is a sight word intervention that is well supported in the literature as being both effective and efficient. However, with growing demands in school systems, there is increased pressure to employ interventions that enhance learning outcomes with less instructional time. WordSheets (WS) was created as a method that could potentially (a) produce better learning outcomes by increasing opportunities to respond, (b) produce better generalization than traditional approaches that present one word at a time by presenting stimuli in connected text format, and (c) be more convenient for educators by eliminating the need to manipulate flashcards. Utilizing a group design, the current study sought to compare the instructional effectiveness and efficiency of TDP and WS with a first-grade sample of students ($n = 27$) identified by their teachers as struggling readers. Results indicated that, although WS led to significantly more opportunities to respond, it was both less effective and efficient than TDP with regard to both retention and maintenance outcomes. However, TDP was no more effective and efficient than WS with regard to the generalization of treatment gains. Limitations and implications of the current study for school-based professionals are discussed.

*Keywords:* Reading intervention; Sight word intervention; Sight word learning
A Comparison of Two Sight Word Interventions: Traditional Drill and WordSheets

Reading is an imperative skill in today’s society that is applied in numerous aspects of our lives (Daly, Chafouleas, & Skinner, 2005). Better reading skills are linked to overall academic success, as well as one’s ability to become a productive member of society (Adams, 1990; Good, Simmons, & Kaméenui, 2001). Poor reading skills often have debilitating effects on academic and social development (Rathvon, 2008). Unfortunately, a significant number of children in the United States struggle with reading (i.e., approximately, 38% of fourth graders and 29% of eighth graders are reading below basic levels; NCES, 2009). In fact, difficulties in reading are the most frequent cause of teacher referrals to school psychologists (Bahr et al., 2006; Bramlett, Murphy, Johnson, Wallingsford, & Hall, 2002).

It is not surprising that reading challenges are so prevalent, as it is a rather complicated skill that encompasses a wide variety of prerequisite skills (Daly et al., 2005). For example, necessary achievements in reading include: (a) phonemic awareness (i.e., specific skill that requires attention to and manipulation of individual sounds of spoken words), (b) alphabetic principle (i.e., knowing that there is an association between letters and sounds or that there are letter-sound relations), (c) reading fluency (i.e., the rate at which text is translated into spoken language), (d) vocabulary (i.e., knowing the meanings of words and concepts), and (e) reading comprehension (i.e., the comprehension or understanding of text; Joseph, 2006; Joseph, 2008; U.S. Department of Education, 1999). One skill that many American children struggle with is the ability to recognize words accurately and fluidly without hesitation (Daley et al., 2005; Joseph, 2006; Pikulski & Chard, 2005; Pinnell et al., 1995). This skill is commonly known as reading fluency (Samuels, 2002) and is the bridge between decoding and reading comprehension.
SIGHT WORD INTERVENTIONS

(Pikulski & Chard, 2005). For instance, a fluent reader is able to effortlessly and automatically decode text, allocating mental resources to comprehension (Joseph, 2006; Samuels, 2002). However, a non-fluent reader may laboriously decode text (e.g., 20 words per minute), utilizing valuable mental resources, sparing very little for comprehension (Daley et al., 2005; Samuels, 2002).

The skill or ability to accurately and automatically read words is critical as it has been demonstrated to increase reading fluency and comprehension (Tan & Nicholson, 1997). Given that reading fluency is directly related to a child’s ability to derive meaning from text, several remediation strategies have been developed to enhance this skill (e.g., repeated readings, phase drill, and listening while reading; Joseph, 2008). One such strategy is the flashcard drill method, which is used to supplement developmental reading programs (e.g., phonemic awareness and phonics). Flashcard drill methods, while sometimes depicted as monotonous, allow for high repetition or practice of unknown words, which helps children to read words accurately and quickly. Several flashcard drill techniques exist for teaching children to read whole words or sight words (i.e., words that a child is able to read accurately and automatically; Volpe, Mulé, Briesch, Joseph, & Burns, 2011), which vary both in terms of complexity and the time required for implementation.

Two of the most well-supported sight word interventions within the literature have been Incremental Rehearsal (IR) and Traditional Drill and Practice (TDP). IR is an interspersal flashcard drill method that incrementally introduces unknown words to students (e.g., 1U, 1K, 1U, 1K, 2K, 1U, 1K, 2K, 3K, etc., where 1U is the first unknown word, and 1K is the first known word); whereas TDP is a non-interspersal flashcard drill technique in which 100% of the words presented to the student are unknown (e.g., 1U, 2U, 3U, 4U, 5U, [shuffle], 4U, 2U, 1U,
5U, 3U, [shuffle], etc, where U is an unknown word). The procedural differences between these two instructional conditions (i.e., the ratio of unknown stimuli to known stimuli) have resulted in notable contrasts in their rates of *effectiveness* (i.e., the degree to which children can acquire the skills that the intervention purports to address) and *efficiency* (i.e., the rate at which students achieve the skills as a function of instructional time; Daly et al., 2005; Joseph & Schisler, 2007).

IR has been repeatedly cited in the literature as being a highly effective method for targeting sight word recognition as it allows children repeated practice of unknown words (Burns & Boice, 2009; Burns & Sterling-Turner, 2010; MacQuarrie, Tucker, Burns, & Hartman, 2002; Nist & Joseph, 2008). However, while it has been successful in helping children learn words, it has fairly consistently been noted to be the least efficient flashcard drill method for initial acquisition or next day retention (Burns & Sterling-Turner, 2010; Joseph & Schisler, 2007; Nist & Joseph, 2008; Volpe et al., 2011). These findings are not surprising because while IR provides increased opportunities to practice unknown words (Greenwood, Delquadri, & Hall, 1984), the procedure itself takes much longer to implement than non-interspersal methods (e.g., TDP) as it dedicates instructional time to rehearsing known words (e.g., IR takes approximately 5-15 minutes to implement, whereas TDP takes only 1-4 minutes; Burns & Sterling-Turner, 2010; Matchett & Burns, 2009; Nist & Joseph, 2008).

Interestingly, when opportunities to respond or to practice target words have been allowed to vary, TDP is less favorable than IR in terms of instructional effectiveness (Burns & Boice, 2009; Burns & Sterling-Turner, 2010; MacQuarrie et al., 2002; Nist & Joseph, 2008). However, when opportunities to respond were held constant, TDP has been shown to be just as effective as IR (Volpe et al., 2011). Additionally, it has been consistently cited as the most efficient flashcard drill method for teaching sight word recognition (Burns & Sterling-Turner,
Because of the sheer number of children with reading problems, school psychologists and educators are challenged to find and implement techniques that are both effective and efficient. The contrast between IR and TDP's instructional effectiveness and efficiency has left school psychologists and educators with a dilemma. Although school psychology practice standards (NASP, 2008) require that practitioners employ effective strategies in order to help students succeed academically, effectiveness alone no longer seems sufficient. Specifically, because of increased demands on educators' and school psychologists' time, there is a need to implement interventions that are both effective and efficient (i.e., given comparable effectiveness, the more efficient method should be implemented; Sterrett & Imig, 2011; Walker, 2004). Appropriately, increased attention has been directed toward making already effective interventions more efficient and further, developing interventions that satisfy the need for both effectiveness and efficiency.

Utilizing interventions that are both effective and efficient will allow more children to receive intervention services. Additionally, because learning difficulties are often characterized as a problem with the rate at which children learn, utilizing efficient techniques will allow struggling readers to catch up to their peers at a faster rate (Skinner, 2008). To this point flashcard drill methods have been used because they are effective and more efficient than other methods (e.g., phonics analysis); however, it is possible for new interventions to build upon the instructional effectiveness and efficiency observed in current flashcard drill methods.

WordSheets (WS) was devised as a method that could perhaps exceed the instructional effectiveness and efficiency observed in flashcard drill techniques (Volpe, 2011a). WS is a sight
SIGHT WORD INTERVENTIONS

word intervention that closely resembles the instructional procedures of TDP. For example, (a) only unknown words are presented to the student; (b) the interventionist models the unknown words and then asks the student to read them independently; (c) target words are presented in random order; and (d) error correction is provided for any miscues and hesitations. However, there is one significant distinction in the presentation; that is, instead of using flashcards which can be cumbersome and time consuming to manipulate, WS is delivered on a piece of 8 x 11 paper (see Figure 1). As a result of the intervention format, the time that is required to cycle through flashcards in TDP is eliminated; therefore, theoretically leaving more time for students to practice target words.

In addition to the enhanced number of opportunities to practice, it is also hypothesized that there are other related advantages to WS. For example, WS presents stimuli in a connected text format, which may result in better generalization to authentic reading conditions (e.g., reading a book). Additionally, WS is student paced, which may accelerate the pace of instruction, or at least pace instruction at a rate that the student is comfortable with. Finally, WS stimuli are computer generated and are thought to be a more convenient intervention for educators to prepare (see Figure 2 for WS Generator). Although both WS and flashcard drill methods require some level of individualization, targeted words in WS can simply be selected through a computer software program that randomizes the order of the stimuli. Conversely, flashcard drill methods require the interventionist to spend a considerable amount of time preparing materials and also manipulating them during intervention.

The goal of this study was to compare the instructional effectiveness and efficiency of TDP and WS while holding time constant and allowing opportunities to respond to vary. Time was held constant instead of holding opportunities to respond constant because educators
typically have a finite number of minutes in which they can implement an intervention (i.e.,
interventionists would rather provide intervention for 5 minutes rather than count the number of
opportunities to respond that a child is given). Additionally, by holding time constant
instructional efficiency will be provided in directly translatable terms (i.e., the effectiveness
measure will also be the efficiency measure). Specifically, the current study sought to answer the
following research questions:

1. Which procedure (i.e., TDP, WS) affords greater opportunities to respond?
2. Which procedure (i.e., TDP, WS) is associated with greater cumulative number of
   words read correctly on next day retention probes?
3. Which procedure (i.e., TDP, WS) is associated with greater words maintained or
   read correctly at maintenance assessment?
4. Which procedure (i.e., TDP, WS) is associated with greater words generalized or
   read correctly at generalization assessment?
5. Which is the most efficient procedure (i.e., TDP, WS) in regards to next-day
   retention, maintenance, and generalization data?

Methods

Participants and Setting
Participants consisted of 27 first grade students ($M$ age = 6.56 years; Range = 6 to 8
years) who were referred by their classroom teachers for reading difficulties. Students attended
eight different public elementary schools in suburban and urban school districts in the
Northeastern United States. Forty-eight percent ($n = 13$) of the sample were male and 52% ($n =
14$) were female. Additionally, 40.7% of the participants were Caucasian, 25.9% African
American, 29.6% Hispanic, and 3.7% of the students were of Asian descent. Approximately,
one-fifth (18.5%) of the sample were English Language Learners (ELL) and 3.7% were receiving special education services.

Sixteen school psychology graduate students enrolled in an Academic Interventions course served as the interventionists for the study and were responsible for administering the interventions under investigation (described in detail below). Graduate students worked in pairs; thus, one graduate student served as the interventionist while the other observed implementation and served as the procedural integrity and inter-observer agreement (IOA) data collector. Data collectors minimally interacted with the students (e.g., they were advised to sit behind the student so that they were not in view during the delivery of intervention) and the students seemed to be comfortable with them observing their intervention session. Training of interventionists and data collectors consisted of one full class lecture, involving didactic instruction and role play activities. Interventionists were trained to criterion (i.e., the interventionists demonstrated ability to carry out all items indicated on the procedural integrity checklists).

The intervention sessions took place in a semi-private space outside of each student’s classroom in which the interventionist, data collector, and student were the only individuals present. During the intervention, the interventionists and students sat at a right angle from one another at a large table, while the data collector sat behind the student.

**Materials**

Prior to the start of intervention, a pre-assessment was conducted to identify specific target words for instruction. Specifically, known and unknown words were identified through an assessment of 379 high frequency words. Words were printed with a landscape orientation in black ink on white 3” x 4” index cards, shuffled, and presented to the student one at a time. Students were asked to read each word aloud. If a student did not respond within 3 s or
mispronounced the word, the word was determined to be “unknown” and placed in a corresponding unknown pile. Unknown words did not receive corrective feedback. If the student correctly pronounced the word within 3 s, the word was determined to be “known,” and was removed from the possible pool of target words and subsequently placed in a known pile. Verbal praise was given for all words read correctly and within 3 s, as well as for effort. The full assessment was conducted on a second occasion. Words that were discrepant between administrations (e.g., read correctly on Trial 1 but incorrectly on Trial 2 or vice versa) were omitted from the target list of words. Words that were unknown on both assessments were randomly assigned to one of the two instructional conditions: WS and TDP. Six unknown words were randomly assigned to each instructional condition; consequently a total of 12 unknown words were targeted per intervention session. The six target words per condition was decided upon because of ceiling effect concerns (i.e., student might have learned all words within each condition if fewer words were targeted; thus making it difficult to make comparisons between conditions).

Words that were targeted in the WS condition were printed on a single page with a portrait orientation (see Figure 1). Words targeted in the TDP condition were taught using the preassessment index cards mentioned above (i.e., printed in black ink on 3 x 4 index cards with landscape orientation). The index cards were also used for all retention and maintenance assessments.

**Measures**

**Retention.** A next day retention probe including all of the unknown words instructed in the previous session (i.e., 6 unknown words per instructional condition, totaling 12 unknown words across conditions) was administered the day following intervention and served as a
measure of instructional effectiveness. Target words from both the WS and TDP conditions were individually presented in random order with the flashcards described above. Students were prompted by the interventionist to read each word aloud. If the student accurately read the target word within 3 s, the word was considered learned. If the student mispronounced the word or hesitated for 3 s or more, the word was considered unknown.

**Maintenance.** A maintenance probe was administered 1 week after the intervention was completed to determine how many of the learned words (i.e., only words read correctly on the retention probes) were maintained over time (a maximum of 36 words could have been assessed). Flashcards representing words read correctly on the retention probes were presented individually to the students in random order. The students were asked to read each word aloud. Words were recorded as being correct if students were able to pronounce them accurately within 3 s. Words were recorded as incorrect if students mispronounced them or hesitated for 3 s or longer. Students were praised for their effort, and no corrective feedback was provided for any miscues or hesitations.

**Generalization.** Following the maintenance probe, a generalization probe was also administered to assess whether the students could read the targeted words correctly when presented in sentences. Previously constructed sentences from Volpe et al (2011) were used, and included one learned word (based on retention data), as well as four to six monosyllabic words. These sentences were individually presented to the students on an 8 x 11 piece of paper with portrait orientation, and students were prompted to read each sentence aloud. Responses were recorded as correct if the student accurately read the target word in the context of the sentence. Conversely, responses were recorded as incorrect if the student was not able to accurately
pronounce the target word or if the student hesitated for 3 s or longer (Note. Errors made on non-target words [i.e., distracter words] were not factored into this scoring).

**Instructional effectiveness and efficiency.** Instructional effectiveness was defined as the cumulative number of words read accurately on the retention, maintenance, and generalization assessments; whereas instructional efficiency was defined as the cumulative rate of words learned. The learning rates were calculated by multiplying the number of words read accurately (WRA) on the next day retention assessment by 60 s and then dividing the instructional time in seconds spent under each respective condition (WRA x 60 s / Instructional Time [s]). However, in this study instructional time was held constant across intervention conditions at 120 s. The same calculation was applied maintenance and generalization data to determine their efficiency.

**Interobserver agreement.** Interobserver agreement (IOA) data were collected for 50% of the preassessment trials, 33% of the next-day retention assessments, and 100% of the maintenance and generalization assessments. Agreement was examined by aggregating the total number agreements on student correct responses and errors, dividing by the total number of words assessed, and multiplying by 100. Overall, agreement was high for preassessment, retention, maintenance and generalization probes (99.5%, 99.1%, 97.1% and 96.8% respectively). Corresponding kappa values for preassessment, retention, maintenance and generalization were .99, .98, .94, and .92, respectively.

**Procedural integrity.** Procedural integrity was recorded by data collectors for the intervention conditions (i.e., WS and TDP), and retention, maintenance, and generalization assessments. The WS (see Figure 2), TDP (see Figure 3), retention (see Figure 4), maintenance (see Figure 5), and generalization (see Figure 6) checklists consisted of nine, eight, four, four, and two items, respectively. Data collectors placed a checkmark under the “yes” column if the
interventionist implemented a procedural step as indicated on the checklist or placed a checkmark under the “no” column if the interventionist did not implement a procedural step as indicated on the checklist. Procedural integrity was recorded for 33% of the intervention sessions and retention assessments, and for 100% of the maintenance and generalization assessments. According to these data, 100% of procedural steps were followed.

**Procedures**

Students received a total of three intervention sessions that occurred on three consecutive school days. On the first day, only the intervention was delivered. On the second and third days, retention assessments were delivered prior to intervention, and on the fourth day, the next day retention assessment was the only activity (i.e., Day 1 = intervention; Day 2 = retention and intervention, Day 3 = retention and intervention, and Day 4 = retention). One week following intervention, interventionists collected maintenance and generalization data for words that the student read correctly across all retention assessments.

The order of the intervention conditions were counterbalanced across sessions and students (e.g., Student 1 received TDP and then WS for session 1, and WS and then TDP for session 2; whereas Student 2 received WS and then TDP for session 1, and TDP and then WS for session 2). Before the intervention timing began, the interventionist trained the student on each target word. The interventionist then used a stopwatch to record the length of each session; timing began with the presentation of the first flashcard for the TDP condition or the first word read in the WS condition. The interventionist discontinued timing after 2 min elapsed for each instructional condition. In both conditions, students intermittently received verbal praise (e.g., “great work” or “good job”) for their efforts and received corrective feedback for any miscues
SIGHT WORD INTERVENTIONS

(i.e., words read incorrectly or hesitations 3 s or longer). The following sections provide a description of each intervention condition.

**WordSheets.** WS were generated using a computer program designed in LiveCode and developed by Volpe (2011a). The WS generator contained all 379 high frequency words that were used in the preassessment. High frequency words were programmed in drop-down menus within the WS generator program (see Figure 2). The primary researcher selected the randomly assigned target words intended for each session and then pressed the "First Grade Sheet" button on the WS generator screen. Subsequently, the program produced the stimuli seen in Figure 1. Once stimuli were generated the primary researcher provided them to each respective team of graduate students in the Academic Interventions course.

Six target words for each intervention session were printed on a single page with a portrait orientation (in total 18 words were targeted across three sessions). Target words, printed in 20pt Times New Roman font with three spaces in between each word, were displayed at the top of the page and received training. Interventionists referred to a script for the training phase. For example, the interventionist said, “[Student name] I am going to read some words from this WordSheet. Listen carefully as I read each word because I am going to ask you to read them after me. This word is ‘fed.’ What word?” Subsequently, the student read the unknown word aloud. If the unknown word was read correctly the student received verbal praise. If the student mispronounced the unknown word or hesitated for 3 s or longer, the training of that word was repeated. After the student was able to accurately read the first target word, training occurred in this same fashion for the remaining five target words.

After training occurred the student was instructed to move to the first WS phase. Stimuli in the first WS phase were presented in 16 pt Times New Roman font in random order, with 1.5
line spacing (six words per line), and three spaces in between each word (see Figure 1). Again, the interventionists used a script to deliver directions (i.e., “Now we are going to go through the words again, only this time you will read them by yourself. Put your finger under the first word and start reading when I say begin.”). Verbal praise was intermittently provided for effort. Corrective feedback was provided to students if they did not respond to the target word within 3 s and for any words mispronounced. When miscues occurred, the interventionist placed a ( / ) through the word to record errors. This procedure continued until 1 min elapsed. After 1 min elapsed, the interventionist placed a ( ] ) around the last word practiced in order to track opportunities to respond.

Once students completed the first training phase of the WS intervention, they were instructed to move into the second phase (stimuli were presented in the same manner as in the first phase; however, in the second phase there were 12 words per line). Interventionists continued to use a script and said, “Okay, now I want to see how many words you can read in 1 minute. Put your finger under the first word and start when I say begin.” Again, verbal praise was intermittently provided for effort and corrective feedback was provided for errors and hesitations. In this final phase, interventionists recorded errors and opportunities to respond using the same procedures described above.

**Traditional drill and practice.** In each session of TDP, six target words were presented on flashcards to the student in a drill format (in total 18 words were targeted across three sessions). First, the interventionist modeled the correct pronunciation of the target word and then asked to the student to repeat the word before proceeding to the next word. Mispronounced words and hesitations lasting 3 s or longer received corrective feedback until the student was able to read the word correctly. Once the student received training on all target words, the words
were presented in the prescribed TDP method: 1U, 2U, 3U, 4U, 5U, 6U. For each subsequent trial, the interventionist shuffled the cards and presented the words in random order. The interventionist provided intermittent praise for effort and corrective feedback for miscues and hesitations. The procedures above were repeated until 2 min elapsed. While delivering intervention the interventionists recorded errors by placing a tally mark on the TDP record form. Opportunities to respond were recorded by placing a tally mark on the record form every time the student was able to complete one trial of the flashcard stack. If 2 min elapsed before a trial was completed, the interventionist recorded the number of practiced words over six (e.g., if a student practiced two of the six target words when 2 min elapsed the interventionist recorded "2/6").

Results

Statistical procedures were conducted using IBM SPSS Statistics 21. Retention (time 1 - time 3), maintenance, generalization, opportunities to respond, and errors were initially analyzed to determine whether the assumptions for parametric techniques (i.e. normal distributions and homogeneity of variance) had been met. Visual inspection of histograms for each dependent variable supported the assumption of normality. Descriptive data are presented in Table 1 and include estimates of skew and kurtosis. In general, these indicators were within three standard errors. Thus, the data were analyzed parametrically. Finally, within this study, an alpha level of .05 was employed for all tests of statistical significance.

Research Question 1: Which procedure affords the most opportunities to respond?

To address the first question a dependent-samples t-test was conducted to examine difference in opportunities to respond across conditions. Results indicated that WS \( (M = 153.22, SD = 53.17) \) afforded significantly more opportunities to respond than did TDP \( (M = 85.15, SD = \)
Research Question 2: Which procedure led to greater cumulative number of words read correctly on next day retention probes?

To address the second research question a 2 x 3 within-subjects model of analysis of variance (ANOVA) with one between factor, with two levels (i.e., intervention condition [WS and TDP]), and three within factors (Time 1, Time 2, and Time 3) was conducted to assess the differential impact of the two interventions on cumulative next day retention across three time periods. Results of the ANOVA indicated a significant main effect for condition, Wilks Lambda = .78, $F(1, 26) = 7.50, p < .05$, $\eta^2_p = .22$, observed power = .75, suggesting that TDP was significantly more effective than WS (see Table 2 and Figure 3). Additionally, there was a significant main effect for time, Wilks Lambda = .15, $F(2, 52) = 71.57, p < .001$, $\eta^2_p = .85$, observed power = 1.00, with both groups showing a significant increase in the number of retained words over time. Cumulatively, the 27 participants within this study learned 285 words in the TDP condition and 242 words in the WS condition; on average, individual participants learned 10.56 words ($SD = 3.83$) in the TDP condition and 8.96 words ($SD = 4.71$) in the WS condition. Finally, the interaction between type of intervention and time was not significant, Wilks Lambda = .91, $F(2, 52) = .57, p = .29$, $\eta^2_p = .09$, observed power = .25, indicating that there was not a statistically significant difference in the intervention effects by time.

Research Question 3: Which procedure led to greater words maintained or read correctly at maintenance assessment?

The third research question also addressed the effectiveness of the two instructional conditions. Maintenance data were compared across conditions to determine if significant
differences existed between TDP and WS with regard to the maintenance of intervention effects after one week of no treatment. A dependent-samples t-test was conducted to compare the number of words maintained across the two instructional conditions. Significantly more words targeted in the TDP condition \((M = 7.96, SD = 4.17)\) were maintained as compared to those targeted in the WS condition \((M = 6.33, SD = 4.16)\), \(t (26) = 2.65, p < .05\) (two-tailed). The magnitude of the differences in means (mean difference = 1.63, 95% CI: 0.36 to 2.90) was moderate to large \((r^2 = .21)\), suggesting that TDP was significantly more effective than WS with respect to the maintenance of learned words (see Table 1).

**Research Question 4: Which procedure led to greater words generalized or read correctly at generalization assessment?**

A dependent samples t-test was conducted to investigate differences in the number of words generalized across instructional conditions. Differences in generalization across conditions was minor \((TDP M = 7.44, SD = 4.16; WS M = 6.67, SD = 4.21)\), \(t (26) = 1.53, p = .14\) (two-tailed). Indeed, the magnitude of the difference in the means (mean difference = .78, 95% CI: -0.27 to 1.83) was negligible \((r^2 = .08)\), suggesting no difference in the effectiveness of the two intervention conditions with respect to helping participants generalize learned words (see Table 1).

**Research Question 5: Which is the most efficient procedure in regard to next-day retention, maintenance, and generalization?**

The fifth research question concerned the efficiency of the two intervention conditions based on total instructional time taken to learn the target words. Because instructional time was held constant across the two instructional conditions the statistical techniques used to explore effectiveness (using the retention, maintenance, and generalization assessment data) were used
simultaneously to investigate efficiency (number of learned words/2 minutes). With respect to retention, TDP was significantly more efficient than WS, Wilks Lambda = .78, $F (1, 26) = 7.50, p < .05, \eta_p^2 = .22$, observed power = .75. Likewise, in terms of maintenance, TDP was the more efficient method as compared to WS, $t (26) = 2.65, p < .05$ (two-tailed), $r^2 = .21$. Finally, TDP was no more efficient than WS when evaluating generalization assessment data, $t (26) = 1.53, p = .14$ (two-tailed), $r^2 = .08$ (see Table 1).

On average, participants retained 1.76 words per minute in the TDP condition and 1.49 words per minute in the WS condition (see Table 3). Participants maintained an average of 1.33 words per minute of instruction in the TDP condition and 1.06 words in the WS condition. Finally, participants were able to generalize an average of 1.24 words per minute of instruction in the TDP condition, whereas in the WS condition they were able to generalize an average of 1.11 words.

**Exploratory Analysis: Considering opportunities to respond, were significantly more errors committed in one instructional condition than in the other?**

A dependent-samples $t$-test was conducted to compare the number of errors committed across instructional conditions, as increased errors might suggest compromised attention or effort. Significantly fewer errors were committed in the TDP condition ($M = 14.44, SD = 9.61$) compared to the WS condition ($M = 25.52, SD = 10.52$), $t (26) = -5.35, p < .001$ (two-tailed), $r^2 = .52$ (see Table 1). Because of this difference, the WS error data were further studied. An additional dependent-samples $t$-test was conducted to compare the errors committed in Phase 1 and Phase 2 of the WS intervention (recall that in Phase 1 participants received initial learning trials and in Phase 2 they were given an opportunity to enhance fluency of sight word recognition). Results indicated that significantly more errors were committed in Phase 1 ($M =$
14.67, \(SD = 4.89\) than in Phase 2 \(M = 11.30, SD = 6.85\) of the WS condition, \(t (26) = 3.49, p < .05\) (two-tailed), \(r^2 = .32\). Interestingly this was the case even when Phase 2 \(M = 91.11, SD = 29.24\) afforded significantly more opportunities to respond than did Phase 1 \(M = 62.70, SD = 25.66\), \(t (26) = -9.48, p < .001\) (two-tailed), \(r^2 = .77\). Thus, participants read faster in Phase 2, but committed fewer errors.

**Discussion**

The goal of this study was to extend the current research on sight word interventions by comparing the instructional effectiveness and efficiency of two interventions. Specifically, this study sought to draw comparisons between one well-supported sight word intervention (i.e., TDP) and one novel sight word intervention that was hypothesized to lead to superior effectiveness and efficiency outcomes (i.e., WS) due to enhanced opportunities to respond and the connected text format of the stimuli. As expected, the WS condition afforded significantly more opportunities to practice target words than did the TDP condition. In fact, students received almost twice as many opportunities to respond in the WS condition \(M = 153.22; SD = 53.17\) than in the TDP condition \(M = 85.15; SD = 17.62\). These data suggest that students had approximately 25 and 14 opportunities to respond to each target word in the WS and TDP conditions, respectively.

With respect to effectiveness, both TDP and WS were effective interventions for helping students increase their overall sight word knowledge. Cumulatively, over three TDP intervention sessions, the 27 students retained 285 words \(M = 10.56; SD = 3.83\), maintained 215 words \(M = 7.96; SD = 4.16\), and generalized 201 words \(M = 7.44; SD = 4.16\). Similarly, in the WS condition students cumulatively retained 242 words \(M = 8.96; SD = 4.71\), maintained 171 words \(M = 6.33; SD = 4.16\), and generalized 180 words \(M = 6.66; SD = 4.21\). Given
numerous methodological differences between the current study and prior investigations (e.g., the types of interventions evaluated, the number of intervention sessions delivered, the number of words targeted per session and by condition), it is difficult to make direct comparisons with respect to study findings as they relate to effectiveness. However, current effectiveness results for both TDP and WS fall within or just outside of the range of effectiveness reported for other sight word drill methods. For example, across six studies investigating the effects of TDP, the average number of words retained by each participant ranged from 1.40 - 28.70; the average number of words maintained ranged from 4.00 - 52.00; and the average number of words generalized ranged from 4.50 - 12.60 (Burns & Sterling-Turner, 2010; Joseph & Nist, 2006; Joseph & Schisler, 2007; Nist & Joseph, 2008; Schmidgall & Joseph, 2007; Volpe et al., 2011). Likewise, across five studies investigating the effects of IR, the average number of words retained ranged from 1.19 - 28.20; the average number of words maintained ranged from 3.50 - 22.83; and the average number of words generalized ranged from 3.50 - 23.50 (Burns & Sterling-Turner, 2010; Joseph & Schisler, 2007; Nist & Joseph, 2008; Szadokierski & Burns, 2008; Volpe et al., 2011). Therefore, effectiveness data from the current study generally reflects those from other sight word interventions in prior studies. This is particularly noteworthy for WS, as this is the first study to evaluate its effectiveness.

Although WS effectiveness was similar to prior works, in this study it was less effective than TDP when retention and maintenance data were examined. Given the numerous additional opportunities afforded by the WS condition, this finding was unexpected and stands in contrast to the extant literature (e.g., Belfiore, Skinner, & Ferkis, 1995; Daly, Hintze, & Hamler, 2000; Greenwood, Delquadri, & Hall, 1984; MacQuarrie et al., 2002; Szadokierski & Burns, 2008).

---

3Reported effectiveness ranges for TDP and IR are large because studies used a varied number of intervention sessions and targeted a varied number of words per session.
which suggests that increased opportunities to respond have a significant and positive effect on learning (MacQuarrie et al., 2002; Szadokierski & Burns, 2008). Further unexpected were the generalization results, perhaps one of the most important outcomes of intervention, which revealed no significant differences across intervention conditions. This finding fails to make a consistent case for TDP over WS and is particularly interesting, as interventions with superior rates of retention and maintenance have typically resulted in superior rates of generalization (Nist & Joseph, 2008; Volpe et al., 2011). However, it may be that because the WS stimuli were formatted in a connected-text format, and were more properly matched or aligned with the generalization assessment than was TDP, that gains may have been more observable (Shapiro, 2004).

In terms of efficiency, both TDP and WS were efficient in helping students learn. In the TDP condition, students respectively retained, maintained, and generalized an average of 1.76, 1.33, and 1.24 words per instructional minute; and in the WS condition students respectively retained, maintained, and generalized an average of 1.49, 1.06, and 1.11 words per instructional minute. Efficiency results from the current study are consistent with, and in some cases exceed, the rate of learning supported by other well-supported sight word interventions in the literature (e.g., TDP and IR). For example, the same studies mentioned above evaluated TDP's efficiency and supported a range of 0.34 - 2.79 words retained per instructional minute; 0.21 - 2.38 words maintained per instructional minute; and 0.19 - 0.90 words generalized per instructional minute (Burns & Sterling-Turner, 2010; Joseph & Nist, 2006; Joseph & Schisler, 2007; Nist & Joseph, 2008; Schmidgall & Joseph, 2007; Volpe et al., 2011). Additionally, studies evaluating IR's efficiency report retention, maintenance, and generalization rates ranging from 0.31 - 0.69, 0.15 -

---

4 In previous studies comparing opportunities to respond, the nature of the stimuli were consistent (i.e., flashcard drill methods were compared to one another) and the WS stimuli vary considerably from these. Therefore, the effects of opportunities to respond may not be directly comparable.
0.46, and 0.15 - 0.21, respectively (Burns & Sterling-Turner, 2010; Joseph & Schisler, 2007; Knight et al., 2003; Nist & Joseph, 2008; Volpe et al., 2011). Worthy of note, is that rates of learning in the WS condition exceeded those of IR and were within a similar range as TDP. This is particularly impressive and suggests that WS may be a competitive alternative to other sight word drill methods.

Although efficiency rates for WS were strong in relation to the sight word literature, efficiency rates were significantly better for TDP than for WS when retention and maintenance data were assessed. This was an unexpected finding given that the length of instructional time was held constant across conditions and opportunities to respond were greater in the WS condition. Thus, WS was expected to lead to greater gains in sight word recognition, and to superior rates of learning. Although results differ from expected findings, previous studies have consistently cited TDP to be a leading flashcard drill intervention with respect to efficiency (e.g., Cates, Skinner, Watson, Meadows, Weaver, & Jackson, 2003; Burns & Sterling-Turner, 2010; Joseph & Nist, 2006; Nist & Joseph, 2008; Schmidgall & Joseph, 2007; Volpe et al., 2011); therefore, we cannot say that this study finding is entirely surprising. Finally, since there were no statistically significant differences between TDP and WS when generalization data were examined for effectiveness, there were also no differences in efficiency.

Although effectiveness and efficiency outcomes from the current study were generally positive and closely aligned with the extant literature, as mentioned above there were several unexpected findings. Given this, exploratory analyses examining error frequency data were conducted with the hope that they might help to explain why the additional opportunities to respond afforded by the WS intervention were not associated with enhanced learning outcomes and rates of learning. Results from these analyses indicated that students made significantly more
errors in the WS condition than they did in the TDP condition. Also interesting and worthy of note, despite having significantly more opportunities to practice in Phase 2 of the WS condition, participants committed significantly fewer errors in this phase than in Phase 1. This finding may suggest that Phase 1 afforded more initial learning trials and that students benefited from the error correction they received in this phase; thus, committing fewer errors in Phase 2. Likely, significantly more errors were committed in the first half of the TDP delivery as well; however, while TDP error data were collected, they were not collected with this level of specificity. The error frequency data may also support greater initial adjustments to the presentation of stimuli in Phase 1 (i.e., students may have had more difficulty attending to the stimuli in Phase 1 because of the intervention's novelty but then may have adapted by Phase 2 of the intervention).

Additionally, it may support the idea of moving directly to a brief fluency trial immediately after new facts are acquired. Results further suggest that despite the increased pressure to read as many words as they could in Phase 2 of the WS condition, participants appear not to have traded accuracy for speed and were able to build upon their sight word recognition fluency.

There are several possible explanations as to why participants did not realize the benefits in learning that should have been afforded by the additional opportunities to respond in the WS condition. First, it is possible that TDP provided an adequate number of opportunities to respond and that additional opportunities for practice beyond that had no added benefit. However, the number of opportunities to respond afforded by WS is comparable to the number of opportunities to respond afforded by IR (i.e., with a ratio of 9 knowns: 1 unknown, one target word could be practiced up to 26 times in IR), which has consistently been linked to superior effectiveness. Furthermore, this hypothesis stands in contrast to the extant literature mentioned above, which suggests that opportunities for practice is the main instructional factor that affects retention of
target words (MacQuarrie et al., 2002; Szadokierski & Burns, 2008). Consequently, while possible, this hypothesis does not seem likely.

As opposed to the quantity, an alternative explanation for why opportunities to respond did not lead to enhanced effectiveness and efficiency outcomes relates to the quality of those opportunities to respond. As mentioned above, WS afforded nearly double the amount of opportunities to respond than those afforded by TDP, and in the same amount of instructional time (i.e., 2 minutes). On average, participants read 1.28 words per second in the WS condition (i.e., \( M \) opportunities to respond \( [153.22] \)/instructional time \( [120 \text{ s}] \)) and 0.71 words per second in the TDP condition (i.e., \( M \) opportunities to respond \( [85.15] \)/ instructional time \( [120 \text{ s}] \)). Thus, the sheer volume of content processed in such a short time span, in the WS condition, may have contributed to the significant number of errors committed and thus, may have negatively impacted learning outcomes. As mentioned above, while WS offers a similar amount of opportunities to practice as does the IR intervention, the rate at which students read in the WS condition was much faster than one could read in IR (e.g., IR can take upwards to 15 minutes to implement [Burns et al., 2004; Matchett & Burns, 2009]). Thus, the number of words processed per second would be far fewer than in the WS condition and perhaps would be more meaningful.

An additional consideration is that the WS condition was student paced, which was thought to offer the advantage of eliminating delays between the presentation of words. However, in the TDP condition students were cued by the interventionist to read one word at a time. This consistent cuing may have increased attention and behavioral engagement; therefore, enhancing intervention outcomes for TDP. Additionally, the error correction (i.e., feedback) procedures varied only slightly (i.e., the TDP condition afforded more eye and facial contact [i.e., social reinforcement] from the interventionist), but may have varied enough to impact the
effectiveness of the feedback given to participants. Thus, social reinforcement may have strengthened participants’ ability to attend to the TDP condition (Allen, Henke, Harris, Baer, & Reynolds, 1967). Given the procedural differences between conditions, it is possible that the TDP condition may have benefited participants with attentional difficulties or participants in need of additional interventionist reinforcement, as the intervention provides individual prompts for each target word and affords slightly enhanced error correction procedures.

Finally, an additional alternative explanation relates to the presentation of the WS stimuli. Whereas the TDP condition presented stimuli to students one word at a time, in the WS condition the individual target word the student was attempting to read was presented in a field of other target words, which may have served as distractors. At least in Phase 1 of the WS condition where participants were acclimating to the novel stimuli, they may have had difficulty attending to individual words. As such, the WS intervention may well have required increased cognitive resources to filter out extraneous stimuli, leaving fewer resources available for processing the relevant stimuli. Although previous studies have noted improvements in sustained attention with the addition of stimulation in simple tasks (e.g., Zentall, 1985), studies of intratask stimulation in more complex tasks suggest that the addition of stimuli may actually draw attention away from relevant stimuli (e.g., Adams, Hayden, & Canter, 1974; Denton & McIntyre, 1978; Rosenthal & Allen, 1980; Rosenthal & Allen, 1980).

Despite some of the unexpected effectiveness and efficiency results of this study, two valuable lessons were learned. First, although research has emphasized the quantity of opportunities for practice, it seems likely that the quality of opportunities should be given greater consideration (Topping, Samuels, & Paul, 2007). Topping et al. (2007) caution researchers from over-simplifying the learning process and misleading the educational community to believe that
simply increasing opportunities for practice will enhance reading achievement. Specifically, their work suggests that gains in reading achievement can only be achieved with the combination of quantity and quality of practice and that pure quantity of reading, in the absence of quality, has diminishing marginal returns (Topping et al., 2007). As a consequence, practitioners and researchers are urged to consider other mediating variables (e.g., student behavioral engagement, the quality of visual stimuli, the quality of formative feedback on performance, and quality of interventionist-participant interactions) that may affect the quality of opportunities to respond, as they may be more influential in the learning process than sheer volume of reading (Berliner, 1990; Topping et al., 2007; Topping & Sanders, 2000).

The current study demonstrates the importance of exploring these mediating variables in order to develop drill procedures with enhanced efficacy. Specifically, evaluating the quality of participant behavioral engagement (e.g., on- and off-task time and number of errors committed), the visual quality of stimuli (e.g., differences in size of text, spacing, number of distracters, single word presentation versus connected text of random words, presenting stimuli on paper versus a computer screen, and color of print), the quality of interventionist and participant interactions (e.g., collaborative bond between interventionist and participant), and the quality of formative feedback on performance (e.g., immediacy of error correction, positive-reinforcement, eye contact, and facial recognition) are significant and important variables that require further exploration in order to better understand how to maximize the effects of drill procedures.

One final implication of the current study is the importance of individualized intervention selection. Specifically, several studies have focused on determining which intervention condition is most effective or superior, so that this knowledge can be applied more broadly across students with reading difficulties (e.g., Burns, Zaslofsky, Kanive, & Parker, 2012; Joseph & Schisler,
2007; MacQuarrie et al., 2002). However, findings from this study indicate that neither intervention approach was consistently superior across students and highlights the importance of selecting interventions that are well suited for an individual child, taking into consideration mediating variables, such as attention and behavioral engagement. The goal of individualized intervention selection can be accomplished through brief experimental analysis (BEA), which allows researchers and practitioners to assess the effectiveness of intervention strategies for individual children by manipulating environmental and/or instructional variables (Riley-Tillman & Burns, 2009).

**Limitations and Directions for Future Research**

Although findings suggest interesting implications for practitioners and researchers with regard to sight word interventions, implications should be tempered by the potential impact of limitations. First, although WS was designed to be a more convenient method of intervention, given its electronic generator and method of recording data, we did not formally collect social validity data from interventionists or participants. Consequently, it would be beneficial to gain a better understanding of how interventionists and participants perceived the WS intervention, especially given its novelty to the field. Additionally, future lines of research should consider assessing the applied usage of intervention methods in the classroom (i.e., the actual frequency with which intervention methods are implemented outside of research contexts), as this may actually be the best indicator of intervention approval (Nist & Joseph, 2008).

Second, we did not expect to observe significant differences in the errors committed between instructional conditions, but suspect that the mediating variables discussed above (e.g., quality of engagement, visual stimuli, interventionist-participant interactions, and interventionist feedback) may have contributed to the observed differences (Berliner, 1990; Preston, Heaton,
Unfortunately, we did not collect data with respect to these variables, which would have been useful in better understanding the study findings (e.g., identifying whether students with attentional difficulties benefited from one intervention over another and if they generally responded less well to intervention across conditions). Therefore, future studies should compare the effects of WS and TDP while also collecting mediating variable data. For example, with respect to behavioral engagement, researchers could use the Academic Intervention Rating Scale (AIRS; Volpe, 2011b), a brief eight-item scale that allows interventionists to rate the presence of inattentive and hyperactive behaviors during the delivery of intervention. In addition to collecting behavioral data, researchers might consider manipulating the visual quality of WS in order to optimize learning. For example, facets of the intervention that should be examined experimentally include: the number of words presented per line, the spacing of words on each line, font size of printed target words, and the color of printed target words. Researchers could also introduce the use of a piece of paper to cover stimuli that serve as distracters to students, a traditional accommodation that assists with visual tracking. Lastly, future studies could also examine the quality of interventionist and participant interaction, as well as the quality of formative feedback on performance as mediating variables that might influence attainment of sight word knowledge.

Third, as Phase 2 of WS focused on enhancing the speed at which students could read sight words, it would have been beneficial to have included reading fluency as an outcome measure. It is plausible to believe that WS may have resulted in greater fluency benefits than TDP given the training students received in Phase 2 of the WS intervention; however, because fluency did not serve as a dependent variable we cannot say for sure if such a benefit was
obtained. Future studies could include fluency as an outcome measure by using reading probes that mirror the presentation of WS. By adding a fluency measure, the differential effects of WS and TDP can be further refined and more precise recommendations for future use elucidated (e.g., if WS does result in enhanced reading fluency benefits interventionists might select it as a method of intervention over TDP if they are targeting intervention with a child with this specific skill deficit).

Fourth, we did not employ the use of control words (words not targeted by any intervention) in this study. Future work should consider incorporating control words in retention, maintenance, and generalization assessments to establish a standard against which instructional condition outcome measures can be compared. Incorporating control words will also help to ensure that observed effects can truly be attributed to the instructional conditions rather than to chance learning (e.g., learning taking place in the classroom or at home).

Finally, the decision to target a total of 12 words each day (across both intervention conditions) was made despite the fact that previous literature (i.e., Burns & Dean, 2005) suggests targeting only two to four words with struggling readers at this grade level. This decision was reached because of ceiling-effect concerns. Specifically, if fewer words were targeted, participants might have learned all the target words presented within each intervention condition; thus, making it difficult to compare intervention effects. It is possible that targeting so many words across each session influenced participants’ response to intervention. Additionally, it is also possible that a WS intervention with fewer target words, and therefore a less cluttered WS stimuli, might have led to superior effects compared to those found in the current study. To address this shortcoming, it would be beneficial to compare intervention effects with a between-group design (i.e., two independent groups with a larger overall sample size). This
methodological change would allow for fewer words to be targeted in each session (i.e., target words would be cut in half since each participant is only being exposed to a single condition) and would eliminate the concern that participants would reach a ceiling, allowing researchers to make a fair comparison across conditions.

**Implications for Practice**

This investigation has important implications for school professionals who are involved in selecting, designing, implementing, and evaluating sight word interventions. This study strongly suggests that although increasing opportunities for students to practice unknown material is an important ingredient of academic interventions, other factors must be considered. In our study, WS led to significantly more opportunities for practice but was generally less effective than TDP. Findings demonstrate the importance of recognizing mediating variables in the intervention context and further stress the importance of tailoring interventions to a student's specific needs.

In addition, as school psychology continues to be involved in academic remediation, the instructional technology that we use to address deficits must be continuously refined (Burns & Boice, 2009). This study was an attempt to examine a sight word intervention that could meet the need for both effectiveness and efficiency, above what current practices (i.e., TDP) have demonstrated to date. Although effectiveness and efficiency results were somewhat unexpected, especially in light of the opportunities to respond afforded by each condition, conducting studies of this kind is critical in the development of novel interventions that might realize enhancements in effectiveness and efficiency (Nist & Joseph, 2008). This study is only the first investigation to evaluate the effects of WS; future studies will determine if manipulating facets of the
intervention (e.g., number of targets words, arrangement of words, and adding context/meaning to strings of words) will lead to gains in effectiveness and efficiency.
References


Appendix

Figure 1. Example WordSheet Intervention Stimuli

meal link cave once down peep

link once peep cave down meal
peep meal once cave down link
link cave once down peep meal
once peep meal link cave down
link peep down meal cave once
down once peep meal link cave

-----------------------------------------------
meal down once peep link cave once peep meal link down cave peep
meal down link once cave meal down once peep cave link link down
meal once peep cave cave link down once peep meal meal peep
down link once cave peep once down cave meal link cave peep once
link meal down cave meal down link peep once peep cave meal
down link once peep once down link cave meal once peep link down
cave meal once meal peep down cave link meal peep once down link
cave down meal peep link once cave link cave peep once down meal
peep once meal link cave down meal down peep link cave once
down cave meal once peep link peep down once cave meal link link
cave once down peep meal peep link down cave once meal once
meal down cave peep link once cave peep down link meal
Figure 2. WordSheet Generator
Table 1

*Instructional Differences in Opportunities to Respond, Maintenance, and Generalization (n = 27)*

<table>
<thead>
<tr>
<th></th>
<th>TDP</th>
<th></th>
<th>WS</th>
<th></th>
<th>t</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Skew</td>
<td>Kurtosis</td>
<td>Mean (SD)</td>
<td>Skew</td>
<td>Kurtosis</td>
</tr>
<tr>
<td>Opportunities to</td>
<td>85.15 (17.62)</td>
<td>1.64</td>
<td>3.67</td>
<td>153.22 (53.17)</td>
<td>0.39</td>
<td>-0.50</td>
</tr>
<tr>
<td>Respond</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retention-Time 1</td>
<td>3.81 (1.55)</td>
<td>-0.41</td>
<td>0.19</td>
<td>3.11 (1.91)</td>
<td>-0.17</td>
<td>-1.08</td>
</tr>
<tr>
<td>Retention-Time 2</td>
<td>7.22 (2.83)</td>
<td>0.10</td>
<td>-0.95</td>
<td>6.11 (3.23)</td>
<td>0.05</td>
<td>-0.73</td>
</tr>
<tr>
<td>Retention-Time 3</td>
<td>10.56 (3.83)</td>
<td>-0.20</td>
<td>-0.40</td>
<td>8.96 (4.71)</td>
<td>0.11</td>
<td>-0.73</td>
</tr>
<tr>
<td>Maintenance</td>
<td>7.96 (4.17)</td>
<td>0.28</td>
<td>-0.66</td>
<td>6.33 (4.16)</td>
<td>0.97</td>
<td>0.73</td>
</tr>
<tr>
<td>Generalization</td>
<td>7.44 (4.16)</td>
<td>0.54</td>
<td>-0.69</td>
<td>6.67 (4.21)</td>
<td>0.83</td>
<td>0.56</td>
</tr>
<tr>
<td>Errors</td>
<td>14.44 (9.61)</td>
<td>0.86</td>
<td>-0.24</td>
<td>25.52 (10.52)</td>
<td>0.14</td>
<td>0.20</td>
</tr>
</tbody>
</table>

*Note.* TDP = Traditional Drill and Practice; WS = WordSheets

Standard error of the estimate is 0.45 for skew and 0.87 for kurtosis

* = p < .05, ** = p < .001.
Table 2

*Instructional Condition x Time Repeated Measures ANOVA (n = 27)*

<table>
<thead>
<tr>
<th>Effect</th>
<th>df</th>
<th>F</th>
<th>(\eta_p^2)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Instructional Condition</td>
<td>1</td>
<td>7.50</td>
<td>.22</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>(B) Time</td>
<td>2</td>
<td>71.57</td>
<td>.85</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>A x B (interaction)</td>
<td>2</td>
<td>1.30</td>
<td>.09</td>
<td>.29</td>
</tr>
<tr>
<td>Error (within groups)</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3

WS and TDP Retention, Maintenance, and Generalization Efficiency

<table>
<thead>
<tr>
<th>Group</th>
<th>WR</th>
<th>RR</th>
<th>WM</th>
<th>MR</th>
<th>WG</th>
<th>GR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDP</td>
<td>10.56</td>
<td>1.76</td>
<td>7.96</td>
<td>1.33</td>
<td>7.44</td>
<td>1.24</td>
</tr>
<tr>
<td>WS</td>
<td>8.96</td>
<td>1.49</td>
<td>6.33</td>
<td>1.06</td>
<td>6.67</td>
<td>1.11</td>
</tr>
</tbody>
</table>

Note. TDP = Traditional Drill and Practice; WS = WordSheets; WR = average number of words retained; RR = average rate of words retained (number of words retained/instructional time); WM = average number of words maintained; MR = average rate of words maintained (number of words maintained/instructional time); WG = average number of words generalized; GR = average rate of words generalized (number of words generalized/instructional time).

Instructional time was held constant across conditions (2 minutes).
Figure 3. Retention Means Plot

Note. Condition 1 = Traditional Drill and Practice (TDP) and Condition 2 = WordSheets (WS)
Appendix A

High Frequency Words
1. a 47. blue 93. did 139. full
2. able 48. body 94. dim 140. funny
3. about 49. bold 95. divide 141. garden
4. above 50. bolt 96. do 142. gate
5. afraid 51. bore 97. dock 143. gave
6. after 52. both 98. does 144. get
7. again 53. break 99. doll 145. girl
8. against 54. bring 100. don’t 146. give
9. all 55. brown 101. done 147. giving
10. already 56. bug 102. door 148. goal
11. also 57. build 103. dot 149. goes
12. always 58. but 104. down 150. going
13. am 59. butter 105. draw 151. gone
14. an 60. buy 106. drink 152. good
15. and 61. by 107. dug 153. got
16. animal 62. cab 108. eat 154. green
17. any 63. cake 109. edge 155. had
18. are 64. call 110. eight 156. happy
19. arms 65. came 111. else 157. hard
20. around 66. can 112. enough 158. has
21. as 67. car 113. evening 159. he
22. ask 68. carry 114. ever 160. head
23. ate 69. cash 115. every 161. hear
24. away 70. cat 116. eye 162. help
25. baby 71. caught 117. face 163. her
26. back 72. cave 118. fade 164. here
27. bake 73. children 119. fall 165. hide
28. bank 74. clean 120. family 166. him
29. bat 75. climb 121. fan 167. his
30. bay 76. coat 122. far 168. hold
31. bean 77. coin 123. fast 169. hook
32. bear 78. cold 124. father 170. horse
33. because 79. color 125. fed 171. hot
34. bed 80. come 126. feel 172. hour
35. beef 81. cool 127. find 173. house
36. been 82. core 128. first 174. how
37. before 83. could 129. five 175. hungry
38. began 84. cow 130. flower 176. hurt
39. begin 85. cut 131. fly 177. I
40. bent 86. dab 132. follow 178. idea
41. best 87. dance 133. for 179. if
42. better 88. dash 134. forest 180. in
43. big 89. day 135. found 181. into
44. bill 90. dear 136. four 182. is
45. bird 91. deep 137. friend 183. it
46. black 92. dice 138. from 184. its
185. job 231. near 277. ring 323. then
186. joke 232. nest 278. rob 324. there
187. jump 233. net 279. room 325. these
188. just 234. never 280. round 326. they
189. keep 235. new 281. run 327. think
190. kick 236. no 282. said 328. this
191. kind 237. not 283. sale 329. those
192. king 238. note 284. save 330. though
193. know 239. now 285. saw 331. thoughts
194. lace 240. ocean 286. say 332. three
195. lack 241. of 287. school 333. tiny
196. lake 242. off 288. second 334. tip
197. lamp 243. old 289. see 335. to
198. leaf 244. on 290. sent 336. together
199. learn 245. once 291. seven 337. too
200. let 246. one 292. sharp 338. try
201. lick 247. only 293. she 339. turn
202. light 248. open 294. shoe 340. two
203. like 249. or 295. shout 341. under
204. link 250. other 296. show 342. upon
205. little 251. our 297. sing 343. us
206. live 252. over 298. sit 344. use
207. lock 253. own 299. six 345. very
208. long 254. paper 300. sleep 346. wake
209. love 255. part 301. small 347. walk
210. made 256. peep 302. so 348. wall
211. main 257. people 303. soap 349. want
212. make 258. person 304. sob 350. warm
213. may 259. pick 305. some 351. was
214. me 260. picture 306. soon 352. wash
215. meal 261. piece 307. soup 353. watched
216. mice 262. play 308. start 354. water
217. milk 263. please 309. stop 355. we
218. mine 264. pool 310. sure 356. wear
219. minute 265. present 311. table 357. web
220. moat 266. pretty 312. take 358. well
221. mob 267. pull 313. talk 359. went
222. moon 268. put 314. tall 360. were
223. more 269. rain 315. teacher 361. what
224. morning 270. ran 316. tell 362. when
225. most 271. rank 317. ten 363. where
226. mother 272. read 318. thank 364. which
227. much 273. ready 319. that 365. white
228. must 274. red 320. the 366. who
229. my 275. ride 321. their 367. why
230. myself 276. right 322. them 368. will
369. wish 370. with 371. work 372. world 373. would
374. write 375. yellow 376. yes 377. you
378. your 379. zip
Appendix B

TDP Intervention Script
TRADITIONAL DRILL (TDP) PROCEDURES

**Materials:**
- Flashcards
- Intervention Record Form
- Stopwatch
- Pen

**TD INTERVENTION SCRIPT:**

**Teaching Phase**

1. Interventionist states, “[Student name] I am going to read some words from these flashcards so listen carefully as I read each word because I’m going to ask you to read them after me.”

2. Interventionist reads aloud the first of six target words and asks the student to repeat the word following their model by saying, “**What word?**” before moving on to the next target word.

3. If read correctly, interventionist makes one of the following comments, “**Nice work, good job, excellent.**” If read incorrectly, the examiner says, “**Listen again, this word is. . .**” and repeats the word.

4. Repeat the process until all six target words have been modeled and read correctly by the student.

**Intervention Phase**

5. The examiner shuffles the index cards for random ordering and states, “**Now, we’re going to go through the words again, only this time you will be reading them by yourself. Ready?**”

6. Interventionist shows index cards one by one to the student and starts stopwatch once the student begins reading the first flashcard. If read correctly, move to next card and occasionally give praise. **If read incorrectly or there is no response within 3 seconds, the interventionist provides corrective feedback by saying, “No that word is X, what word?”**
(child says word). Simultaneously, interventionist places a tally mark (denoting the error) on their intervention record form.

7. Once the student has read all six target words, shuffle index cards and present them again to the student. Repeat process for duration of intervention.

8. Interventionist denotes the number of opportunities to respond by placing a tally mark (I) on the intervention record form for each time the student cycles through the flashcards. If the student is only able to complete part of cycle before one minute elapses, then indicate this by writing “X/6.” For example, if the student reads 3 out of 6 words when the timer goes off, write 3/6.

9. Repeat intervention until 2 minutes has elapsed. When two minutes has elapsed, ask the student to stop reading by saying “stop.” Then say, “Wow, you read a lot of words.”
Appendix C

WS Intervention Script
**WORDSHEET (WS) PROCEDURES**

**Materials:**
- Two copies of WordSheet intervention (examiner copy and student copy)
- Stopwatch
- Pen

**WS INTERVENTION SCRIPT:**

**Teaching Phase**

1. Interventionist states, “[Student name] I am going to read some words from this word sheet (first 6 words on sheet). Listen carefully as I read each word because I’m going to ask you to read them after me."

2. Interventionist reads first word aloud and asks the student to repeat the word by saying, “What word?”

3. If the student reads a word correctly, the examiner makes one of the following statements, “Nice work, good job, or excellent.” If read incorrectly, the interventionist says, “Listen again, this word is “X,” now you read it.”

4. Repeat procedure until all six target words have been modeled and read correctly by the student.

**First Section of Intervention**

5. Next, say “Now we’re going to go through the words again, only this time you’ll read them by yourself. Ready?”

6. Interventionist points to the first target word in the first section, asks the student to begin reading by saying “Begin.” As soon as the first word is read (even if incorrectly) the interventionists starts stopwatch and continues until one minute has elapsed. (Avoid pointing to each word. If the student stops, say “go on” or if they pause say “read the next word.”)
7. If words are read correctly, occasionally provide praise by saying, “good job.” If words are read incorrectly or there is no response within 3 seconds, provide corrective feedback by saying, “Stop, that word is “X.” What word?” Mark the error on the interventionist form by placing a ( / ) through the word.

8. When one minute has elapsed, ask the student to stop reading by saying “stop.” Indicate where the student stopped reading by placing a bracket ( ] ) around the last word read. If the student reads all of the words in this section in less than one minute, ask them to return to the top of the section. Indicate on the examiner copy where the student stopped and if they had to begin at the top again.

**Second Section of Intervention**

9. After completing the first section of the WordSheet, move to the second part and say, “Okay now I want to see how many words you can read in 1 minute - get ready, begin.”

10. If words are read correctly, occasionally provide praise by saying, “good job.” If read incorrectly or there is no response within 3 seconds, provide corrective feedback by saying, “Stop, that word is “X.” What word?” Mark the error on the interventionist form by placing a ( / ) through the word.

11. Draw a bracket ( ] ) after last word read and say “Wow you read a lot of words, good job!” If the student reads all of the words in this section in less than one minute, ask them to return to the top of the section. Indicate on the examiner copy where the student stopped and if they had to begin at the top again.
Appendix D

TDP Procedural Integrity Checklist
**Student:** _________________________

### TRADITIONAL DRILL PROCEDURAL INTEGRITY CHECKLIST

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Interventionist uses flashcards for traditional drill condition</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Interventionists models each target word</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Corrective feedback is provided for words read incorrectly or not read within 3s</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Interventionist places tally marks for errors and opportunities to respond on the traditional drill record form</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Verbal praise is occasionally provided for words read correctly</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Interventionist shuffles cards after each trial of traditional drill</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Interventionist conducts traditional drill intervention for 2 minutes</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Interventionist conducts intervention conditions in correct order (e.g., Traditional Drill then WordSheets OR WordSheets then Traditional Drill)</td>
<td></td>
</tr>
</tbody>
</table>

**Total =**
Appendix E

WS Procedural Integrity Checklist
**Student:** _________________________

**WORDSHEETS PROCEDURAL INTEGRITY CHECKLIST**

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Interventionist presents WordSheet for WordSheet condition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Interventionist models each target word</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Interventionist conducts first section of WordSheet intervention for 1 min</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Corrective feedback is provided for words read incorrectly or not read within 3s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Verbal praise is occasionally provided for words read correctly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Interventionist conducts second section of WordSheet intervention for 1 min and reads instructions on the intervention script verbatim</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Interventionist marks errors by placing a (/) through the word on the interventionist form</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Interventionist draws a bracket ([ ] ) after last word read on the interventionist form</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Interventionist conducts instructional conditions in correct order</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Total =**
Appendix F

Example Retention Assessment Sheet
## RETENTION 1 FORM

**Student:** _____________________

<table>
<thead>
<tr>
<th>Unknowns from Both Instructional Conditions</th>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. sale (TD - #283)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. edge (TD - #109)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. both (TD - #52)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. afraid (TD - #5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. else (TD - #111)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. run (TD - #281)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. present (WS - #265)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. coat (WS - #76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. after (WS - #6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. ready (WS - #273)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. fed (WS - #125)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. know (WS - #193)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Total =**
Appendix G

Example Maintenance Assessment Sheet
MAINTENANCE FORM

Student: _________________________

<table>
<thead>
<tr>
<th>Unknowns from Both Instructional Conditions</th>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. both (TD - #52)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. else (TD - #111)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. run (TD - #281)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. present (WS - #265)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. coat (WS - #76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. after (WS - #6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. ready (WS - #273)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. fed (WS - #125)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. call (TD - #64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. much (TD - #227)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. feel (TD - #126)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. five (WS - #129)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. teacher (WS - #315)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. giving (TD - #147)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. open (TD - #248)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. said (TD - #282)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. cave (WS - #72)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. peep (WS - #256)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix H

Example Generalization Assessment Sheets
GENERALIZATION FORM

Student: __________________________

<table>
<thead>
<tr>
<th>Unknowns from Both Instructional Conditions</th>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. We both went to play.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. What else is here?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I can run fast.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. She likes the present.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Put on your coat.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Go out after him.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Are you ready to go?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. I fed the cat.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Can you call me?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. I ate too much.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. I feel sad now.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. I am five today.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. The teacher said so.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. He is giving her a car.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Open the door.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. She said no way.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. The cave was dark.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Do not say a peep.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total =
1. We both went to play.

2. What else is here?

3. I can run fast.

4. She likes the present.

5. Put on your coat.

6. Go out after him.

7. Are you ready to go?

8. I fed the cat.

9. Can you call me?

10. I ate too much.

11. I feel sad now.

12. I am five today.

13. The teacher said so.

14. He is giving her a car.

15. Open the door.

16. She said no way.

17. The cave was dark.

18. Do not say a peep.
Appendix I

Retention Assessment Procedural Integrity Checklist
**Student:** _______________________

**Date:** _______________________

---

**RETENTION PROCEDURAL INTEGRITY CHECKLIST**

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1.</strong> Interventionist presents retention flashcards in random order to the student</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2.</strong> Interventionist presents correct target words (i.e., words targeted in preceding intervention session) within retention assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3.</strong> Interventionist records words read (in)correctly on the retention form</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>4.</strong> Verbal praise is occasionally provided for words read correctly</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Total =**
Appendix J

Maintenance Assessment Procedural Integrity Checklist
Student: _________________________
Date: ___________________________

MAINTENANCE PROCEDURAL INTEGRITY CHECKLIST

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1.</strong> Interventionist presents maintenance flashcards in random order to the student</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>2.</strong> Interventionist presents correct target words (i.e., words read correctly on retention probes) within maintenance assessment</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>3.</strong> Interventionist records words read (in)correctly on the maintenance form</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>4.</strong> Verbal praise is occasionally provided for words read correctly</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Total =</strong></td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
Appendix K

Generalization Assessment Procedural Integrity Checklist
Student: ______________________

Date: ______________________

**GENERALIZATION PROCEDURAL INTEGRITY CHECKLIST**

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Interventionist presents generalization sentences to student</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Interventionist records words read (in)correctly on the generalization form</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total =