A comparison of three tablet-based sight word flashcard interventions

Brandon Paul Sinisi

University at Albany, State University of New York, brandon.sinisi@gmail.com

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Educational Psychology Commons, Educational Technology Commons, and the Language and Literacy Education Commons

Recommended Citation

https://scholarsarchive.library.albany.edu/legacy-etd/2386

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
A COMPARISON OF THREE TABLET-BASED
SIGHT WORD FLASHCARD INTERVENTIONS

by

Brandon P. Sinisi

A Dissertation
Submitted to the University at Albany, State University of New York
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Psychology

School of Education
Department of Educational and Counseling Psychology
2019
ABSTRACT

The purpose of the present study was to compare the efficiency of three tablet-based flashcard sight word interventions with varying ratios of known to unknown material that have historically been administered in paper-based form. To control for level of intervention duration, which tends to vary due to the presence of different amounts of known items across the conditions, the OTR for each flashcard procedure across intervention sessions were fixed. Additionally, this study investigated four specific outcome variables for each intervention: (a) the rate of learning new sight words (i.e., acquisition per minute of instruction), (b) skill maintenance of learned sight words at two-week follow-up, (c) the generalizability of newly-learned sight words to paper-based reading, (d) the degree to which the interventions were administered with integrity. Overall, results were positive. One condition, DI, was eliminated due to technical glitches in the software mid-intervention and will not be discussed further. For the other two conditions, acquisition, retention, and generalization of sight words was comparable to that of other studies that used paper-based cards and manual administration. Efficiency was comparable amongst the two remaining interventions. The computerized application evaluated presently stands to improve the fidelity of implementation of responsive intervention, thereby increasing the quality of instructional decisions in schools.
DEDICATION

Dedicated to my parents, Pasquale and Joan Sinisi, whose love, encouragement, and support allow me to remain in constant pursuit of my dreams.
ACKNOWLEDGEMENTS

This work would not have been possible without the patience, commitment, and expertise of Flash Fluency’s creator, Dhruvil Kotak. I am especially indebted to Dr. Benjamin Solomon, Chairman of my Dissertation Committee, who ensured this project had everything it needed to succeed from its inception. My sincere appreciation goes to the graduate students of Dr. Solomon’s research lab: Carmela Battista, Kayla Campana, Shannon Killar, Erin Marr, Lexy Payne, Sophie Parks, Renee Rymarz, Alexander Silva, and Sam Sutton. I wish to thank Dr. David Miller for his support and guidance as a committee member. I would also like to thank Dr. Callen Kostelnik for initially serving as a committee member prior to her maternity leave. I am incredibly grateful for Dr. Matthew LaFave agreeing to step in as a committee member in Dr. Kostelnik’s place. A very special thanks to my field experience supervisors: Dr. Carrie Trimarchi, Claire Marcus, Dr. Sheila McLean, Dr. Elizabeth Brown, Dr. Yiping Sherer, and Dr. Steve Marcal. While not a part of my dissertation committee, they have enriched my knowledge as a researcher and practitioner. Most of all, thank you to the elementary students who participated in this study.

Finally, I could not have completed this dissertation without the support and encouragement from my friends and family: Charlie Reutemann, Chelsea Reutemann, Kelly Rossman, Michael Rossman, Breanna Sinisi, Ginger Sinisi, Kaylee Sinisi, Madison Sinisi, Ryan Sinisi, and Toby Sinisi. Thank you to my mother and father, Pasquale and Joan Sinisi, for always supporting me with anything I set my mind to. Last but not least, thank you to my loving wife, Sarah Mae Sinisi, for believing in me and inspiring me to always aim higher.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter 1: Introduction</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose Statement</td>
<td>3</td>
</tr>
<tr>
<td>Possible Implications</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 2: Review of Relevant Literature</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response to Intervention</td>
<td>6</td>
</tr>
<tr>
<td>The Instructional Hierarchy</td>
<td>9</td>
</tr>
<tr>
<td>Application of the Instructional Hierarchy for Reading Intervention</td>
<td>10</td>
</tr>
<tr>
<td>Effective Intervention Components for Sight Word Reading</td>
<td>16</td>
</tr>
<tr>
<td>Flashcard Intervention Techniques</td>
<td>19</td>
</tr>
<tr>
<td>Traditional Drill and Practice</td>
<td>20</td>
</tr>
<tr>
<td>Direct Instruction</td>
<td>21</td>
</tr>
<tr>
<td>Incremental Rehearsal</td>
<td>24</td>
</tr>
<tr>
<td>Comparative Studies</td>
<td>25</td>
</tr>
<tr>
<td>Advancements in Flashcard Intervention Technology</td>
<td>29</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 3: Methodology</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
<td>33</td>
</tr>
<tr>
<td>Participants</td>
<td>33</td>
</tr>
<tr>
<td>Materials</td>
<td>33</td>
</tr>
<tr>
<td>Pre-Intervention Assessment</td>
<td>34</td>
</tr>
<tr>
<td>Intervention</td>
<td>34</td>
</tr>
<tr>
<td>Post-Intervention Maintenance Assessment</td>
<td>35</td>
</tr>
<tr>
<td>Procedure</td>
<td>36</td>
</tr>
<tr>
<td>Design and Data Analysis Procedures</td>
<td>36</td>
</tr>
<tr>
<td>Treatment Integrity</td>
<td>38</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 4: Results</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive Trends across Participants</td>
<td>39</td>
</tr>
<tr>
<td>Statistical Analyses</td>
<td>39</td>
</tr>
<tr>
<td>Visual Analysis</td>
<td>41</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>4.1 Number of Words Learned, Mean Time of Instruction, Total Time of Instruction, Rate of Words Learned, Number of Words Maintained, Percentage of Words Maintained, Rate of Words Maintained, and Percentage of Words Generalized for Each Student by Instructional Condition</td>
<td>22</td>
</tr>
<tr>
<td>4.2 Results of Non-linear Model with the Logit Link Function</td>
<td>23</td>
</tr>
<tr>
<td>5.1 Mean Rate of Words Learned, Mean Rate of Words Maintained, Mean Rate of Words Generalized</td>
<td>24</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Details of Traditional Drill and Practice Procedure</td>
<td>22</td>
</tr>
<tr>
<td>2.2 Details of Direct Instruction Procedure</td>
<td>23</td>
</tr>
<tr>
<td>2.3 Details of Incremental Rehearsal Procedure</td>
<td>24</td>
</tr>
<tr>
<td>4.1 Number of words Sarah acquired and retained across sessions</td>
<td>46</td>
</tr>
<tr>
<td>4.2 Number of words Madison acquired and retained across sessions</td>
<td>47</td>
</tr>
<tr>
<td>4.3 Number of words Toby acquired and retained across sessions</td>
<td>48</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

The three-tiered model of Response-to-Intervention (RtI) emphasizes a data-based, scientific approach to assess the academic progress of students to inform decisions for appropriate academic intervention (Bineham, Shelby, Pazey, & Yates, 2014; Brown-Chidsey & Steege, 2005). The literature overwhelmingly supports the notion that monitoring academic outcomes and remaining responsive to the individual needs of each student is necessary for ensuring academic success, particularly when it comes to fostering at-risk children’s foundational reading skills (e.g., Deno, 1985; Fuchs & Fuchs, 1986; Fuchs, Deno, & Mirkin, 1984; Shinn, 2007).

Haring and Eaton’s (1978) conceptualization of the Instructional Hierarchy (IH) outlines four stages of learning by which academic skills are consolidated: (a) skill acquisition, (b) fluency-building, (c) generalization, and (d) adaption. This approach to learning facilitates educators’ instructional practices by deconstructing reading interventions to their individual treatment components. In other words, by accurately assessing academic skills in accordance with the IH, teachers are able to select interventions that are efficient in addressing skill-deficits at specific stages of the IH. Unfortunately, educators are likely to possess negative perceptions and attitudes toward school-wide implementation of RtI (Castro-Villarreal, Rodriguez, & Moore, 2014). In a survey of teachers, Tillery, Varjas, Meyers, and Collins (2010) found that the majority of respondents indicated they did not possess an adequate understanding of RtI. Connecting these findings with research related to implementation integrity, it can be reasonably assumed that in order for students to receive adequate instruction consistent with the data-driven, evidence-based approach of RtI, schools are in need of innovative strategies to simplify and streamline assessment and intervention practices.
Flashcard intervention procedures are commonly recommended within the RtI framework, typically as a tier II intervention. Initially, these interventions were intended to be administered using paper-based index cards. However, the convenience, speed, and adaptability of modern technological devices holds potential for students to benefit from flashcard interventions under the broader model of RtI. Flashcard intervention techniques have been shown to be an invaluable tool in facilitating student learning across academic areas (e.g., January, Lovelace, Foster, & Ardoin, 2016; Skarr, Zielinski, Ruwe, Sharp, Williams, & McLaughlin, 2010; Szadokierski & Burns, 2008). However, as pedagogical strategies have become further integrated with the use of technology, targeted evidence-based interventions have failed to adapt to the modern classroom at a similar pace (Jameson, Thompson, Manuele, Smith, Egan, & Moore, 2012). As a result, the fate of empirically-supported academic interventions may be at risk, as they may be replaced by more convenient, technology-based educational programs that lack peer-reviewed evaluation.

Therefore, there is a critical need for current research to develop and evaluate academic interventions that have traditionally been administered in paper form, but can be adapted to a technology-based platform. However, this paradigm shift should not be limited to simply investigating the generalization of paper-based intervention effectiveness to technological devices. Rather, the blending of technology with academic intervention should be viewed as a new frontier for comparative studies of flashcard intervention techniques, as modern programming allows for tightly controlled treatment component comparisons of flashcard intervention techniques that follow complex algorithms.

 Likely drawing from Vygotsky’s (1978) theory of the Zone of Proximal Development and Skinner’s (1966) notion of the Experimental Analysis of Behavior, Gickling and Armstrong
(1978) recognized the importance of matching students’ skill level with instructional material that is neither too easy nor too challenging to maximize learning potential. This concept has been a major focus among the literature related to flashcard intervention techniques. That is, many studies have sought to compare the effectiveness of different techniques, which tend to vary in their ratios of known to unknown material. However, several researchers have recognized the role of instructional time as it relates to the feasibility of administering evidence-based interventions, which resulted in a shift in focus from effectiveness to efficiency (Skinner, Belfiore, & Watson, 1995). Excluding studies that match instructional time between techniques, the literature tends to favor interspersed techniques (e.g., IR, DI) over traditional drill and practice. The former includes both known and unknown material while the latter consists only of unknown material. Conversely, studies that compared techniques under time constraints favored traditional drill and practice.

Daly, Murdoch, Lillenstein, Webber, and Lentz (2002) recognized that the divergence of these findings is best explained by the theoretical principles of the IH, as interspersed techniques typically provide students with more opportunities to respond (OTR) to target words than traditional drill and practice. Currently, there is a lack of experimental studies involving technology-based academic interventions that equalize the number of OTR across conditions, despite (a) a growing reliance on technology to enhance student learning, and (b) initial evidence that transfer of learning from stimuli presented on illuminated screens to paper should not be assumed (Duhon, House, Poncy, Hastings, & McClurg, 2010; Hirsch-Pasek, Zosh, Golinkoff, Gray, Robb, & Kaufman, 2015).

**Purpose Statement**
The purpose of the present study was to compare the efficiency of three tablet-based flashcard sight word interventions with varying ratios of known to unknown material that have historically been administered in paper-based form. To control for level of intervention duration, which tends to vary due to the presence of different amounts of known items across the conditions, the OTR for each flashcard procedure across intervention sessions were fixed. Additionally, this study investigated four specific outcome variables for each intervention: (a) the rate of learning new sight words (i.e., acquisition per minute of instruction), (b) skill maintenance of learned sight words at two-week follow-up, (c) the generalizability of newly-learned sight words to paper-based reading, (d) the degree to which the interventions were administered with integrity.

**Possible Implications**

This study has potential for testing the practical utility of technology-based academic interventions. Under the Individuals with Disabilities Education Improvement Act (2004), school districts are mandated to identify academically at-risk students and use evidence-based interventions to remediate skill deficits. However, successful systems-level changes such as these are often contingent upon educators’ degree of readiness for change and may require several years to effectively alter the established culture of the school (Fullan, Miles, & Taylor, 1980). Even educators that are fully committed to altering instructional approaches may struggle, as research has demonstrated that even the most robustly validated interventions may not be effective if the complexity of the intervention hinders educators’ adherence to treatment fidelity (Carter, 2008).

Time—another potential obstacle for the implementation of evidence-based interventions—has also been shown to influence treatment fidelity. The more time that is
required to prepare and administer an intervention, the less likely educators are to implement the intervention with adequate integrity (Elliott, Witt, Galvin, & Peterson 1984; Martens, Elliott, & Darveaux, 1985). In other words, educators may be less likely to provide necessary academic supports to their at-risk students if they perceive a prescribed evidence-based intervention to be unfeasibly implementable (Glover & Albers, 2007). Fortunately, the automatization of virtually all intervention procedures in the current study simplifies data recording, eliminates the need to manually calculate student progress or apply decision rules, and automatically charts student progress across sessions. As such, the computerized application evaluated presently stands to improve the fidelity of implementation of responsive intervention, thereby increasing the quality of instructional decisions in schools.
Chapter 2: Review of Relevant Literature

The purpose of the current study was to both establish the relative efficacy and effectiveness of three tablet-based sight word flashcard interventions. The following chapter provides background information for the current study, including: (a) the need for effective, socially valid academic interventions within a Response to Intervention (RtI) approach; (b) the Instructional Hierarchy (IH) as a conceptual framework for selecting appropriate interventions, such as flashcard-based procedures; (c) effective intervention components for building sight word reading skills; and (d) the emerging empirical support for the technological adaptation of evidence-based flashcard procedures for increasing sight word recognition.

Response to Intervention

The RtI paradigm provides an organizing platform for schools to use formative and summative assessment data to inform decisions related to the selection and delivery of tiered, evidence-based interventions. It is a continuum that uses increasingly intensive interventions for academically struggling students until an effective intervention is identified (Daly, Persampieri, McCurdy, & Gortmaker, 2005). Although its tenets are rooted in decades of behavior-based assessment and intervention research (Deno, 1985; Fuchs & Fuchs, 1986; Fuchs, Fuchs, & Hamlett, 1989; Fuchs et al., 1984), the complete framework of RtI was unfamiliar to many prior to it being thrust into the national spotlight following the passage of the Individuals with Disabilities Education Improvement Act (IDEIA) of 2004 (Gresham, 2005; Shinn, 2007).

In addition to granting freedom to local education agencies to use RtI for the identification of students with specific learning disabilities (SLDs), the IDEIA also allowed agencies to use up to 15% of their federal special education funding to provide supplementary evidence-based academic intervention services to general education students to prevent later
classification (Gresham, Reschly, & Shinn, 2010), which recognized the significant influence of early identification and intervention. In effect, schools were permitted to use federal special education funding to establish universal assessment and intervention practices (Tier I) for all students, regardless of their status as a student in general or special education.

Conceptually, students that do not adequately respond to Tier I intervention are provided targeted intervention (Tier II) to remediate specific skill deficits. In contrast to the use of the ability-achievement discrepancy model for learning disability identification, which has historically failed to accurately predict which students would benefit from intervention (Gresham & Witt, 1997; Machek & Nelson, 2010; Stuebing, Fletcher, LeDoux, Lyon, Shaywitz, & Shaywitz, 2002), Tier II of the RtI model allows students whose skill deficits can be easily remediated through evidence-based intervention to avoid unnecessarily being classified as students with disabilities. Importantly, Tier II interventions are meant to be brief, low cost, and administered with relative ease. Flashcard interventions have been shown to fulfill this niche well (Nist & Joseph, 2008; Volpe et al., 2011). Successful skill remediation in Tier II results in the student returning to Tier I.

Students that do not adequately respond to evidence-based intervention in Tier II are provided with individualized, intensive intervention (Tier III), which is often provided through special education services. Flashcard-based procedures are appropriate here as well (Brasch, Williams, & McLaughlin, 2007; Higgins, McLaughlin, Derby, & Long, 2012; Ruwe, McLaughlin, Derby, & Johnson, 2010), commonly using more intensive procedures relative to their Tier II counterparts. As a whole, the RtI model allows relevant stakeholders to regularly monitor the progress of all students and appropriately intervene when necessary. The RtI model’s use of formative evaluation to tailor instruction has also demonstrated a large effect on
academic achievement (Fuchs & Fuchs, 1986), whereas special class placement fails to consistently demonstrate any positive effect on academic achievement (Kavale & Forness, 1999).

Response-to-intervention may be considered one of the most revolutionary advancements in the large-scale delivery of educational services (Merrell & Buchanan, 2006). However, the implementation of RtI has proven to be challenging. Unfortunately, it is unlikely that most teachers possess an accurate understanding of RtI processes (Sansosti, Telzrow, & Noltemeyer, 2010; Tillery et al., 2010), which has likely contributed to the negative perceptions and attitudes that teachers often hold toward it (Castro-Villarreal et al., 2014). In their survey of elementary and secondary education teachers, Castro-Villareal et al. (2014) found that teachers perceived lack of training, time, and resources as the most common barriers to successful RtI implementation. In turn, these issues may lead to inadequate implementation integrity for evidence-based assessment and intervention practices. Although the RtI model has established effectiveness for improving student outcomes, its lack of emphasis on feasibility within the classroom is a common point of criticism made by teachers (Bineham et al., 2014).

As a result, RtI procedures may not always be implemented in a manner that is consistent with the evidence-base, which decreases the likelihood that students will demonstrate improvements in their academic performance (Witt, Noell, LaFleur, & Mortenson, 1997). Due to these obstacles, the field has consistently failed to liberate itself from the traditional refer-test-place model (O’Connor & Witter Freeman, 2012). Therefore, there is a critical need for innovative solutions that aim to strengthen the procedural integrity of academic interventions. Such solutions must be perceived as simple, quick, and user-friendly by relevant stakeholders (e.g., educators, parents, paraprofessionals, reading specialists) because procedural integrity is
inversely related to intervention complexity (Brown-Chidsey & Steege, 2005; Glover & DiPerna, 2007). It is also equally important to ensure such solutions do not stray from the basic tenets of RtI (i.e., evidence-based intervention implemented with high fidelity; Sanetti & Kratochwill, 2009) and follow an empirically-supported theoretical model for instruction (Ardoin & Daly, 2006; Daly, Lentz, & Boyer, 1996). The advancement of computer-based interventions may compliment RtI implementation in this regard. Digital intervention procedures have the potential to assure the standardized, high-fidelity implementation necessary to produce meaningful data concerning each individual student’s measured response to academic intervention.

The Instructional Hierarchy

Haring and Eaton’s (1978) conceptualization of the IH provides a systematic approach to instructional planning that assumes learning is a graded progression of skill development with four stages: (a) acquisition, (b) fluency-building, (c) generalization, and (d) adaption (Ardoin & Daly, 2006; Daly et al., 1996; Daly & Martens, 1994). The power of the IH is in its ability to inform the selection of intervention, conditional on pre-assessment data of the target skill. This developmental approach to instruction draws from a behavioral analytic framework to systematically increase the strength of students’ academic responding over time with empirically-supported instructional techniques (e.g., demonstration, model, prompting, drill, reinforcement; Ardoin & Daly, 2006; Codd & Poncy, 2010; Daly et al., 1996; Greenwood, Delquadri, & Hall, 1984). Its utility for increasing learning potential is rooted in applied behavior analysis; employing specific components of instruction to affect change in the student’s behavior (i.e., academic responding; Greenwood et al., 1984; Haring & Eaton, 1978). Within the IH, academic responding is operationalized as an observable and measurable dependent variable that educators seek to influence with empirically-supported instructional practices (Ardoin &
Daly, 2006). As students progress within the IH, specific techniques are employed to generate specific outcomes in academic responding (i.e., accuracy, maintenance, fluency, generalization).

**Application of the Instructional Hierarchy for Reading Intervention**

Daly and Martens (1994) were among the first to use the IH model to systematically conceptualize the effects of academic intervention on students’ learning behavior. Following the realization that students struggling to achieve grade-level expectations would likely need a more tailored approach to remediation than simply increasing the amount of time spent engaged in learning (Ysseldyke & Christenson, 1993), Daly et al. (1996) presented an adaptation of the IH for linking academic responding to specific intervention treatment components. By identifying students’ most relevant response targets (e.g., low reading accuracy, adequate reading accuracy with low fluency), educators are able to select intervention approaches that deliberately elicit the desired type of response that corresponds with the IH model. The IH model dictates which types of interventions should be used at what times to optimize student learning.

**Acquisition.** The IH requires the use of empirically-supported practices to teach instructionally appropriate academic skills. The first phase, skill acquisition, initiates the process of learning with methods of explicit instruction (i.e., demonstration, model, prompting; Daly et al., 1996; Ferkis, Belfiore, & Skinner 1997; Haring & Eaton, 1978) presented within three-term contingency trials (i.e., antecedent-response-consequence; Albers & Greer, 1991; Belfiore, Skinner, & Ferkis, 1995). The educator’s objective in this stage is to produce academic responding that is accurate in isolation (Espin & Deno, 1989).

For example, to demonstrate sight word reading, a teacher may present a word (e.g., “cat”) and read the word aloud. Demonstration has often been used interchangeably with the term “modeling” (e.g., Gersten, Woodward, & Darch, 1986; Volpe et al., 2011), but Haring and
Eaton’s (1978) IH differentiates these terms. Demonstration involves the performance of a desired response, whereas a model is a visual exemplar that can be referenced for the purpose of self-monitoring. A prompt may serve as the antecedent or consequence within a three-term contingency trial. As the antecedent, a visual and verbal prompt signals the expectation of a response (e.g., presentation of a sight word on a flashcard, “What is this word?”). For incorrect responses, additional prompts may take the form of phonetic prompting (e.g., “The word starts with the /c/ sound”) or whole-word prompting (e.g., “The word is ‘cat’”). Other terms have been used to describe the consequence following an incorrect response within the three-term contingency (e.g., word drill, corrective feedback, error correction; Daly et al., 1996, Ferkis et al., 1997; Martin-Chang, Levy, & O’Neil, 2007). The current study will use the term “corrective feedback” to describe whole-word prompting that takes place following hesitation or an incorrect response. Verbal reinforcement is given as a consequence within the three-term contingency trial following a correct response (e.g., “Correct”, “That is right”).

One alteration appears to have emerged subsequent to the initial presentation of the IH model. Although the effectiveness of drill has been established for fluency-building, Skinner, Fletcher, and Harington (1996) made a compelling case for using drill to enhance correct responding in the acquisition stage as well. Considering the amount of instructional time that educators are able to engage in administering Tier II interventions is likely limited, students may be able to demonstrate quicker gains when intervention techniques maximize student response rates (i.e., rapid drilling) proportional to three-term contingency trials (e.g., flashcard procedures; Belfiore et al., 1995). Although this may appear to contradict the IH, Haring and Eaton (1978) acknowledged that future research would likely dictate when it would be appropriate to present
stage-specific strategies concurrently. In other words, interventionists must balance the explicit instruction of discrete skills with the need to have students repeatedly practice those skills.

**Fluency.** The second stage of the IH aims to use techniques to increase the speed at which the skill is performed in context. The concept of skill maintenance, which is the retention of fluency over time, is also considered an objective of attaining fluency (Haring & Eaton, 1978). In order to meaningfully apply a learned skill, the IH recognizes that the student must be able to perform the skill with automaticity, where automaticity is defined as the relative ease and speed at which a behavior is produced (LaBerge & Samuel, 1974). For instance, a student that is able to consistently identify the sounds of corresponding printed letters and letter combinations and combine this information to sound-out words has mastered the skill of decoding. However, this skill must be applied with a sufficient pace to effectively extract meaning from words and sentences. A student’s ability to comprehend written text may be affected if she or he requires significantly more time to read a sentence in comparison to her or his peers (Daly & Martens, 1994), with too much effort dedicated to decoding, rather than deriving meaning from text (Knight-Teague, Vanderwood, & Knight, 2014; Paleologos & Brabham, 2011).

While the definition of fluency may vary based on the skill that is being taught (Haring & Eaton, 1978), empirical research has helped to identify age- and grade-level standards for skills related to literacy and mathematics that must be demonstrated consistently over time (Shinn, 1989). To build fluency in this phase, educators primarily use the drill technique, which provides students with repeated OTR and reinforcement (e.g., verbal praise, tokens to later exchange for a prize; Daly et al., 1996). When selecting a drill technique, interventions that maximize OTR, which is the number of presentations of, for example, sight words within a session, are more likely to lead to the retention of newly learned words (Burns, 2007). This is
further enhanced through overlearning, which involves practicing a skill well beyond the point at which adequate fluency is demonstrated (Haring & Eaton, 1978). However, the degree of practice that would constitute adequate overlearning is subject to debate, as Skinner et al. (1996) would argue that educators should counterbalance overlearning procedures with the introduction of new material to maximize overall student learning.

LaBerge and Samuels’s (1974) model of automatic information processing has often served as a rationale for enhancing sight word recognition among struggling readers. The model implies that students must develop their sight word recognition skills to the point that they no longer require conscious attention. This concept is similar to Haring and Eaton’s (1978) regarding the role of overlearning in achieving automaticity. Students that continue to rely on phonetic decoding of words, while their peers transition to reading words by sight, are likely to demonstrate deficits in reading comprehension, as most of their effort is placed on accurate decoding (Tan & Nicholson, 1997). Similarly, students that must consciously review their lexicon prior to correctly identifying a word (i.e., greater than two seconds) will struggle with later comprehension. While this stage mainly emphasizes proficiency and maintenance, prolonged repetition of single tasks may not be sufficiently engaging for students. Therefore the inclusion of reinforcement procedures in this stage is often beneficial in maintaining motivation and promoting an appropriate pace (Ardoin & Daly, 2007).

**Generalization.** The third stage of the IH involves generalization of the skill, which is defined as the ability to perform the target skill accurately and fluently in novel situations (Martin-Chang et al., 2007; Schmidgall & Joseph, 2007; Stokes & Baer, 1977). Therefore, students are expected to be able to demonstrate their skills across contexts. In Shahan and Chase’s (2002) review of the literature related to generalization, they determined that students
are more likely to generalize skills when they possess strong accuracy and fluency skills. Consequently, students demonstrating difficulty in this stage may require mastery of an earlier phase of the hierarchy, suggesting Haring and Eaton’s (1978) sequence of skills is relevant for instructional planning. However, generalization should not be assumed once a solid foundation of accurate and fluent responding is present (Daly et al., 1996; Stokes & Baer, 1977). Although the use of flashcard procedures is likely to improve students’ sight word recognition accuracy and fluency, such gains have inconsistently generalized to improved reading comprehension (Fleischer, Jenkins, & Pany, 1979; Just & Carpenter, 1987; Tan, Moore, Dixon, & Nicholson, 1994), and generalization is inconsistently assessed in the empirical literature.

Depending on the skill, generalization likely will need to be programmed into the learning process across contexts through the use of two types of practice: discrimination and differentiation (Baer, Wolf, & Risely, 1968; Haring & Eaton, 1978). Discrimination training teaches students to produce a specific response when particular discriminative stimuli are presented, but not when other stimuli are presented (e.g., producing the /s/ sound for sock, sit, and still, but not for shot, ship, or show; Eckerman, 1970). Differentiation practice involves teaching the student to respond similarly to varying forms of a stimulus (e.g., “there”, “their”, “they’re”; Kerckhoff, 1986). Successful generalization may only be claimed when both discrimination and differentiation can be demonstrated consistently while performing the skill (Stokes & Baer, 1977).

Consequently, particularly for computer-based flashcard interventions, it is important to consider generalization potential when selecting appropriate interventions. Although there is some evidence that gains from computer-based sight word interventions transfer to paper-based word lists (Musti-Rao, Lo, & Plati, 2014) and hand-printed sight words presented on index cards
(Yaw, Skinner, Orsega, Parkhurst, Booher, & Chambers, 2012), empirical support of computer-based sight word interventions is still emerging. It is therefore critical to ensure students are able to demonstrate the same level of proficiency with sight words in alternate forms involving components that may affect generalization (e.g., font, size, letter casing; Musti-Rao et al., 2014; Yaw et al., 2012).

**Adaption.** According to the IH, a skill is fully learned when the student is able to independently and flexibly apply it across contexts (Haring & Eaton, 1978). Adaption heavily emphasizes the student’s ability to problem-solve by modifying the application of a learned skill to successfully accomplish a task. Progression to this phase requires a shift in instruction toward the provision of multiple opportunities to apply the modified skill across differing naturalistic settings. The IH model described by Daly et al. (1996) does not include the adaption stage.

The IH provides educators with a deconstructed understanding of the various treatment components of reading intervention, which facilitates appropriate intervention selection for each level of the hierarchy (Daly et al., 1996). Conceptually, the elements of flashcard procedures are consistent with promoting learning in the first three tiers of the IH, but are most strongly rooted to the acquisition stage. For sight word reading, students that consistently demonstrate less than 90% accuracy (Fuchs, Fuchs, & Hollenbeck, 2007; Johnson, Jenkins, Petscher, & Catts, 2009) would likely benefit from an evidence-based flashcard procedure that provides the essential components outlined in the acquisition stage of the IH. The refinement of analyzing students’ low academic achievement in this manner is thought to improve the accuracy of instructional decision-making, as it relies more on a scientific approach using objective data and an evidence-based heuristic, as opposed to professional judgment, which may be influenced by bias (Daly & Murdoch, 2000).
Effective Intervention Components for Sight Word Reading

As a cornerstone of reading proficiency, students’ sight word recognition serves an important function toward higher-order reading skills (National Institute of Child Health and Human Development [NICHD], 2000). Appropriate sight word knowledge is commonly and explicitly taught in elementary school. Under the RtI approach, students who are not able to read words with sufficient accuracy in comparison to their same-aged peers are provided with targeted intervention to prevent successive reading difficulties. As previously mentioned, there are empirically-supported approaches to identifying students that require intervention. However, as recognized by Gersten (1985), it is important to conduct component analyses to determine which parts of the intervention meaningfully contribute to specific improvements in performance. In terms of the IH, many studies have aimed to compare sight word intervention strategies within and between the levels of the hierarchy (Belfiore et al., 1995; Espin & Deno, 1989; Jenkins, Larson, & Fleisher, 1983; Skinner, Smith, McLean, 1994). Others have also sought to examine how the adjustment of frequency, intensity, or timing of specific components influences skill acquisition (Joseph & Nist, 2006; Skinner et al., 1996, Skinner, Belfiore, & Watson, 2002).

Espin and Deno’s (1989) study compared the provision of whole-word corrective feedback (i.e., interventionist reads the sight word aloud) and phonetic prompting (i.e., interventionist says the beginning of the sight word) following hesitations or incorrect responses. The results indicated that whole-word corrective feedback was more effective than prompting for increasing students’ accurate responding. In a replication and extension of Espin and Deno’s (1989) study, Barbetta, Heron, & Heward, 1993) reexamined the effectiveness of each component with students with developmental disabilities. They also added a demonstration
component when words were initially presented for both the whole-word corrective feedback and phonetic prompting conditions. In support of Espin and Deno’s (1989) findings, Barbetta et al. (1993) found that whole-word corrective feedback was more effective. In contrast to these conclusions, Carnine (1980) found that prompting was more effective. However, in Carnine’s (1980) study, sound-identification and blending training preceded the presentation of the target words, which may have primed the students to use phonetic analysis over whole-word reading. Together, these studies indicate that the initial teaching procedure (i.e., phonic analysis, whole-word reading) should inform the selection of an appropriate approach to error correction (i.e., whole-word corrective feedback, prompting; Barbetta et al., 1993; Carnine, 1980). In the context of sight words, it appears whole word feedback is more sound.

Jenkins et al. (1983) compared the effectiveness of whole-word corrective feedback and response repetition to a combination of techniques (i.e., corrective feedback, response repetition, drill). In the first condition, the interventionist demonstrated how to read the word following an incorrect response and the student repeated the word. For the second condition, the same procedure was used in addition to drill, which involved continuing to present the error words again until correct responses were provided for all words. During the drill procedure, whole-word corrective feedback and response repetition were used following the provision of incorrect responses. The results suggested that drill was responsible for significantly higher accuracy scores upon follow-up. However, it is important to recognize that the condition involving drill also provided more three-term contingency trials. Therefore, the significant difference in word recognition is not definitive verification that drill alone improves accuracy.

Belfiore et al. (1995) compared the effects of response- and trial-repetition on the acquisition of sight words among three elementary students with learning disabilities in reading.
Each student received sight word training in both conditions during each intervention session. The trial-repetition condition involved an antecedent-response-feedback sequence, which included: (a) prompting the student to read the word, (b) the student’s correct, incorrect, or absent response, and (c) confirmation that the word was read correctly or the provision of corrective feedback. This sequence was repeated until five interspersed three-term contingency trials were provided to the student for each sight word. In the response-repetition condition, only one antecedent-response-feedback sequence was completed for each sight word, but was followed by prompting the student to repeat the word four additional times. In other words, the first condition provided 25 three-term contingency trials and the second condition provided five three-term contingency trials with 20 word repetitions. Their findings indicated that students mastered more words in the trial-repetition condition, which supported their hypothesis that using three-term contingencies trials for all OTR likely leads to more favorable outcomes for sight word recognition.

In a comparison of two flashcard interventions that differed only in the duration of intervals between trials (i.e., immediate, five-seconds), Skinner et al. (1994) found that increases in students’ sight word reading accuracy did not differ based on condition. In both groups, students were provided with response prompts, corrective feedback for incorrect responses, and reinforcement (i.e., token that could be traded for stickers at the end of each session) for correct responses. Skinner et al.’s (1994) findings support maximizing students’ OTR during intervention sessions. Although it may be counterintuitive to maintain a brisk pace when administering interventions to struggling readers, other studies have also demonstrated that students are likely to benefit from rapid pacing (e.g., Darch & Gersten, 1985; Rhymer, Dittmer, Skinner, & Jackson, 2000).
In contemporary intervention research, it is important to incorporate the lessons learned from previous studies to build the empirical backing for interventions. Therefore, in addition to using the IH model as a guiding framework, the current study incorporated empirically-supported extensions of the model. Each OTR was provided in the form of a three-term contingency trial, which has been found to enhance intervention effectiveness (Belfiore et al., 1995). Hesitations and incorrect responses were followed by whole-word corrective feedback, rather than phonetic prompting (Espin & Deno, 1989). Additionally, items were presented with rapid pacing to provide an adequate number of OTR within reasonably brief intervention sessions that can be manageably completed in a student’s typical school day schedule.

**Flashcard Intervention Techniques**

Various types of flashcard procedures have demonstrated effectiveness for the acquisition of sight words. However, as mentioned earlier, there is a need for academic interventions to also be efficient in order to enhance treatment acceptability within the RtI model (Nist & Joseph, 2006; Skinner et al., 1995). Instructional efficiency, or learning rate, is the measurable effectiveness of an intervention that accounts for the amount of time required to achieve the desired change in behavior (Skinner, 2008). Instructional efficiency for sight word flashcard interventions is typically determined by dividing the total number of words learned by the total amount of instructional time expended (Cates & Rhymer, 2003; Mulé, Volpe, Fefer, Leslie, & Luiselli, 2015; Nist & Joseph, 2008; Volpe et al., 2011).

Maximizing instructional time for all students has been recognized as a pillar of academic success (Skinner et al., 1996), and identifying efficient interventions carries important and practical utility. For this reason, Cates & Rhymer (2003) argued that measures of efficiency should replace (e.g., growth per minute of intervention), rather than supplement, measures of
intervention effectiveness (e.g., mean differences of scores from pre- to post-test). As a result, a number of studies have sought to compare the instructional efficiency of sight word flashcard procedures (January et al., 2017; Joseph & Nist, 2006; Mulé et al., 2015; Nist & Joseph, 2008; Schmidgall & Joseph, 2007). However, a number of studies that have examined instructional efficiency have failed to include measures of maintenance, generalization, and social validity, which yield critically important information with respect to informing decisions about appropriate intervention selection (Volpe et al., 2011).

Although a variety of flashcard procedures are present in the literature, three notable techniques include: traditional drill and practice (TDP), direct instruction (DI), and incremental rehearsal (IR). Each of these procedures draw from principles of behavioral analytic learning and have garnered empirical support for their effectiveness with respect to increasing sight word accuracy. TDP uses massed practice by including only unknown words, whereas DI and IR intermix varying ratios of both known and unknown items. That is, known words are presented between unknown words to improve retention. All procedures incorporate effective components of reading intervention to teach unknown words (i.e., demonstration, prompting, whole-word corrective feedback, reinforcement) and provide multiple OTR (i.e., drill). It is important to note that sight word training is not intended to replace the practice of teaching the basic literacy skills of phonemic awareness and letter-sound relationships. Rather, interventions targeting sight word recognition aim to increase the accuracy and speed at which students are able to read specific words that are consistent with developmentally-appropriate, grade-level expectations (Tan & Nicholson, 1997). Below, specific differences between the procedures are discussed.

**Traditional Drill and Practice.** The TDP procedure exclusively uses unknown words within each set of flashcards. Upon being presented with an unknown word for the first time, the
interventionist demonstrates how to read the word accurately and then prompts the student to read the word aloud. After all of the words included in the set have been taught, the student receives repeated practice. Although some studies have used specific criteria to determine when an unknown word may be considered to be a known word (e.g., correct response on three consecutive trials, correct response on nine consecutive trials, correct responses to all words within a set; Brasch et al., 2008; Burns & Sterling-Turner, 2010; Sterling & Burns, 2010), other studies have chosen to use timed sessions and retention probes. When students demonstrated the ability to correctly respond to items presented in the previous session within two seconds, those items were considered learned (Brasch et al., 2008; Burns & Sterling-Turner, 2010; Mulé et al., 2015; Nist and Joseph, 2008; Volpe et al., 2011).

With the exception of Mulé et al. (2015), most major studies that have systematically compared TDP to one or more other flashcard techniques chose not to shuffle the flashcards after each trial (e.g., U1, U2, U3, U4, U1, U2, U3, U4, U1, U2, U3, U4). In the current study’s TDP condition, unknown words will be randomized after each trial (Figure 2.1) to encourage students to remain visually fixated on the stimulus and to avoid a sequence effect. Corrective feedback is provided when the student does not read the word accurately or the student does not respond correctly within two seconds. The effectiveness of TDP has been examined using three to six words within each session (i.e., set size; Burns & Sterling-Turner, 2010; Mulé et al., 2015; Nist & Joseph 2008; Volpe et al., 2011). A ratio of 0:6 was used in each session for the current study, where 0 represents the number of known sight words and 6 represents the number of unknown sight words.

**Direct Instruction.** The DI technique incorporates both known and unknown words into its procedures. Similar to TDP, the initial presentation of the unknown words includes
demonstration and prompting. A hallmark of the DI procedure is that the order of the flashcards is responsive to the student’s performance. More specifically, a word that is accurately read aloud within a fixed delay interval (i.e., two seconds) is placed at the back of the deck, whereas an incorrect response or hesitation (i.e., more than two seconds) results in the word being placed two to three cards back from the front of the deck. That word then enters “drill mode,” which means it will continue to be placed as the second or third card in the set each time it is presented to the student until the student accurately responds to the word three consecutive times.

Although there is some consensus in the literature pertaining to the previously mentioned procedure, the ratio of known to unknown words is inconsistent (e.g., Skarr et al., 2014; Ruwe, McLaughlin, Derby, & Johnson, 2011). Brasch et al.’s (2008) study also experimented with different ratios across sessions (e.g., 12:3, 10:5, 11:4, 8:7). The current study, however, used a 6:6 ratio in each session, where 6 represents the number of known words and 6 represents the number of unknown words. To allow for direct comparison to the other flashcard techniques, the

Figure 2.1. Schematic of the TDP procedure. Following the completion of each set presentation, the set of four unknown words is randomized and presented again. U = unknown sight word.
current study used a fixed delay interval of two seconds. Additionally, to maintain consistency in the current study’s procedures of DI, incorrect responses always resulted in the word being inserted as the third card of the set (see Figure 2.2).

Figure 2.2. Schematic of the DI procedure. At OTR 1, the student provided a correct response to K1. As a result, K1 is placed at the back of the set. The same occurs for K2 at OTR 2. At OTR 3, an incorrect response was given for U1. Therefore, U1 was placed as the third card in the set. Following the next two OTR, K3 and K4 are placed at the back of the set because correct responses were provided within three seconds. At OTR 6, U1 will again be placed third in the set regardless of the student’s correct or incorrect responding, as U1 will remain in “drill
mode” until the student accurately responds to the card three times, consecutively. Following the third consecutive accurate response to U1, the card will exit “drill mode” and be placed at the back of the set. At OTR 7, the student correctly responds and U2 is placed at the back of the set. OTR = opportunity to respond; K = known sight word; U = unknown sight word.

**Incremental Rehearsal.** Similar to the TDP and DI procedures, the IR technique begins with the interventionist using demonstration and prompting to teach unknown words. The flashcards are then presented in a fixed sequence in which a single unknown word is gradually paired with multiple known words (see Figure 2.3). After the completion of a trial, the unknown
Figure 2.3. Schematic of the IR procedure. The first set presentation pairs an unknown word with a known word. Subsequent set presentations incrementally fold-in one known word until a predetermined ratio of known to unknown words is reached (e.g., 5:1, 7:1, 9:1). In this example, U2 is introduced after the ratio of 5:1 is reached. Once that occurs, K5 is removed from the set and U1 becomes K1, as it is now considered to be known. Consequentially, K1 becomes K2, K2 becomes K3, K3 becomes K4, and K4 becomes K5. K = known sight word; U = unknown sight word.

This process continues for the duration of the intervention session until the predetermined number of unknown words are folded-in. Once all known words have been added to the set, Burns and Sterling-Turner (2010) argued that the unknown word should be considered learned if the sequence was completed without error. However, other studies have used retention probes, which were administered at the beginning of the next session, to determine whether a word should be considered learned (Mulé et al., 2015; Nist & Joseph, 2008; Volpe et al., 2011). A number of ratios for IR have been explored (e.g., 8:6, 5:3, 9:3, 9:6; Mulé et al., 2015; Nist & Joseph, 2008; Burns and Sterling-Turner, 2010; Volpe et al., 2011), but the current study used a 5:6 ratio, where 5 represents the number of known sight words and 6 represents the number of unknown sight words.

Comparative Studies

The problem-solving cycle, which includes assessing students’ discrete academic skills and selecting interventions that contain active treatment components that correspond with specific intervention targets, has been established as an evidence-based practice (Ardoin & Daly, 2007; Codding, Shiyko, Russo, Birch, Fanning, & Jaspen, 2007; Daly et al., 1996; Daly &
Martens, 1994; Tilly, 2002). This process of instructionally matching students’ skill level with the difficulty level of the materials used for intervention was originally proposed by Gickling and Armstrong (1978). In their study of first- and second-grade students, they sorted assignments into three categories based on each student’s skill level. Each student completed reading assignments containing material at the frustration-level (i.e., less than 90% known material), instructional-level (i.e., between 93% and 97% known material), and independent-level (i.e., more than 97% known material). Gickling and Armstrong (1978) observed the highest rates of on-task behavior, task completion, and task comprehension when students were provided with instructional-level material. However, the literature reflects a lack of consensus in terms of the optimal ratio of known to unknown material that should be assigned to maximize the effectiveness and efficiency of learning, particularly amongst flashcard intervention procedures. Simply put, it is currently unclear as to the number of known words (if any) to include in sight word interventions to maximize learning remains debatable.

Nist and Joseph (2008) compared the effectiveness and efficiency of IR and TDP. Each condition included six unknown sight words. The IR condition used a 9:6 ratio, where 9 represents the number of known sight words. Their findings indicated that students demonstrated significantly higher retention for the IR condition than the TDP condition. Conversely, the TDP intervention was more efficient, as students learned more words per minute when accounting for the amount of time required to complete the intervention. In terms of Nist and Joseph’s (2008) findings related to effectiveness, they are consistent with the IH, considering the IR condition possessed a folding-in component that provided students with more than three times as many OTR to target words per session (i.e., 169) than TDP (i.e., 54). Regarding efficiency, it is unsurprising that the IR condition was less efficient, as each session included 323
OTR to known and unknown sight words. Therefore, in reference to the current study, it was of particular importance to appropriately compare efficiency by equalizing the number of OTR across conditions. Matching either OTR or instructional time represents two different ways to control for obvious confounds when engaging in comparative research on flashcard interventions.

Volpe et al. (2011) made an effort to closely match the potential efficiency of IR and TDP by decreasing the number of unknown sight words to three and reducing the known to unknown ratio in IR to 5:1. Additionally, in a separate condition, the administration time for each procedure was held constant (i.e., three minutes). Volpe et al. (2011) stated that to fairly compare effectiveness, the folding-in component of the IR procedure was removed, which allowing the total number of OTR to remain constant for both conditions (i.e., 60). Consistent with previous findings, TDP was more efficient than IR when time was held constant. However, Volpe et al. (2011) found that both conditions were comparably effective when the number of OTR was held constant, which is at odds with the IH and inconsistent with the literature. Due to the inclusion of known words during the IR condition, it was expected that the TDP condition would be more effective when holding both time and OTR constant. Although Volpe et al.’s (2011) study appears to have produced a remarkable finding (i.e., comparable effectiveness between TDP and IR with OTR held constant) there are a number of methodological concerns that likely influenced the results. For example, the study may be critiqued for its heavily modified IR procedure (i.e., eliminating the folding-in component) and the learning cap placed on the TDP condition (i.e., three target words per session).

Aiming to clarify the disparate findings of previous research, Mulé et al.’s (2015) study compared the effectiveness and efficiency of TDP and IR with three target words per session.
The folding-in component was not removed from the IR procedure and the known to unknown ratio remained 9:1. The time for each intervention was held constant (i.e., three minutes) to assess efficiency. Their results indicated that TDP was more effective and efficient than IR, but it was also noted that, on average, the TDP condition provided nearly twice as many OTR on unknown words than the IR condition. As a result, it is unlikely that the second and third target words were even presented during most IR sessions.

From a cognitive psychology perspective, Varma & Schlesman (2014) argued that flashcard procedures that involve spaced practice are more likely to lead to accurate recall of items because they present unknown items in slightly different contexts (i.e., increased time between the presentation of unknown items) in comparison to TDP. However, Cates et al. (2003, p. 610) warned that, “minor instructional alterations that have intuitive appeal and appear to have no side effects, in fact may decrease learning when learning is measured with regard to instructional time.” Although support for Varma & Schleisman’s (2014) hypothesis exists (Burns & Dean, 2005; MacQuarrie, Tucker, Burns, & Hartman, 2002), Szadokierski and Burns (2008) explained that past studies have often conflated the influence of drill ratios with the influence of OTR within a predetermined period of time. In other words, comparative studies that have examined effectiveness have often attributed results to the differing drill ratios of the procedures. However, such studies have often failed to equalize the number of OTR, which has led to more favorable conclusions for interspersed techniques (e.g., Burns, 2004; Cates & Rhymer, 2003; Nist & Joseph, 2008). Therefore, the results of Volpe et al.’s (2011) study likely requires further validation from future studies that do equalize OTR across methods.

In terms of maintenance and generalization, the literature indicates interspersed flashcard techniques, particularly IR, are more likely to lead to higher rates of long-term recall and
stronger oral reading fluency skills (Dean & Burns, 2002; Roberts & Shapiro, 1996). However, similar to the issue of unequal OTR among studies that compared effectiveness and efficiency, these findings are better explained by Daly et al.’s (2002) study, which linked skill maintenance with OTR (MacQuarrie et al., 2002). That is, the majority of comparative studies may favor interspersed flashcard techniques over TDP simply because such studies provided more OTR for interspersed conditions. Therefore, students in those studies were likely better able to maintain and generalize their skills because they had a stronger foundation of accuracy and fluency through overlearning. This demonstrates further support for the IH, as generalization is facilitated by maximizing the OTR to develop adequate accuracy and fluency skills.

Beyond the comparison of outcomes, there is evidence that suggests that the successful completion of discrete tasks is often reinforcing for students (Skinner, 2002). Social validity, which is defined as the overall acceptability of a given intervention, has most often been found to be higher for interspersed procedures when compared to TDP (Cates & Skinner, 2000; Cates, Skinner, Watkins, Rhymer, McNeil, McCurdy, 1999; Skinner, 2002). The repeated practice of items that a student perceives as difficult, such as in the case of TDP, may lead to feelings of frustration. Consequently, Skinner and Shapiro (1989) found that students actually chose longer assignments with interspersed easy items (i.e., 3-6) and challenging items (i.e., 15) over shorter assignments with only challenging items (i.e., 15). Nist and Joseph (2008), however, found that students favored TDP because it took less time.

**Advancements in Flashcard Intervention Technology**

Education in the 21st century has recognized the utility and convenience of technology in promoting the academic achievement of students (Tamim, Bernard, Borokhovski, Abrami, & Schmid, 2011). Not only are current technological devices and associated applications now able
to be programmed to use empirically-supported approaches to instruction (e.g., demonstration, model, prompting, corrective feedback), they also have the capacity to monitor students’ progress and dynamically modify approaches based on student responding (Nakata, 2008; Pyc & Rawson, 2007). Although the use of technology in the classroom has not always been welcomed and supported by the educational community (Skinner, 1958), the integration of technology in the classroom has increased over the last several decades as teachers have become increasingly more confident and comfortable with using it (Liu, Ritzhaupt, Dawson, & Barron, 2017).

Considering teachers’ often demonstrate negative attitudes toward the implementation of evidence-based academic intervention within an RtI framework, the observed shift in educators’ acceptability of technology integration is promising. Moreover, the increased social validity of technology integration may prove to be a functional platform toward the promotion of empirically-supported computer- and tablet-based academic interventions for struggling students. It is reasonable to assume that the strength of any evidence-based academic intervention is inextricably related to the simplicity of its administration (Daly et al., 2005). In the same vein, relevant stakeholders should be wary of educational programs and applications on digital platforms that purportedly facilitate student learning in the absence of quality, peer-reviewed evaluation (Duhon et al., 2010; Hirsch-Pasek et al., 2015).

Computer-based flashcard interventions have demonstrated encouraging results for at-risk general education students, as well as students with limited English proficiency, attention-deficit/hyperactivity disorder, autism, and intellectual disabilities (Yaw, Skinner, Parkhurst, Taylor, Booher, & Chambers, 2011, Yaw et al., 2012). Beyond generalization across subjects, Yaw et al.’s (2012) study also showed generalization over time (i.e., eight week follow-up) and across stimuli (i.e., handwritten words on index cards). Additionally, both teachers and students
are more likely to rate computer- and tablet-based interventions higher on measures of social validity than traditional paper-based sight word intervention approaches (Jameson et al., 2012).

Chapter Summary

The review of the literature clearly reflects a growing interest in emerging technology-based academic interventions. It is also fortunate that the literature indicates educators are becoming increasingly more accepting of technology within the classroom because this shift holds potential for streamlining RtI processes, making it more likely that at-risk students receive necessary academic supports. However, as mentioned above, schools must exercise caution in their selection of technology-based instructional tools, as there are commercially available programs that may emphasize consumer appeal over the theoretically-grounded approach to instruction outlined in Haring and Eaton’s (1978) hierarchy.

In relation to technology-based flashcard intervention techniques to increase sight word recognition, theory-driven research has helped to identify the specific active treatment components that are necessary to maximize student learning potential (e.g., demonstration, prompting, corrective feedback) within three-term contingency trials. However, the effectiveness of a flashcard intervention may not mean much to educators if the intervention cannot be feasibly implemented in the classroom or is not perceived as socially valid. Therefore, contemporary research must account for these variables as traditionally paper-based interventions with abundant empirical support are adapted for use on modern technological devices. However, the momentum of comparative studies in the current literature must also be maintained, as continuing to investigate specific treatment components of various flashcard intervention techniques will further expand our understanding of how they interact with the
individual learning characteristics of students. Specifically, the current study sought to answer the following research questions:

1. Which flashcard intervention technique (i.e., TDP, DI, IR) was associated with the greatest number of acquired words (target words read correctly across two consecutive retention checks)?

2. Which flashcard intervention technique (i.e., TDP, DI, IR) was associated with the greatest number of words maintained (acquired words read correctly at two-week follow-up)?

3. Which flashcard intervention technique (i.e., TDP, DI, IR) was associated with the greatest percentage of maintained words read correctly at generalization assessment (paper-based presentations of words maintained)?

4. Which was the most efficient flashcard intervention technique (i.e., TDP, DI, IR) in terms of learning rate, skill maintenance, and generalization data?
Chapter 3: Methodology

Overview

The purpose of this chapter is to explain the research design of the study, including methods and procedures. In addition, a description of the participants and materials is provided.

Participants

Three second grade students from a suburban elementary school in the northeastern United States participated in the study. They were identified by their classroom teacher as having low academic achievement in reading. The school’s total population is 659, 47% of which are female and 51% of which are persons of color. Seventy-six percent of the students qualify for free or reduced-price lunch.

Materials

Assessment and instruction was conducted using an iPad-based application (i.e., Flash Fluency, v. 1.01, Kotak, Sinisi, & Solomon, 2019) specifically designed for the current study. Two hundred and twenty words used in the application were drawn from the Dolch Sight Vocabulary List. Similar to paper-based flashcard interventions, words were presented in a landscape orientation with a black font color on a white background. The interventionist allowed two seconds for the student to provide a response. To record an accurate response provided within two seconds, the interventionist tapped a circle at the bottom-left of the screen. Inaccurate or absent responses were recorded by tapping a circle at the bottom-right of the screen. Self-corrections (e.g., the student initially provided an incorrect response, but then immediately provided a correct response) that occurred within the two-second interval were scored as correct to avoid disrupting the pacing of the intervention. The presentation of the next sight word occurred as soon as the interventionist scored the response.
Pre-Intervention Assessment. A pre-intervention assessment (pre-IA) component of the application identified known and unknown words to be used for intervention sessions. The first 179 words from the Dolch Sight Vocabulary List (i.e., pre-k, kindergarten, first grade, second grade) were included in the pre-IA. The current study defined known sight words as words that the student consistently read accurately within two seconds across three pre-IA sessions. The order of words within the pre-IA were randomized across sessions. Unknown sight words are words that the student consistently failed to read accurately (i.e., incorrect response, absent response) across three pre-IA sessions. The application was designed to then use the student’s pre-IA results to establish sets of known and unknown sight words. Discrepant performance on sight words across pre-IA sessions (e.g., the student correctly read “apple” within two seconds during the first and second pre-IA sessions, but provided an incorrect response during the third pre-IA session) were removed and not included in the intervention phase. The application then distributed the student’s sets of known and unknown sight words across the three flashcard conditions. All three procedures drew from the results of the pre-IA results, but the application was designed to randomly assign each unknown sight word to only one of the three conditions: (a) TDP, (b) DI, and (c) IR. Therefore, any given unknown sight word did not appear in more than one condition.

Intervention. Each flashcard procedure targeted six unknown sight words on each day of instruction (i.e., 18 total unknown words targeted per intervention session). Following the first session, a retention check occurred prior to instruction for all subsequent sessions. The retention check involved the presentation of each target word reviewed in the previous session, without any feedback, to test whether the student was able to provide an accurate response within two seconds for each target word. The results from retention checks were used to inform
instructional decisions for same-day instruction. In other words, the application automatically determined which target words required further instruction and the number of new unknown words to introduce. This process included a two-step verification that required the student to accurately respond to previously reviewed target words across the retention checks of two consecutive intervention sessions. Sight words that were correctly read within two seconds for two consecutive retention checks were removed and new unknown sight words took their place. Therefore, data during the intervention phase was not available until the retention checks of the third intervention session was complete. Target words that did not pass the two-step verification continued to be included in instruction until they were learned. After the fifth intervention session, students were shown time-series graphs of their progress for each flashcard method prior to starting each intervention session.

**Post-Intervention Maintenance Assessment.** Following the last intervention session, the post-intervention assessment (post-IA) measured skill maintenance by presenting students with all previously unknown mastered sight words. The number of words from the Dolch Sight Vocabulary List included in each student’s post-IA was dependent on the number of target words that passed the two-step verification process. For each of the three conditions, skill maintenance over time was determined by comparing the total number of learned words within a condition to the number of correct responses provided within two seconds for that same condition on the post-IA. The post-IA occurred two weeks after the last intervention session. Three post-IA sessions occurred across three separate sessions within a seven-day timeframe so as to determine data stability.

To test generalization from tablet-based text to paper-based text, participants’ performance reading from a paper-based list of all words learned across intervention sessions
was also recorded. The words were printed in random sequence in 12 point Times New Roman font, arranged like an oral reading passage (i.e., text from left to right). Finally, generalization assessments were counterbalanced with maintenance assessments, as both were completed within the same session.

**Procedure**

The administration of the three flashcard procedures were counterbalanced across sessions and participants (e.g., TDP, DI, IR for the first session; DI, IR, TDP for the second session, …). Instructional time was recorded passively by the app and separated for each condition. After the retention check and initial presentation of target words, time was kept for each of the three conditions. The application was programmed to automatically start and stop timing at the beginning and end of each intervention session. Timing began when the first word of each procedure was presented and stopped once the student reached 120 OTR for each respective procedure. Instructional time across procedures was often similar because OTR was equalized. However, because the TDP procedure contained only unknown sight words, it often ran slightly longer due to the increased opportunities to receive corrective feedback following incorrect responses. The interventionist and participant were seated at the corner of a table or desk to allow the interventionist to more easily position the tablet in a manner that made the screen viewable for both parties. The interventionist initially prompted students for a response, and provided consistent verbal reinforcement for correct answers. This was thinned, however, after the first few trials as the students gained momentum and sped up their responses within session.

**Design and Data Analysis Procedures**
An adapted alternating treatments design nested within an AB design was used to compare the effects of TDP, DI, and IR (Sindelar, Rosenberg, & Wilson, 1985), where baseline was the pre-assessment of unknown words, followed by the intervention phase, which rapidly alternated conditions. The experimental control associated with the embedded rapid replication component of this design allowed for a high degree of confidence in determining the functional relation between each discrete intervention and their respective outcomes (Kratochwill et al., 2010). This design is considered fully experimental within subject, which has become the standard among empirical studies that have sought to compare the effectiveness of two or more flashcard methods (e.g., January et al., 2007; Joseph and Nist, 2006; Mule et al., 2015; Nist and Joseph, 2008; Volpe et al., 2011).

Experimental sessions occurred three school days per week for six weeks, typically Monday, Wednesday, and Friday. The number of intervention sessions ranged between seven and 14 among participants, which exceeds the five replications needed to meet What Works Clearinghouse evidence standards for alternating treatment designs (Kratochwill et al., 2010). Missed sessions (e.g., student absence, school closure) were made up the next day, to the greatest extent possible. The primary dependent variable, number of cumulative learned target words, was calculated by adding the total number of learned target words for each condition, summed across sessions, based on retention assessment performance. Students’ performance across conditions was then compared. A target word was considered learned if it was accurately read within two seconds on two consecutive retention checks.

The progress of each student was plotted on time-series graphs and subjected to visual and statistical analyses. Visual analysis targeted differences in slope, level, consistency across subjects, and variability (Kratochwill et al., 2010). Statistical analysis was conducted
concurrently across subjects using hierarchical linear regression to estimate the relative likelihood of identifying a given word correctly across conditions (see Moeyaert, Ferron, Beretvas, & Van den Noortgate, 2014; Shadish, Kyse, & Rindskopf, 2013). In this latter analysis, the higher the likelihood, the greater the success of the intervention.

**Treatment Integrity**

All sessions were completed by school psychology graduate students. A graduate student or faculty member collected treatment integrity data for approximately 20% of the intervention sessions. A checklist listing the sequence of each required step of the intervention administration session was used to record treatment integrity data. The training of interventionists consisted of an hour-long session with the primary experimenter. Training sessions included instruction and role-play. Treatment integrity was calculated for each participant by dividing the total number of steps completed by the interventionists by the total number of steps indicated by the checklist across intervention sessions.
Chapter 4: Results

The design of the study was executed as intended and, as a whole, the independent variable was effective in improving sight word knowledge for all three students. One of the conditions (i.e., Direct Instruction) experienced fatal technical errors during the administration. It was discovered mid-experiment that it was no longer working as intended. The data for that condition, therefore, were deemed invalid and will not be discussed further.

Descriptive Trends across Participants

Table 4.1 provides the total number of acquired sight words, average intervention session length, total time of instruction, and the rate of words learned per minute for TDP and IR. Additionally, the total number of words maintained, percentage of words maintained and the rate of words maintained for each condition are presented. The data revealed that the total number of words students learned using TDP was comparable to the total number of words learned using IR. Also, the cumulative learning rate for each condition was identical across subjects. On average, students acquired .35 words per minute of instruction and 3.63 words per session. In other words, both TDP and IR worked and did so quickly.

Together, Sarah, Madison, and Toby learned 61 words in the TDP condition and 59 words in the IR condition. Each student, on average, learned 20.33 words in the TDP condition and 19.67 words in the IR condition. In terms of skill maintenance, data from each condition was compared to determine whether a significant difference existed following two weeks without treatment. The data indicated students maintained words relatively equally in both conditions (on average, 61% of words from the TDP condition and 66% of words from the IR condition). When the students were asked to read paper-based sight words they had acquired during intervention,
### Table 4.1

*Number of Words Learned, Mean Time of Instruction, Total Time of Instruction, Rate of Words Learned, Number of Words Maintained, Percentage of Words Maintained, Rate of Words Maintained, and Percentage of Words Generalized for Each Student by Instructional Condition*

<table>
<thead>
<tr>
<th></th>
<th>Traditional Drill &amp; Practice</th>
<th>Incremental Rehearsal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W</td>
<td>MT</td>
</tr>
<tr>
<td>Sarah</td>
<td>22</td>
<td>5.38</td>
</tr>
<tr>
<td>Madison</td>
<td>10</td>
<td>4.37</td>
</tr>
<tr>
<td>Toby</td>
<td>29</td>
<td>5.69</td>
</tr>
<tr>
<td>Group</td>
<td>61</td>
<td>5.28</td>
</tr>
</tbody>
</table>

*Note:* W = number of words learned; MT = mean time of instruction; Time = total time of instruction (number of minutes); Rate = rate of learning (number of words read accurately/time); WM = number of words maintained; %WM = percentage of words maintained; MR = rate of words maintained (number of words maintained/time); %GZ = percentage of words generalized.
the group correctly read, on average, 79% of the words from TDP and 69% of the words from IR across three trials.

**Statistical Analyses**

Target words were considered acquired when a correct response was provided within two seconds during the retention check across two intervention sessions. Therefore, for each trial during retention checks, responses were recorded as either correct or incorrect. The dependent variable for this data is then a proportion (correct trials out of total trials) from a fixed number of binary (0, 1) observations. Therefore, a binomial distribution was assumed when analyzing the data and a binomial link function used to analyze the data. HLM estimates were produced on a logit scale and were converted to an odds ratio for interpretation. The main predictor, \( \beta_{10} \), condition, was dummy coded (0 = TDP, 1 = IR) to allow for comparisons across conditions of the likelihood of correct responding on a given retention check.

A parameter was also included to address possible change in the learning rate across sessions, \( \beta_{20} \). These were the only two predictors aside from the intercept. All predictors were permitted to vary across subjects. Results are shown in Table 4.2. None of the fixed effects were significant, as the odds ratio for each coefficient suggested no difference. Therefore, it can be concluded that the two intervention conditions were effective, based on the descriptives and visual inspection of the data, however, the effectiveness of the two conditions were not statistically different from one another. Further, the rate of learning appeared steady across intervention sessions. The random effects suggested that these findings were fairly constant across subjects.

**Visual Analyses**
Table 4.2

*Results of Non-linear Model with the Logit Link Function*

<table>
<thead>
<tr>
<th>Fixed effect</th>
<th>Coefficient</th>
<th>Standard error</th>
<th>t ratio</th>
<th>df</th>
<th>p</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{00}$</td>
<td>- 0.06</td>
<td>0.32</td>
<td>- 0.19</td>
<td>2</td>
<td>0.867</td>
<td>0.94</td>
</tr>
<tr>
<td>$\beta_{10}$</td>
<td>- 0.01</td>
<td>0.31</td>
<td>- 0.03</td>
<td>2</td>
<td>0.978</td>
<td>0.99</td>
</tr>
<tr>
<td>$\beta_{20}$</td>
<td>- 0.08</td>
<td>0.04</td>
<td>- 2.03</td>
<td>2</td>
<td>0.180</td>
<td>0.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effect</th>
<th>Standard deviation</th>
<th>Variance component</th>
<th>df</th>
<th>$x^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_0$</td>
<td>0.31</td>
<td>0.10</td>
<td>2</td>
<td>2.19</td>
<td>0.335</td>
</tr>
<tr>
<td>$r_1$</td>
<td>0.34</td>
<td>0.11</td>
<td>2</td>
<td>2.97</td>
<td>0.225</td>
</tr>
<tr>
<td>$r_2$</td>
<td>0.03</td>
<td>0.00</td>
<td>2</td>
<td>0.81</td>
<td>&gt; 0.500</td>
</tr>
</tbody>
</table>

*Note.* Values in this table are rounded to two decimals, but computations reported in text, such as conversions of log odds ratios to odds ratios and proportions, were done on results to six decimals. If such conversions are done on the numbers in the table, the results will differ due to rounding error. df = degrees of freedom.
**Sarah.** Visual analysis of Sarah’s data in Figure 1 shows a clear and immediate increase in sight word recognition with sufficient skill maintenance after two weeks without intervention. Her rate of improvement (ROI) during the intervention phases of each condition was comparable. On average, she acquired 2.11 and 2.38 words per session in the TDP and IR conditions, respectively. Sarah correctly read 18, 18, and 19 of the words she acquired from the TDP condition across the first, second, and third post-intervention assessments. Across the assessments, she read 16 of the 22 (73%) words she learned from the TDP condition consistently across all three assessments. From the IR condition, she correctly read 20, 22, and 20 learned words in the first, second, and third post-intervention assessments, respectively. Across the post-intervention assessments, she read 19 of the 25 (76%) learned words from the IR condition consistently across all three assessments. When presented with a randomized paper-based word list of all acquired words for three separate trials, she correctly read 91% of the words she learned from TDP and 80% of the words she learned from IR, on average. In total, Sarah received 131.52 minutes of total instruction across 24 intervention sessions, each taking an average of 5.24 minutes to complete.

![Graph](image)

*Figure 4.1.* Number of words Sarah acquired and retained across sessions.
**Madison.** As can be seen in Figure 2, Madison’s sight word knowledge increased immediately for both sets of stimuli. Visual analysis showed a significant increase in the number of sight words she could read within two seconds for both TSP and IR. Her ROI for each condition indicated she typically acquired 1.89 and 2.18 words per session from the TDP and IR flashcard strategies, respectively. She learned an average of .33 words per minute from TDP sessions and .43 words per minute from IR sessions.

Two weeks after the interventions were withdrawn, she was able to recognize most of the words she had acquired. Madison correctly read 9, 10, and 9 of the words she acquired from the TDP condition across the first, second, and third post-intervention assessments. Across the assessments, she read 8 of the 10 (80%) words she learned from the TDP condition consistently across all three assessments. From the IR condition, she correctly read 11, 12, and 13 learned words in the first, second, and third post-intervention assessments, respectively. Across the post-intervention assessments, she read 10 of the 13 (77%) learned words from the IR condition consistently across all three assessments. When presented with a randomized paper-based word list of all acquired words for three separate trials, Madison consistently recognized all 10 words.

*Figure 4.2. Number of words Madison acquired and retained across sessions.*
she learned from TDP and 12 of the 13 words she learned from IR. Combined, Madison received 60.92 minutes of instruction across 14 sessions, each taking an average of 4.35 minutes to complete. Compared to Sarah and Toby, Madison completed fewer intervention sessions, which translated to learning fewer words. This occurred because Madison fluently read more sight words than anticipated at baseline. As a result, there were not enough unknown sight words for Madison to continue past seven sessions for each method.

**Toby.** In contrast to the other two students, a visual analysis of Toby’s sight word acquisition (depicted in Figure 3) demonstrates a clear distinction between TDP and IR. While it is evident that growth was achieved in both conditions, his ROI from TDP (2.31 words per session) indicates he acquired more words per session than from IR (1.71 words per session). However, when considering the amount of instructional time over 14 sessions, the difference in intervention efficiency shrinks, as he learned an average of .36 words per minute from TDP and .29 words per minute from IR. This occurred because the total amount of instructional time Toby was engaged in TDP was 79.62, which was 9.47 minutes more than the 70.18 minutes he was engaged in IR. In terms of retention, Toby maintained approximately half of the words he acquired from both conditions.

Toby correctly read 21, 24, and 22 of the words he acquired from the TDP condition across the first, second, and third post-intervention assessments. Across the assessments, he read 16 of the 29 (55%) words he learned from the TDP condition consistently across all three assessments. From the IR condition, he correctly read 12, 16, and 15 learned words in the first, second, and third post-intervention assessments, respectively. Across the post-intervention assessments, he read 10 of the 21 (48%) learned words from the IR condition consistently across all three assessments. In terms of generalization, he consistently read 66% of the words he
acquired from TDP and 43% of the words he acquired from IR when the words were presented on paper. In total, Toby received 149.8 minutes of instruction across 28 sessions, each taking an average of 5.35 minutes to complete.

![Graph showing the number of words Toby acquired and retained across sessions.](image)

**Figure 4.3.** Number of words Toby acquired and retained across sessions.

**Treatment Integrity**

All graduate students that administered the interventions were observed by a licensed psychologist during at least 20% of sessions. The observer used a procedural checklist to monitor intervention fidelity (See Appendix). A checkmark was placed beside each applicable item when the instructor implemented a procedural step as indicated on the checklist. The treatment integrity data indicated the interventions were administered as they were intended to be administered 100% of the time.
Chapter 5: Discussion

The goals of the current study were to: (a) establish the effectiveness of three tablet-based sight word interventions (i.e., TDP, DI, IR) that have historically been evaluated in paper-based form and (b) compare the relative efficiency of the same three sight word interventions. Each tablet-based sight word intervention was administered to the three second-grade students that were identified by their teacher as being at-risk for low academic achievement in reading. Assessments of sight word recognition were completed on the tablet-based application, skill deficits were confirmed, and the results were used to inform instruction during intervention sessions. Six unknown sight words were targeted in each session for each condition (18 unknown sight words in total per session) and each condition took approximately five minutes to complete (approximately 15 minutes in total per day). Any given unknown sight word appeared in only one of the three conditions. Retention checks were completed prior to each session. Maintenance and generalization measures were administered two weeks following the last intervention session. Treatment integrity was collected using a procedural checklist during at least 20% of sessions and found to be very high.

Overall, results were positive. One condition, DI, was eliminated due to technical glitches in the software mid-intervention and will not be discussed further. For the other two conditions, acquisition, retention, and generalization of sight words was comparable to that of other studies that used paper-based cards and manual administration. Efficiency was comparable amongst the two remaining interventions.

General Impressions of Tablet-Based Intervention

The decision to adapt paper-based flashcards to a tablet-based application was made in an attempt to enhance the feasibility of implementing evidence-based academic intervention within
the classroom. Time is undoubtedly a precious commodity for all school-based professionals. Therefore, it is unsurprising that the amount of time required to prepare and administer an intervention significantly predicts teachers’ perceptions of an intervention in terms of its appropriateness, practicality, and effectiveness (Eckert & Hintze, 2000; Elliott et al., 1984; Gresham, 1989; Martens, Elliott, & Darveaux, 1985;). By eliminating the need for educators to focus on anything other than facilitating learning (e.g., preparation, timing, marking correct/incorrect responding, calculating learning rates, graphing student progress), the evolution of empirically supported intervention to technological devices holds promise for increasing students’ access to quality instruction.

Flashcard interventions, despite their abundance in the peer reviewed literature, are arduous to prepare. They require careful and repeated organization of hundreds of flashcards. Considering there was no preparation prior to intervention sessions or manual analyses to complete after intervention sessions during the study, the amount of time a student received intervention (approximately five minutes on average) would be virtually the same amount of time an educator would need to provide these evidence-based interventions to a struggling reader during the school day in applied practice. That is, the tablet did all the preparation, seamlessly preparing both assessment, intervention, and conducting instructional decision-making, thereby substantially reducing the amount of teacher time necessary to implement, and therefore likely increasing the likelihood of implementation fidelity.

In addition to increasing the feasibility of implementing intervention within the classroom, it is important to recognize that the tablet-based application was also created in a manner that closely aligns with empirically supported instructional components. A cursory search for applications currently available for download would show that there is a plethora of
applications that claim to target academic skill development (e.g., phonemic awareness, sight word recognition, math facts). However, educators must exercise caution against the allure of many technological tools that are not designed and tested to meaningfully impact student learning outcomes (Falloon, 2013). More and Travers (2013, p. 17) warned that “many of these [educational] apps may be lacking essential design, instruction, content, accessibility, and individualization features that are characteristic of high-quality education software.”

The application developed for the current study was designed to increase sight word recognition by: (a) maintaining the 1:1 support needed to maximize student success (Skinner, 2002), (b) including effective instructional components consistent with the Instructional Hierarchy (Daly et al., 2002; Haring & Eaton, 1978), and (c) using formative assessment to alter instruction (Sadler, 1989).

**Comparative Results Between Conditions**

As recommended by Volpe et al. (2011), the number of OTR (i.e., 120) was held constant for each intervention session, eliminating a confound that exists in prior studies. With OTRs controlled, the two conditions differed in two ways. First, IR included both known and unknown words while TDP included only unknown words. Because the number of OTR was held constant, TDP allowed for approximately 36% more three-term contingency trials on unknown words. Second, by design, IR included a folding-in component with known words, which allowed for spaced practice within session which is theorized to foster retention and promote engagement through frequent success (Dempster, 1991; Varma & Schleisman, 2014).

**Statistical Analyses.** Across participants, there was no significant difference between conditions in terms of the likelihood of correct responding on a given retention check. It is important to note that the odds ratios were not an indication of whether or not the either
condition worked more generally. Rather, the multilevel analysis showed that, for a given retention session, the likelihood of retaining a target word was about the same across the two conditions. Furthermore, the random effects showed that this was likely the case across participants. The multilevel analysis also showed that retention rates were comparable session-to-session within and across students.

**Visual Analyses.** Notable differences emerged for two of the three participants upon visual analyses of the data. For Toby, TDP appeared to be more efficient. The number of words acquired from TDP consistently remained above the number of acquired words from IR. Further, the gap between the number of acquired words for TDP and IR continued to widen over time. A visual analysis of Madison’s data indicated IR was slightly superior, although there were fewer data points to judge from compared to the other participants. For Sarah, there was no discernible difference between the efficiency of TDP and IR. That is, intervention effects appear idiosyncratic across students. In terms of maintenance and generalization rates for each condition, the data indicated similar results to those observed during the intervention phase across the three participants. These results are not necessarily inconsistent with the statistical analyses. The former speaks to accumulated acquired words, whereas the latter speaks to the probability of a given word being retained within session and it also considered variation in responding across sessions. As such, they speak to different aspects of the data.

Taken together, the statistical and visual analyses provide context to the interaction that may have occurred between intervention and participant. Consistent with the ranges in results reported in Volpe et al.’s (2011) study, the findings from the current study indicated that the success of each condition may have, in part, been determined by the individual learning characteristics of the student (e.g., attention, motivation, frustration tolerance). However, both
Intervention conditions resulted in clinically meaningful effects on sight word knowledge for all students. A final important observation was that these effects generalized to paper-based word lists, and thus students would likely apply this knowledge easily to passages and curriculum texts.

**Convergence with Prior Research**

In alignment with the recommendation of Cates and Rhymer (2003), the current study sought to compare the interventions in terms of efficiency, rather than more rudimentary comparisons of effectiveness (i.e., analyses of effects without considering implementation time). That is, the analysis of interventions were carefully controlled such that they presented growth rates in equivalent units of time. The identification of efficient interventions carries practical utility in maximizing instructional time for students (Skinner et al., 1996). Among studies that have examined the efficiency of TDP and/or IR (e.g., January et al., 2007; Joseph and Nist, 2006; Mule et al., 2015; Nist and Joseph, 2008), the results of the current study are most consistent with the instructional conditions of Volpe et al. (2011), in which instructional time was held constant. Although Volpe et al.’s study also examined TDP and IR conditions in which OTR was held constant, they defined OTR as opportunities to respond to unknown sight words. In the current study, OTR was defined as opportunities to respond to any word, both known and unknown.

Several limitations from Volpe et al.’s study include its heavily modified IR procedure (i.e., eliminating the folding-in component) and the learning cap placed on the TDP condition (i.e., three target words per session). Interestingly, the current study produced similar findings, despite addressing the methodological flaws identified in Volpe et al.’s study. However, differing conclusions may be drawn from each due to these procedural differences. The current
study retained the folding-in component of IR within sessions and doubled the number of words targeted per session, and yet achieved similar results. A plausible explanation for this is that ceiling effects occurred in both conditions, despite the larger set size (i.e., learning cap). In other words, it appears both the IR and TDP conditions may have reached their maximum effect given the number of OTRs presented, and thus additional OTRs served no beneficial effects for the subjects. If true, increasing the number of OTR and repeating the study may not produce any significant increases in acquired words for either condition. Rather, differences between conditions may be observed if IR and TDP were tested under the circumstance of fewer OTRs per word per session, although this may be dependent on the initial skill level and learning rate of the student. That is, every student has an optimal number of OTRs required to master words and going beyond this level serves little benefit (Burns, 2004; Burns & Dean, 2005; Szadokierski & Burns, 2008). Of great challenge is identifying this “sweet spot” for a given students and is an area in need of further research.

The average learning and maintenance rates of IR from the current study were relatively consistent with what has been reported in previous studies (January et al., 2007; Mule et al., 2015; Nist and Joseph, 2008; Volpe et al., 2011). However, the average learning and maintenance rates of TDP were lower than what would be expected based on the findings of the majority of past studies (Joseph and Nist, 2006; Mule et al., 2015; Nist and Joseph, 2008). A similar discrepancy also emerged when comparing the current study’s generalization data to those of Mule et al. (2015). The findings of prior studies, relative to the results of the current study, are presented in Table 5.1.

In terms of learning rates, the observed differences may be related to the current study’s more stringent procedure for target words to be considered learned. Joseph and Nist (2006)
considered unknown words to be learned following three accurate responses to a target word during intervention sessions. Most studies, however, deemed unknown words as learned following an accurate response on only one next-day probe (Mule et al., 2015; Nist and Joseph, 2008; Volpe et al., 2011), whereas the current study considered unknown words to be learned following correct responding across two retention probes. Therefore, consistent with the hypothesis that ceiling effects had occurred, it is possible that most target words continued to be reviewed during intervention sessions after they would be considered learned by the standards of the previously mentioned studies, stretching out instructional efficiency. In turn, this may have placed limitations on the students’ opportunities to be exposed to new target words. Importantly, these time estimates do not factor in intervention preparation time. It is likely that the present experimental conditions were superior simply because there was no intervention preparation required on a session-by-session basis, as was previously discussed. Further, time was precisely

<table>
<thead>
<tr>
<th></th>
<th>Traditional Drill &amp; Practice</th>
<th>Incremental Rehearsal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rate</td>
<td>MR</td>
</tr>
<tr>
<td>Current Study</td>
<td>.35</td>
<td>.21</td>
</tr>
<tr>
<td>Mule et al. (2015)</td>
<td>.83</td>
<td>.60</td>
</tr>
<tr>
<td>Volpe et al. (2011)</td>
<td>.37</td>
<td>.18</td>
</tr>
<tr>
<td>Nist and Joseph (2008)</td>
<td>1.13</td>
<td>.93</td>
</tr>
<tr>
<td>January et al. (2007)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Joseph and Nist (2006)</td>
<td>2.79</td>
<td>2.38</td>
</tr>
</tbody>
</table>

*Note: Rate = mean rate of words learned (number of words read accurately/time); MR = mean rate of words maintained (number of words maintained/time); GR mean rate of words generalized (number of words generalized/time).*
measured in the current study, as it was tracked by the app itself. It is not known how reliable these estimates are for the comparative studies, as observer agreement rates were not reported.

The discrepancies of maintenance and generalization results from the current study compared to other studies may be due to differences in the period of time between the last intervention session and follow-up maintenance and generalization assessments. Joseph and Nist (2006) administered maintenance assessments one day after the last intervention session. Nist and Joseph (2008) completed maintenance and generalization assessments five and six days following the last intervention session, respectively. Other studies (January et al., 2007; Mule et al., 2015; Volpe et al., 2011) waited a period of one week to administer maintenance and generalization assessments. The current study, however, completed maintenance and generalization assessments two weeks following the last intervention session. Considering the maintenance and generalization data of the IR condition in the current study are comparable to the same measures that took place at one-day and one-week follow-ups in other studies, this may indicate that the long-term stability of IR’s maintenance and generalization remains relatively constant.

By comparison, the maintenance and generalization data of the TDP condition in the current study did not mirror the sizeable superiority over IR that has been demonstrated in the literature (January et al., 2007; Joseph and Nist, 2006; Nist and Joseph, 2008). That is to say, although TDP may be deemed to be the more efficient method for improving sight word recognition within one week of the conclusion of intervention, the current study’s findings suggest students may not consistently retain the skills learned from TDP at two-week follow-up. Therefore, when considering past studies, the lower-than-expected maintenance and
generalization results from the TDP condition may indicate the efficiency of TDP and IR equalize over time.

To summarize, similar to the results of Volpe et al. study, both TDP and IR were effective in increasing sight word recognition, but TDP underperformed in comparison to other studies (Joseph and Nist, 2006; Mule et al., 2015; Nist and Joseph, 2008). Although similarities in results were observed when compared to the results of Volpe et al., differences in methodologies indicated the more stringent procedure for deeming target words learned and potential ceiling effects may have stunted learning rates for both conditions. Lastly, the longer duration of time between intervention sessions and follow-up assessments in the current study in comparison to the periods of time reported in other studies (January et al., 2007; Mule et al., 2015; Volpe et al., 2011) revealed that the skills gained from TDP may fade over time, at least relative to past studies, whereas the skills gained from IR appears to hold steady. However, because these comparative studies were single-case, there may have been idiosyncratic differences across methods, subjects, and experimenters that cannot be specifically isolated without further study.

**Future Research**

To accurately assess children’s social validity of flashcard techniques, they need to be able to distinguish between conditions. At the conclusion of the current study, it was realized that the elements of the research design that aimed to make the conditions comparable (e.g., counterbalancing, equal number of OTR, similar fonts and presentation) likely would have interfered with the children’s ability to indicate their preferences for each. Consequently, researchers may want to consider pairing specific visual cues prior to and following each flashcard procedure during each session. For example, researchers could introduce TDP as the
“star” method and present a star symbol before and after each administration of the flashcard technique. Similarly, researchers could introduce IR as the “diamond” method and present a diamond symbol before and after each administration.

Alternatively, an adaptation of Volpe et al.’s (2011) approach would involve the use of colored backgrounds for each condition to assist participants in discriminating among methods. By doing so, participants would be better able to complete measures of social validity at the conclusion of the study. If future studies also find that the tablet-based versions of TDP and IR are comparable on all measures of effectiveness and efficiency, social validity data may become a critical factor for educators to consider when selecting interventions.

The feasibility of integrating the administration of an intervention into teachers’ daily schedules should always be considered in the development of targeted academic interventions (Eckert & Hintze, 2000). The current study utilized graduate students to implement the interventions. It would likely be beneficial for future research to replicate the current study with teachers administering the interventions within the classroom. Such studies would provide invaluable information related to the real-world generalization of the current study’s findings.

Peer-assisted learning has long been recognized as an efficient and practical solution to providing 1:1 instructional support to more students in less time (Cook, Scruggs, Mastropieri, & Casto, 1985; Skon, Johnson, & Johnson, 1981). Treatment integrity data from the current study indicated all interventionists were able to administer both conditions with complete fidelity. Therefore, researchers may consider investigating the effectiveness of these tablet-based interventions when using peer-assisted learning within the classroom. Although peers may not be able to provide consistently accurate feedback following responses, this may be remedied by including a function within the application that would allow students to listen to the
pronunciation of words immediately following student responses. That way, peer tutors would be able to listen to the recording of any given word to: (a) determine whether the tutee’s response was accurate, and (b) provide praise and corrective feedback for correct and incorrect responses, respectively. This may also result in incidental benefit to the tutor.

Measures of generalization are important for all studies that examine the effectiveness and efficiency of flashcard intervention techniques (Musti-Rao et al., 2014; Yaw et al., 2012). To assess for generalization, the current study used paper-based word lists of all acquired tablet-based words for each participant. While the results indicated that the participants’ learning did generalize to paper-based words, researchers may want to consider examining the generalization of words that were acquired using tablet-based flashcard techniques to paper-based sentences or passages. This may include target words embedded in narrative text or a measure of general oral reading fluency (Ardoin, Christ, Morena, Cormier, & Klingbeil, 2013; Daly & Martens, 1994; Daly, Martens, Kilmer, & Massie, 1996; Daly, Murdoch, Lillenstein, Webber, & Lentz, 2002).

Despite the demonstrated success of the DI condition’s functionality during an earlier pilot study, a programming error in the application was observed after several intervention sessions had been completed. Data from the DI condition was deemed unusable and was excluded from analysis for the current study. As a result, a gap remains in the literature. It would therefore be worthwhile to repeat the current study following a software update that fixes the DI programming error.

Summary

As academic intervention techniques adapt to maintain pace with the integration of technology in the modern classroom, educational researchers must assist in ensuring educators have access to peer-reviewed tablet- and computer-based evaluation research. Studies that both
verify the efficiency of adapted intervention techniques and build upon the research base are vital in continuing to provide educators with evidence-based tools that maximize instructional time and facilitate student learning, particularly at a time when instructional technology is so thinly researched (Duhon et al., 2010; Hirsch-Pasek et al., 2015). Tablet- and computer-based interventions hold great potential for effectively and efficiently remediating skill deficits using feasible, user-friendly practices. The current study showed that such interventions, when properly designed and in alignment with extant research on best practice in instruction and intervention, are both effective and efficient when used to facilitate sight word instruction. Nonetheless, more research is required to better understand how such technology can be effectively integrated into the classroom.
Appendix

Integrity Checklist

<table>
<thead>
<tr>
<th>Student:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Interventionist:</td>
<td></td>
</tr>
<tr>
<td>Date:</td>
<td></td>
</tr>
</tbody>
</table>

- Interventionist and student sat on opposite sides of a table or desk corner.
- Tri-fold stand was used to keep the iPad resting on the table for the duration of the session.
- The intervention procedures were administered in the order specified in the counterbalancing schedule for the current session.
- Student was shown her or his graphed progress prior to starting the retention check of each method.
- Directions and prompts were delivered verbatim.
- For each retention and intervention item, the student was allotted no more than two seconds to respond.
- Praise and/or corrective feedback were not provided on retention check items.
- Praise was provided following correct responses on intervention items.
- Whole-word corrective feedback was provided when students responded incorrectly or did not respond within two seconds on intervention items.
- The verbal prompt “What’s this word?” was delivered prior to presenting each of the first 10 words for each procedure.

<table>
<thead>
<tr>
<th>Number of Items Completed:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of Items:</td>
<td>10</td>
</tr>
</tbody>
</table>
References


