Single Versus Repeated Demonstration Criteria on Reading Fluency and Retention

Lacy M. Knutson

A Dissertation Submitted to the Faculty of

The Chicago School of Professional Psychology

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in Applied Behavior Analysis

Single Versus Repeated Demonstration Criteria on Reading Fluency and Retention

---

A Dissertation Submitted to the Faculty of

The Chicago School of Professional Psychology

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in Applied Behavior Analysis

---

Lacy M. Knutson

2018

Approved By:

Julie Ackerlund Brandt, PhD, Chairperson
Assistant Professor, The Chicago School of Professional Psychology

---

Susan Flynn, PhD, Member
Assistant Professor, The Chicago School of Professional Psychology

---

Kathryn Kalafut, PhD, Member
Department Faculty, The Chicago School of Professional Psychology

Acknowledgements

I would like to express my sincerest thanks and offers of gratitude to my dissertation chair, Dr. Julie Ackerlund Brandt, and my committee members for their advisement, guidance, and feedback on this work. This process would not have been achievable without you. I want to thank Andrea Murray for her assistance in data collection and analysis, and Dr. Rick Kubina and the Chartlytics team for the use of Chartlytics to monitor and display my data. My thanks go to the families and participants who kindly donated their time to participate in my research. To my support system, Laura Lapen Parkhurst and Conny Raaymakers, you ladies have become family and I want to thank you for being a part of this journey with me. I want to thank Recairdo Mangrum for his steadfast support and strength, and for pushing me to succeed when I needed it. Finally, I would like to send my appreciation to my family members, especially my father and brother, who believed in me, supported me, and never let me give up on myself.

Abstract

Learning has been defined as changes in performances over time (Graf & Lindsley, 2002; Johnson & Street, 2013; Kubina & Yurich, 2012). Learning is not, however, a passive activity. It requires systematic programming and proper evaluation for success. However, in the educational and behavior analytic literature, definitions of terminal learning outcomes or mastery performance have varied dramatically. Recently, the behavior analytic literature has revived its exploration of the topic. The present study sought to expand this area of literature, specifically by manipulating the frequency demonstration requirements of a predetermined level of performance. Results indicated that all participants met preselected high-level terminal fluency aims, supporting prior literature from both the educational and behavior analytic fields. Minimal differences were observed across the two frequency demonstration requirements manipulated, suggesting that valuable educational time can be saved by setting high mastery criteria and requiring only a single demonstration before moving onto more complex behaviors. Additionally, this study explored the role of participant preference regarding evaluative procedures and found single demonstration criteria to be most preferred. Future research should continue to explore the concept of mastery, its role in educational and non-educational activities across various populations, and the role of student preference in evaluative measurement systems.

Table of Contents

List of Tables

List of Figures

Chapter 1: Nature of the Study

Background

Learning has been described as changes in behavioral frequency that occur over time (Graf & Lindsley, 2002; Johnson & Street, 2013; Kubina & Yurich, 2012). These behavioral changes may be either an increase (e.g., increase in frequency of words read correctly) or decrease (e.g., reduction in calling out during class) in the frequency of responding; however, how one defines the conclusion of the learning phase (e.g., mastery) and eligibility to begin more complex material has varied. For example, researchers have explored the various uses of subjective terms such as, "Try your best," to more stringent and clear definitions regarding one's level or quality (i.e., 50%, 80%, or 100% correct) in defining mastery performance (Johnston & O'Neill, 1973; Keller, 1968; Semb, Hopkins, & Hursh, 1973). A preselected level and frequency of performance provides direction in educational decision making regarding if a skill is at "mastery level" (Bloom, 1968; Fienup & Brodsky, 2017; Motamedi & Sumrall, 2000, Zimmerman & Dibenedetto, 2008); however, only recently have researchers in behavior analysis revisited the variables effecting mastery (Fuller & Fienup, 2018).

Two dimensions of behavior are commonly included in a behavioral definition of mastery. These include the level of performance at which an individual must achieve, and a frequency at which this level must be displayed (Fuller & Fienup, 2018). The dimension of behavior that has had the most analysis is on the level of performance. Both the educational and behavior analytic fields have explored this concept (Bloom, 1968; Johnston & O'Neill, 1973; Keller, 1968; Semb, 1973); and although each of these works explored the concept of mastery from different levels of performance (e.g., 50% vs. 90%), the frequency at which this performance was displayed was held constant. Most often, participants only needed to display

the prescribed level of performance once before moving onto the next phase or material. This is limiting, as it fails to account for repeated displays of performance at a specific criteria level.

Another commonality of these studies includes the selected system of measurement. The majority of previous studies used a percent correct accuracy ratio to measure performance (Johnson & Street, 2013; National Center for Education Statistics, 2011; White & Haring, 1980). Although a common measurement system in the educational setting, percent correct is not without limitations. First, percent correct measures provide an incomplete picture of a learner's performance. A percentage is a ratio of performance established by the number of times a certain response occurred as compared to the total number of opportunities, and multiplying it by 100 (Johnston & Pennypacker, 2009; White & Haring, 1980). Therefore, measuring accuracy in isolation reduces instructional decision making, and distances educational data from the sensitive measurement that was the foundation of the experimental analysis of behavior (Binder, 2011).

Second, researchers have demonstrated that accuracy alone, regardless of level, is a poor predictor of mastery (Kubina & Yurich, 2012; White, 1985). For example, accuracy cannot provide information with regard to how fluently the learner can emit the skill. Responding taught to a level of fluency is often said to appear effortless, automatic, and graceful and is naturally reinforcing (Binder, 1996; Johnson & Street, 2004; Kubina & Yurich, 2012). A concert pianist, a professional football player, and a spelling bee champion are all said to perform fluently. The concept of fluency in educational performance can be traced back to O. R. Lindsley, a doctoral student of B. F. Skinner's who founded the field of Precision Teaching. Lindsley sought to extend the work of operant conditioning to the educational setting, carrying forth the primary datum of rate of responding to measure behavioral growth and emphasizing a sensitive data monitoring system (Binder, 2011; Lindsley, 1963, 1964, 1991, 1996). Considering that all

behaviors occur in time, Lindsley emphasized the role of timeliness of behavior in addition to performance accuracy. When performances are fluent, they are effortless, and this effortlessness is considered a hallmark of performance mastery (Binder, 1996, 2011; Johnson & Street, 2004).

A broadly accepted goal within education is that students will acquire the information presented, and that this knowledge will have long-lasting effects on their future (Johnson & Street, 2013; Wood, Murdock & Cronin, 2002). Although it is widely accepted that building accuracy is a necessary piece to the learning process, accuracy alone does not result in performance retention (Kubina & Yurich, 2012; White, 1985). There have been ample demonstrations that building a skill to fluency does result in beneficial learning outcomes such as improved retention (Binder, 1996; Fabrizio & Moors, 2003; Johnson & Street, 2004; Kubina & Yurich, 2012); therefore, accuracy building is only "the first step in mastery" (Johnson & Street, 2004; p. 106). Considering this limitation, additional measurement strategies, such as rate of responding (i.e., fluency), should be explored in the context of defining mastery.

To successfully transition students through the educational process, a clear definition of mastery-level performance must be explored, as well as the variables that contribute to it. For example, a clear level of performance must be obtained, and the frequency of demonstrations at that level should be set. Additionally, the extended effects of the various performance levels and frequencies will provide additional valuable information to the concept of skill mastery. Finally, if the goal of education is to develop self-sufficient lifelong learners, the role of student preference in the evaluation process should be taken into consideration.

Purpose of the Study

The purpose of the current study was to extend the behavior analytic literature base on the role of defining mastery criteria by evaluating the frequency at which a set level of performance must be displayed in the development of reading fluency and the emergence of fluency retention. This was done by evaluating celerations of frequency building towards preselected terminal criteria under single versus repeated demonstration requirements, evaluating the impact of single versus repeated demonstration requirements on fluency retention, and evaluating the preference for demonstration conditions of child participants. By doing so, this study built upon the literature on mastery learning theory, applied behavior analysis, and precision teaching. Additionally, it extended the literature on mastery by examining the effects of mastery criteria on the emergence of skill retention and participant choice.

Summary

The role of defining mastery performance is one closely related to the habilitative and rehabilitative fields of education and applied behavior analysis (ABA). Therefore, it was, and continues to be essential that a clear understanding of the concept of mastery be explored, as well as the variables leading to the highest levels of performance outcomes. By examining the effects of the frequency at which a level of performance is displayed, the concept of mastery can be better understood, increasing the success of educational and therapeutic programs. By taking participant preference for evaluative measures into consideration, educators can not only be successful in progressing students through curricula, but may do so in a way that is highly acceptable by those being evaluated.

Chapter 2: Literature Review

Introduction

To effectively discuss the concept of mastery and its role within ABA, an analysis from the educational field must first be explored. Developers of the mastery learning approach (e.g., Bloom, 1968; Carroll, 1963) are historically relevant, as many of their concepts continue in defining mastery in today's educational system (Zimmerman & Dibenedetto, 2008). Furthermore, much of their work can be traced to the improved understanding of human behavior derived from early behaviorists such as B. F. Skinner and Fred S. Keller (Motamedi & Sumrall, 2000). In the early 1960s, behavior analysts explored the concept of mastery (Keller, 1968); however, after approximately a decade, the research in this area slowed, leaving many questions unanswered. It wasn't until more recently that behavior analysts again began systematically exploring the role of mastery (Fienup & Brodsky, 2017; Fienup & Fuller, 2018).

Defining Mastery from an Educational Perspective

The concept of mastery has been discussed most frequently from the perspective of mastery learning (Bloom, 1968; Carroll, 1963); however, some of the basic principles involved in mastery learning can be traced back to ancient philosophers such as Aristotle (Motamedi & Sumrall, 2000). Traces of mastery learning can be found in the educational system as early as the 1920s, and much of its revival can be attributed to behaviorists (Motamedi & Sumrall, 2000). A prime example is Keller's (1968) work in programmed instruction.

According to Bloom (1968), the educational evaluation system was failing our students. Broadly speaking, students were classified into five categories of performance, and these categories were based on a normal curve which inherently restricts the majority from performing at the highest rank (e.g., A). As such, this provides students with a preconceived notion that only

a select few can perform successfully in the educational setting, leaving the remaining students to become accustomed to being labeled as "average." Furthermore, student outcomes that deviate from this pattern are often attributed to the nature of the instructor being "too tough" or "too easy," rather than a reflection on student performance.

Bloom (1968) also discussed the limitations of the use of a *normal distribution* in the role of education. The normal distribution is mathematical construct and most appropriate when describing chance levels of performance of a group (Bloom, 1968; Herrnstein & Murray, 1994). Learning, especially in an educational system, should not be considered a chance event; rather, learning must be designed and programmed in order for an educational system to be described as effective. An educational system unnecessarily limits students' access to reinforcement when a distribution that only allows a minority to reach the highest level of performance possible dictates educational decision making. Bloom (1968) stated that the school setting must be rewarding for an individual to ensure the proper development of lifelong learning skills. Mastery learning, according to Bloom (1968), eliminates the shackles of the present educational system and creates an opportunity for the majority of students, rather than the minority, to attain high levels of performance in the classroom.

In the era of high-stakes testing-focused initiatives in the educational system, such as No Child Left Behind, Zimmerman and Dibenedetto (2008) compared two common evaluation methods, psychometric and criterion mastery, and appeared to share the distaste for the use of normed assessments. The authors suggested that determining mastery criteria based on large samples, as in the psychometric model, is limiting to the individual performer. It would be a fallacy to make inferences regarding individuals based on aggregated group data (Herrnstein & Murray, 1994). With the increased pressure of national assessments, there is a strain on educators

as they struggle between providing instruction based on student need and teaching to a test. In a review of the literature, Zimmerman and Dibenedetto (2008) found several positive results in favor of the mastery learning approach; a few of the key variables include: (a) the role of a clearly defined level of mastery (most often set at 80% correct or above), (b) providing frequent formative assessments to monitor learning and guide feedback, and (c) the decrease in performance discrepancies across students. Additionally, they reported on a case study of one school which adopted a mastery learning model, and was subsequently designated as a Blue Ribbon School for its improvements in student performance. Teacher interviews indicated that students who previously struggled, flourished under the new contingencies and there was a reported increase in student self-confidence. Student interviews supported these claims and indicated that the clear communication regarding what level of performance was necessary was helpful to the learning process (Zimmerman & Dibenedetto, 2008).

In brief, mastery learning consists of the following elements: a clear and objective definition of mastery level performance, frequent assessments to monitor progress, feedback delivered based on progress, and an emphasis on individualized instruction and pacing. A review of these elements bears a strong resemblance to components found in the field of behavior analysis; therefore, it is not surprising that the field dedicated to the science of human behavior has also explored the role of mastery and its effects on skill development.

Defining Mastery from a Behavior Analytic Perspective

Keller (1968) was one of the first behavior analytic researchers to examine mastery, its operational definition, and its effects on operant behavior. Keller recognized the need for individualized instruction and the ability for all individuals to be able to obtain mastery of material under the appropriate instructional environment. University-level courses functioned as

the setting for these early analyses and many subsequent examinations. One critical element of the learning arrangement was the opportunity for all students to become expert-level performers with the presented material. Although he supported Bloom's (1968) view that all students can achieve mastery, Keller's approach ran contrary to traditional academic programs. Keller recounted early applications of a personalized system of instruction (PSI) in university courses in the early- and mid-1960s. Of the five main features of this learning approach, he emphasized the role of 'unit perfection." Unit perfection required an individual meet a certain level of performance mastery (100%), as determined by a proctor as "passing" on a content area, before proceeding to the subsequent area of content.

Analysis of the outcomes of this learning method included several interesting features. First, the course-grade distribution did not resemble the normal distribution as is found in traditional courses. Rather, grade distribution shifted strongly towards the upper grade limits, with more students passing the course with an A, and fewer students receiving grades of either B or C. Keller (1968) concluded that if students who were traditionally labeled as "average" or "inferior" under a conventional model of instruction were provided ample opportunities to learn, their overall grades would improve. This was in part due to the repeated testing option made available by the proctors. Furthermore, arranging the course contingencies to encourage performance improvement and progress based on the individual's current level supported learning, versus relying on a standard course syllabus and schedule; this also allowed students who were successful with the material to move through the material quickly without the potential of becoming "bored." Additionally, those struggling with the material could take additional time to fully grasp the content before transitioning to more complex material. These outcomes supported Keller's (1968) view that the role of analyzing academic outcomes should focus less

on seeking course outcomes reflective of a normal distribution, and instead seek instructional strategies that increase individual learning outcomes (e.g., mastery).

Several replications have examined various elements of defining mastery that followed Keller's (1968) application of behavioral technologies in the university setting. For example, Johnston and O'Neill (1973) examined the role of performance criteria on student academic performance. Citing Skinner (1968), the authors claimed that the primary role of education is to obtain desired changes in student performance, and to observe such changes required a thorough analysis of the relationship of student performance and instructional methods which resulted in desirable change. Specifically, the role of course grades should be examined, along with definition of grade-level performance and the effects exerted on student performance.

Consistent with Keller's (1968) work, a university course was the setting of Johnston and O'Neill's (1973) experiments with 65 undergraduate students. A total of five experimental manipulations occurred with rate of student performance serving as the primary dependent variable. Under the first experimental condition, no teacher-defined rates were provided, as is consistent with traditional methods of instruction. Rather, students were told that all should provide their best work and that final grades would be based on the course curve. Remaining experimental conditions consisted of variations of defining grade-level performance criteria and sequences. High, medium, and low rates were defined (3.8 correct/0.4 incorrect per min; 3.1 correct/1.1 incorrect per min; 2.5 correct/1.7 incorrect per min, respectively) and were roughly equated to 90%, 75% and 60% correct from an accuracy ratio standpoint. Based on the results of the five experiments, Johnston and O'Neill provided the following conclusions: (a) clearly defining mastery rates reduced variability in performance across students, (b) traditional grade assignments based on a course curve result in higher performance variability, and (c)

performance followed the contingencies of reinforcement. These outcomes supported those of Keller (1968) that all students, if given the proper instructional environment, have the potential to meet mastery levels of performance; meaning, variability in student performance was minimized when clear definitions of what level of performance was required for mastery. Additionally, and perhaps more importantly, performance allocation clearly shifted with the related contingencies of reinforcement. For example, when the definition of mastery was defined with 60% correct receiving a grade assignment of A, student performance dropped to that level regardless if they had previously demonstrated higher performance capabilities. This outcome suggests that when defining mastery-level performance, instructors should require high levels of performance.

Since the 1960s, the effectiveness of other levels of performance have been analyzed. Semb, Hopkins and Hursh (1973) defined mastery as 90% rather than requiring perfect accuracy (100%). In their study, baseline conditions consisted of students earning points for each correct answer. This condition was compared to a differential reinforcement of other responses (DRO) condition and a noncontingent point condition. In the DRO condition, students were informed they would receive points for incorrect responses, whereas points were awarded noncontingently to quiz performance as long as the student was in attendance for class in the noncontingent point condition. Results indicated that under the DRO conditions, correct responses decreased to below-chance levels of performance, to merely 10% correct. This is not surprising considering the reinforcement contingencies for incorrect responding. Additionally, when noncontingent conditions were imposed, an average drop of 16 points was observed, meaning, when the functional relationship between performance quality and outcome (reinforcer delivery) was eliminated, responding allocated accordingly. Students only performed at the level necessary to

meet the mastery criterion. These results suggest that when defining mastery, instructors should base performance off the lowest acceptable rate.

Semb (1974) extended earlier work on examining the role of various mastery criteria (Keller, 1968; Semb et al., 1973). Using a within-group reversal design, Semb (1974) evaluated the effects of mastery criteria and assignment length on student progress and performance. The length of assignment was explored to extend upon Keller's (1968) indication that shorter units of instruction result in superior performance. Similar to previous literature, the participants consisted of 51 students enrolled in undergraduate courses, divided into two groups, each with four content units (Semb, 1974). Each unit included an assigned reading followed by a short-answer essay. Performance was evaluated by percent-correct on unit content and review quizzes. Three criteria manipulations were examined: high-criterion short assignment, low-criterion short assignment, and high-criterion long assignment. High-criterion conditions required a mastery of 100% correct, whereas the low-criterion conditions only required a performance of 60% correct. Short assignments consisted of obtaining the mastery level on both content and review quizzes. This arrangement was considered short, as testing occurred more frequently and on short content quizzes. Long assignments were defined as obtaining the mastery level on the review quiz, while the content quizzes after each unit were excluded.

The results of the study were consistent with earlier literature; Semb (1974) indicated that high mastery criterion requirements resulted in superior performance over low mastery criterions. Low mastery criterions resulted in inferior performance on both content and review quizzes (range: 76.1% – 91.1%) as compared to high mastery criterions (range: 79.6% – 99.2%). The results from this study, in addition to the prior literature, continue to support the use of high criterions when defining mastery for academic performance (Haughton, 1972; Johnston &

O'Neill, 1973; Keller, 1968; Semb et al., 1973). In examination of assignment length, long-assignment conditions resulted in increases of quiz retakes as compared to short-assignment conditions. Semb's (1974) results supported Keller's (1968) hypothesis that shorter assignments result in superior performance as compared to longer assignments. Overall, conclusions by Semb (1974) included the notion that instructors aiming to maximize student performance should use short assignments, and employ a definition of mastery which requires a high level of performance accuracy.

Additional extensions of the mastery criteria literature include the work of Carlson and Minke (1975) on fixed versus ascending mastery criteria. Prior literature indicated that in addition to the many benefits of mastery learning courses, a subset of participants withdrew or failed to maintain a sufficient rate of progress on unit content to successfully pass the course. Carlson and Minke (1975) addressed this limitation by establishing a clear indication of conditions in which successful completion of quizzes increased when presented in a mastery learning college course. Each course was exposed to a different mastery criteria (fixed 90%, fixed 80%, and ascending criteria). The ascending criteria condition schedule consisted of earlier units in the sequence defining mastery with lower performance requirements (60%) and subsequent units increasing the performance requirement (70%, 80%, and 90%). Results indicated that defining mastery at a moderate level (e.g., fixed 80%) resulted in (a) larger proportions of the participants earning a high-passing course grade (A), (b) faster progression through the material, and (c) a larger percentage of the participants successfully completing all units, as compared to those participants in either the high mastery criteria (e.g., fixed 90%) or ascending criteria conditions. One notable result from the ascending condition was that participants were more likely to successfully complete the quiz on the first attempt and required

fewer overall quiz attempts as compared to the other conditions. These results differed from earlier literature emphasizing the need to establish high mastery criteria from the start; however, Carlson and Minke cautioned that these results may not match the overall goal of education, and that lowering the mastery criteria required may not accurately reflect if students who completed the course had truly mastered the content.

After several decades, the topic of defining mastery re-emerged with Fienup and Brodsky (2017). In their study, block and rolling mastery criteria were examined again in the university setting. Block mastery was defined as using a learner's aggregate performance accuracy as the determinant to phase termination. Contrary to block mastery, rolling mastery was a predetermined number of consecutive trials performed correctly. In brief, this could be summarized as comparing an overall performance accuracy (block) to a performance accuracy and frequency (rolling) definition of mastery. The comparison of a large block criteria, which included 12 consecutive correct, was compared to a less stringent, six consecutive correct, and results supported the use of the more stringent (12 consecutive) mastery criteria. Furthermore, when comparing mastery criteria form (block versus rolling), the authors found no significant difference; however, they did report that 72% of participants in the more stringent rolling mastery condition demonstrated immediate emergent relations as compared to 63% of participants in the block condition. The authors suggested that future research should continue to explore stringent mastery criteria, and analyze the generality of these findings on different skill types.

Fuller and Fienup (2018) evaluated the effects of mastery on response maintenance. The authors described two dimensions of behavior that must be included when defining mastery from an accuracy standpoint. These dimensions include level of performance (e.g., 80% correct) and a

frequency at which this level must be observed (e.g., three consecutive sessions). By holding the frequency of performance consistent at a single demonstration, Fuller and Fienup (2018) systematically manipulated various levels of performance (50%, 80%, and 90%). In contrast to the traditional setting and population sample, participants included three children diagnosed with autism receiving 1:1 instruction in the primary school setting. The primary dependent variable was accuracy in reading and early math skills. Results indicated that higher performance mastery criteria resulted in higher performance retention during follow-up probes. Additionally, skills mastered at the higher performance criteria (90%) showed reduced variability in performance. These results are consistent with earlier research on the use and benefits of high criteria definitions of mastery and extend the mastery literature in terms of examining long-term benefits; however, there was not an evaluation of the effects of different frequencies of performance requirements.

Defining Mastery from a Fluency Perspective

Johnston and Pennypacker (1971) were pioneers in evaluating not only the role of defining mastery, but also doing so from a fluency perspective. The authors reported on findings obtained during university courses delivered using strategies based on a PSI. Results from a psychology course indicated that all students gained high cumulative rates of correct performances and low cumulative rates of incorrect performances. A subsequent replication in a personality theory course demonstrated similar findings. In agreement with Keller (1968), the notion of defining performance based on a normal curve was limiting to a large portion of the students. Furthermore, the reliance that academic institutions place on the use of the normal distribution not only limited access to reinforcement for the majority of students, but also marked

a poor reflection on the field's application of behavioral technologies to structure a successful learning environment for all students (Johnston & Pennypacker, 1971).

Fabrizio and Moors (2003) extended the literature on mastery in selecting and measuring instructional outcomes for children diagnosed with autism spectrum disorders (ASD). The authors stated that too often, educational systems place stress on the format of instruction, as opposed to the outcomes that result from instruction. By focusing solely on the format, it is unclear what a student gains in terms of academic growth. Fabrizio and Moors (2003) suggested that learning occurs in two general phases: accuracy building and frequency building. To be successful in frequency building, a learner must obtain a certain level of accuracy; however, accurate performance alone is far from being considered fluent. As discussed previously, accuracy is only part of the equation in defining mastery (Johnson & Street, 2004). Fabrizio and Moors (2003) concluded that performance aims should be selected which reliably predict critical learning outcomes, such as retention, endurance, stability, and application (RESA; Binder, 1996). In this context, learning does not cease to develop following the accuracy building phase, but rather necessitates frequency building to fluent levels.

In much of the Precision Teaching literature, mastery has been discussed with respect to observed beneficial learning outcomes associated with fluent performance. Haughton (1972) identified the comprehensive goal of education as maximizing learner independence; however, there can be confusion between behavioral acquisition and retention (Lindsley, 1964). For a skill to positively impact a learner's future, the learner must retain successful levels of performance and carry them out under naturalistic conditions. In the Precision Teaching literature, research on the critical learning outcome of retention has drawn interest (Berens, Boyce, Berens, Doney, & Kenzer, 2003; Kubina, Amato, Schwilk, & Therrien, 2008). More specifically, researchers have

observed interesting phenomena that occur when skills are taught to certain performance standards, or *frequency aims* (Haughton, 1972; Kubina & Yurich, 2012). Frequency aims describe performance in terms of quantity that should occur within a timing, resulting in critical learning outcomes such as improved retention, application, stability, or endurance (Johnson & Street, 2004; Kubina & Yurich, 2012).

Haughton (1972) first suggested frequency aims (mastery levels of performance) that predicted a learner would be more successful in retaining performance outcomes and applying the skill to novel tasks. Haughton used the acronym RA/PS to stand for retention and application performance standards to refer to these two observed phenomena. Frequency aims (performance mastery standards) are often described in ranges of desired performance frequencies (e.g., 80-120 correct words per min) that typically result in the learner demonstrating these critical learning outcomes. As research in the area continued, endurance, or the ability to perform for longer durations (Binder, 1993, 1996); stability, or the ability to perform in the face of distraction (Johnson & Layng, 1992); and adduction, or the emergence of novel performances (Johnson & Layng, 1992), were added to the outcomes expanding the acronym to RESAA/PS (Johnson & Street, 2013).

Frequency aims provide both the learner and the instructor with nonarbitrary performance goals (Haughton, 1972). As with other educational programming, frequency aims are often individualized based on student need. For example, a learner who is significantly below the target aim band (e.g., 80-120 correct words per minute) may begin with a more attainable aim (e.g., 30-50 correct words per minute). Through achieving small incremental aims, the learner can meet the final target aim. Learners who build performance to these expert levels, and reap the benefits of these critical learning outcomes (RESAA/PS), are often considered to have

mastered the skill (Johnson & Street, 2004, 2013). However, what empirically constitutes mastery performance in the literature outside the use of frequency aims remains inconsistent and is often heavily reliant on percent correct measures.

Additionally, the frequency literature is not clear with regard to how many observations a frequency aim must be met to result in the emergence of the critical learning outcomes. For example, White (1985) suggested that a single demonstration of a high frequency aim is more beneficial than requiring multiple demonstrations as a lower aim. However, conclusions are limited until additional empirical data can be provided to support such claims.

Summary

While commonly used in the educational and therapeutic settings, the concept of mastery continues to remain unclear. When defining mastery, two dimensional qualities must be examined: level and frequency of performance (Fuller & Fienup, 2018). Essentially, instructors must be clear in terms of defining what quantitative level of performance is required, and on how many occasions this level must be observed to be considered mastered. Literature in the field has explored variations of level of performance, but has yet to provide empirical demonstrations of frequency manipulations. Of the various methods reviewed on defining mastery in the academic and therapeutic settings, several gaps in the literature emerged. First, although requiring a certain level of accuracy in performance is appropriate, accuracy alone is only a portion of defining mastery level performance (Johnston & Street, 2004). Individuals who have mastered a skill must be able to perform both accurately and in a timely manner (i.e., rate). Second, earlier research has varied with respect to the frequency at which levels of performance must be observed to be considered mastered. For example, is one observation at a particular level (100%) sufficient to be considered mastered, or must the level be demonstrated two or three times?

Third, few of the analyses on defining mastery criteria examined the long-term effects on learning outcomes. That is, does defining mastery at a certain level (95% or 120 words correct/min) and for a certain frequency (3 consecutive days) result in improved learning outcomes such as retention?

The purpose of the current study was to address these limitations by clearly defining the level and frequency of performance required to be considered mastered, incorporate fluency in addition to accuracy of skill in defining mastery, evaluate the differential effects (if any) of required frequency of performance on the outcome of performance retention, and explore participant preference for various evaluative measures. Evaluating these variables is critical to the educational and behavior-analytic literature, as it will further assist in both academic and therapeutic settings to further understand and conceptualize the role of defining mastery, and the effect(s) it has on learner performance, and provide guidance to instructional strategies on the benefits of various frequency requirements of performance levels.

Chapter 3: Research Design and Method

Chapter Overview

A design well suited to quick alternations of criteria manipulations and examination across participants was necessary to evaluate the effects of manipulating frequency requirements on the definition of mastery criteria. Therefore, evaluation occurred through the combination of a multielement and nonconcurrent multiple baseline design. To effectively monitor learning and retention, the study used the Standard Celeration Chart (SCC) as the primary visual display for frequencies of "see-says" words correctly and "see-says" words incorrectly using the Chartlytics program. This program was fits well with this study, as it supported the online learning setting and was able to be shared through the on screen sharing features of the online conferencing program GoToMeeting. The following paragraphs outline the selection of materials and sequence of the study, including the mastery criteria frequency demonstration manipulations.

Research Design

The present study evaluated the effects of single versus repeated demonstration criteria of mastery-level performance, using a multielement design (Ulman & Sulzer-Azaroff, 1975) embedded within a nonconcurrent multiple baseline (MBL; Singh et al., 2004) design across participants on the development of reading fluency and retention. The primary investigator selected a nonconcurrent MBL design, as it was appropriate for a dependent variable, such as reading fluency, that is prone to irreversibility and allowed participants to join the study as available. Additionally, the natural staggering of baseline to intervention conditions allowed for multiple evaluations both within and across baselines, making this design appropriate to the present line of inquiry (Watson & Workman, 1981). The decision to embed a multielement design within the MBL allowed direct comparison of the two independent variables (single

versus repeated demonstration criteria), and the brief exposure to the varied criteria associated across the word sets was appropriate to the multielement design. To evaluate the efficacy of both the demonstration criteria, a control condition was implemented throughout all phases in which there was no formal demonstration criteria. The control condition also assisted in monitoring generalization effects on non-altered conditions. The combination of these designs provided a thorough analysis of experimental control both within and across participants without the need of a reversal or intervention withdrawal design.

Research Team

A board certified behavior analyst (BCBA), who will be referred to as "the researcher" throughout this dissertation, acted as the primary investigator of the current study and conducted all sessions. A secondary observer (also a BCBA and doctoral student) was enlisted to observe video- and audio-recorded sessions for purposes of interobserver agreement and procedural integrity data collection. Only the research team had access to the data and videotaped sessions. A board certified behavior analyst doctorate (BCBA-D), who was the academic advisor of the primary investigator, supervised the current research project. All research team members completed the Collaborative Institutional Training Initiative (CITI Program) on the application of social and behavioral research prior to participation in the study.

Recruitment and Eligibility

The current study recruited participants between the age of 6 and 8 years old. This population was selected because the skill of reading fluency is typically a curricular requirement at this age, and therefore, appropriate to their learning level. The researcher posted recruitment flyers (see Appendix A) via personal social media accounts (Facebook, LinkedIn, and Twitter). Flyers included information on the target population for inclusion in the study, as well as basic

prerequisite skills (e.g., vocally reads basic words, can follow simple instructions, and able to access electronic devices with minimal support). Individuals were required to not have a known history of photosensitive epilepsy; however, no restrictions were placed on presence/absence of diagnoses (e.g., diagnosed reading disability, developmental delay, autism, ADHD, etc.) or current academic performance (e.g., below- or above-average performance) in reading outside of the eligibility criteria.

The researcher emailed parents who expressed interest in their child participating in the study, providing a link to a private Adobe® Connect™ or GoToMeeting room and scheduling a time to review the components of the study and the informed consent document (see Appendix B). During the informed consent meeting, the researcher verbally reviewed the document and provided responses to any questions that emerged. Additionally, the researcher notified parents that all information would remain confidential and that they may withdrawal at any time without recourse. Following receipt of verbal consent for participation, the researcher emailed parents the informed consent document to sign and return. An initial session was scheduled with the participants to complete the pre-eligibility screening (see Appendix C) and pretest (see Appendix D) following the receipt of the signed informed consent document.

The pre-eligibility screening assisted in determining if the individual met the basic entrance criteria to be eligible for participation (between the ages of 6 and 8 years old, can read basic words, and able to access electronic devices with minimal assistance). Additionally, all participants were required to have access to a personal computer, webcam, microphone, speakers, and reliable internet connection. Participants who had access to the required technology were asked additional questions regarding screen size and resolution, and noted responses on the basic demographic questionnaire (see Appendix E) when available. Although a

specific screen size and resolution were not required, this information was gathered to determine if there were any differences in performance based on visibility of material (e.g., large screen vs. small screen). Additionally, the demographic questionnaire obtained information regarding the participant's race, primary language, and exposure to other languages in the home or school settings, as well as basic computer specifications (e.g., Windows/Mac, desktop/laptop, estimated screen size and resolution, internal/external webcam/speakers, etc.). This information assisted in determining if differences in performance were observed across individuals who have prior exposure to multiple languages in the home or school settings. The researcher notified parents of the purpose of the demographic questionnaire and explained that answering the questions was completely optional. Partial completion of the demographic form (e.g., provides computer information, but not racial information) was also deemed acceptable.

Participants

The current study recruited a total of four participants. Table 1 contains basic participant demographic information, including participant-selected pseudonyms and technology used to participate in sessions. Belle was an 8-year-old female who resided in the Midwest and attended sessions using a Windows laptop computer. She was concluding her second grade year at the onset of the study. As the youngest participant in the study, Jack was a 6-year-old male from the western region of the United States. Jack was finishing Kindergarten when he joined the study and participated in sessions using the family’s iPad. The third participant, Zoe, was an 8-year-old female from the Midwest. She was concluding her second grade year when she joined the study and attended sessions using the family laptop. The final participant, William, was an 8-year-old male from the Rocky Mountain region of the United States. William was concluding the second grade when he joined the study. William was diagnosed with an autism spectrum disorder, but

participated in the general education class at the public school and did not receive any specialized services.

Table 1

Participant Demographics

	Gender	Age (Years)	Technology
Belle	Female	8	Windows Laptop
Jack	Male	6	iPad
Zoe	Female	8	Windows Laptop
William	Male	8	Windows Laptop

Materials

To assess frequency building in reading Dolch sight words, materials included the presentation of Dolch words on Microsoft PowerPoint slides (see Appendix P). Dolch words are a common curricular expectation within the educational setting for students at this age and therefore were selected for use in this study (Common Core State Standards Initiative, 2018). Font style, size, and color were bolded-black Calibri font, size 30pt., which remained consistent throughout the study. A total of 30 or fewer sight words were presented on the screen at any one time throughout the study to maintain consistency with font size presentation.

Video Conferencing Platform

All sessions were scheduled to occur using the online video conferencing program Adobe® Connect™. However, due to issues with connectivity, disrupted audio transmission, and poor recording quality, the use of this program was terminated after four sessions. The researcher

originally selected Adobe® Connect™ because it allows the video and audio recording of both what is shared on the screen (e.g., Dolch words PowerPoint), as well as the researcher and participant webcams. The video recording was initially considered important, as it would assist with data collection and monitoring of participants attending to the words presented on the screen. However, as sessions proceeded, it was observed that verbal communication from the participant and/or parent was captured in the recording, but failed to transmit live, resulting in the researcher not attending to it. Additionally, during review of the recordings, the audio transmission delay resulted in the recording being unclear and unable to be accurately scored by the secondary observer. As such, Adobe® Connect™ was replaced by the online conferencing program GoToMeeting. This program has similar features in that it allows both the researcher and the participant to share webcams, and allows for sharing of materials (i.e., PowerPoint presentation) on the screen. One feature that the GoToMeeting program did not allow for was the simultaneous recording of the webcams and the presented material. GoToMeeting only has the capacity to record the onscreen material that is shared (e.g., PowerPoint) and the audio of a meeting. However, with the improved connectivity, absence of lag in audio transmission, and improved clarity in recording, GoToMeeting was considered a superior program for the needs of the present study.

Online Charting Program

The online charting program Chartlytics was selected to display and monitor participant performance. Chartlytics allows for data entry of words read correctly (displayed as a dot; •) and words read incorrectly (displayed as an X) per sprint in an online worksheet (see Appendix Q). Once data were entered into the worksheet, they were automatically populated on a Standard Celeration Chart (SCC; see Appendix R) for visual inspection and monitoring. This program was

selected because it allowed for the online entry and sharing of data with the assistant of the participant. By accessing the Chartlytics webpage and sharing their screen, the researcher was able to show the worksheet and related SCCs with the participants during the session.

Measurement

The primary dependent variable of the present study was frequency of “see-says” sight words. This included the frequency of “see-says” sight words read correctly and incorrectly per 10 s sprint. A word read *correctly* is defined as a word that is pronounced accurately in the presented sequence, and at a volume appropriate to the online setting (Kostewicz & Kubina, 2010; Shinn, 1989; Valencia et al., 2010). A word read *incorrectly* is defined as a word that is stated without accurate articulation or clarity (e.g., saying "were" instead of "where"), dropping off one or more letters (e.g., saying "pease" instead of "please") or at a volume that is not distinguishable (e.g., whispering) and not related to technical issues (Kostewicz & Kubina, 2010; Shinn, 1989). Self-corrects, where a participant correctly stated the word after making an error, but before transitioning to the next word on the list, were counted as correct. Omissions and substitutions of any words were counted as incorrect (Kostewicz & Kubina, 2010). Participants were told that they could skip a word by saying “skip.” This rule was provided at the beginning of each session. Skipped words (as indicated by saying “skip”) were counted as incorrect. Frequency of responding is a count of behavior over time. Therefore, frequency of “see-says” sight words was calculated by taking a total count of words read correctly (or incorrectly) and dividing that numerical value by the duration of the sprint timing (i.e., 10 s), thus providing a final frequency of “see-says” words read correctly, and “see-says” words read incorrectly.

Interobserver Agreement and Procedural Integrity

All sessions were audio- and video-recorded for the purposes of obtaining interobserver agreement (IOA) and procedural integrity. A second trained observer scored 42% of randomly selected sessions across phases. Agreements were defined on a word-by-word basis, as both observers recorded a word as either correct or incorrect. Disagreements were defined as one observer recording a word as being read correctly and the other observer recording it as being read incorrectly. Total agreement was calculated by dividing the smaller total by the larger total and multiplying the result by 100% (Cooper, Heron, & Heward, 2007). Average total agreement for each participant was calculated (Belle, 99%; Jack, 99%; Zoe, 100%; and William, 97%).

Both the researcher and a trained second observer used a procedural integrity checklist to ensure that all phases of the study were implemented as designed (see Appendix O). This checklist outlined all possible steps the researcher should complete during a session. This included session instructions, presentation of stimuli, feedback, and so forth. A second observer received training on the use of the checklist prior to scoring any sessions. Training consisted of a didactic component provided through GoToMeeting. During the didactic training portion, the researcher and second observer read through the checklists together with opportunities for questions. Procedural integrity calculations were completed by summing the total number of steps the researcher completed correctly within a session, dividing it by the total number of steps possible, and multiplying by 100%. Average procedural integrity for each participant was calculated (Belle, 99%;' Jack, 94%; Zoe, 100%; and William, 95%). Interobserver agreement for the procedural integrity measures averaged 97% (range: 80-100%). For the session that agreement was 80%, the researcher had scored session performance more strictly than the

secondary observer. To minimize further discrepancies in agreement, a brief retraining was held to address the areas of disagreement.

Methods

Preference Assessment

Following receipt of the signed informed consent document, parents received a brief preference inventory (see Appendix F) through email. Parents completed the inventory with the assistance of the participant and returned it to the researcher prior to the first intervention session. With the assistance of the parent, preferred items/activities from the completed inventory were to be provided, contingent upon meeting designated improvement goals.

Pretest

A brief pretest was completed to establish a pool of known words to use during frequency building. Dolch words (see Appendix G) were presented in isolation on a plain PowerPoint slide until 45 words were read correctly. Instructions included having participants read or attempt to read the presented word to the best of their ability. Participants could skip a word by saying, "I don't know" or "Skip," and they had up to 3 s to read or attempt to read each presented word. Positive feedback (e.g., “nice work”) was provided independent of performance on a variable-ratio three (VR 3) schedule. Following the completion of each Dolch list, and/or if the individual appeared fatigued (e.g., looking away from the screen for durations of 5 s or more, expressing disinterest, or refusing to continue), the researcher offered a brief break. As the focus of the study was on frequency building rather than accuracy building, the first 45 words read correctly during the pretest were used during the frequency building phase.

The 45 words culled from the pretest were divided into three sets of 15 words, one for each experimental condition (see Appendix H). Words presented within a set were randomized and repeated on the slide such that a total of 30 words appeared on a single slide to create a *sprint list.* All sprint lists were of relatively equal level of difficulty as measured by the readability statistics within Microsoft Word. The number of words presented (i.e., 30) on the sprint lists was to prevent an artificial ceiling from hindering performance (Pennypacker et al., 2003). For example, during frequency building, it is important to provide ample stimuli to read during timings such that reorganizing of materials (e.g., flipping through cards) does not negatively impact performance measures (i.e., slow the learner's ability to move through the material with speed). By having 30 words presented on the slide at one time, a participant would need to read at a rate of 180 words per minute in order to read them all before the interval elapsed. Although this is physically possible, it was deemed unlikely to occur with this population, as it would require a performance above the 90th percentile for those completing the second grade (Belle, Zoe, and William; Hasbrouck & Tindal, 2017). As an additional safeguard in preventing an artificial ceiling, participants were instructed that if they read all the presented words and the slide had not disappeared yet, they should start reading back at the beginning again and continue until the slide disappeared.

Preteaching

Participants were provided with a brief preteaching opportunity at the beginning of the study. Preteaching consisted of providing a model and practice opportunity for the participant to experience the "see-say" sprint conditions (see Appendix I). Interval sprints are a frequency-building procedure in which brief practice opportunities are provided to

build fluency in skills (Kostewicz & Kubina, 2010). For the current study, sprints consisted of 10 s timings where the participant was asked to read as many of the words presented on the screen as possible before the slide disappeared. During preteaching, the researcher first described what the participant would see on the subsequent slide (e.g., a slide with columns of colored shapes). The participant was told that the slide would remain on the screen for 10 s before disappearing. The goal was to say as many shape names as possible before the slide disappeared. Participants were told to say the shape names in order, starting in the top left corner of the slide, reading down the first column before moving to the second column, and so on. The researcher modeled a 10 s sprint with the colored shapes before providing the participant a practice opportunity. Participants who initiated shape naming within 3 s of presenting the slide transitioned to baseline conditions. Although up to five opportunities were allowed, no participant needed more than one practice opportunity to demonstrate initiation of naming shapes within 3 s of slide presentation.

Procedures

Sessions were conducted two to three times per week, each with two to three word lists. At the beginning of each session, the researcher completed a brief audio, video, and screen sharing review to ensure all technology was working appropriately. After ensuring the technology was working effectively, the researcher obtained assent from the participant by stating something like, "Hi (name), your (mom/dad) said that you could help me with a project on reading. Is that something you can do today?" (see Appendix J). Assent was confirmed if the child stated, “Yes,” “Yup,” “Yeah,” “Sure,” “Uh huh,” nodded their head, or engaged in a similar response. Once the participant provided assent, the session began according to the demonstration-specific procedures (see below).

Baseline

To establish present levels of performance, each participant conducted baseline sprints. Baseline instructions included asking the participant to read as many of the words on their screen as they could, starting with the top left column and reading down before moving to the top of the second column, and so forth. Should a participant read all the presented words and the slide remained showing, the participant was told that they should begin reading again from the top of the left column until the slide disappears. A single 10 s baseline sprint was completed for each word set (i.e., single demonstration, repeated demonstration, and control) each session. Using an online random list generator (www.random.org/lists), the order of the baseline sprints were randomized. Sprint durations across all conditions were controlled for using the "Advance Slide" function within the PowerPoint program. Baseline sessions were absent of any goal setting or demonstration contingency. Following the sprint, the participant received general praise for completion (e.g., "Good work."); however, no response specific feedback (e.g., "You reached your goal.") was provided. Baseline sessions ended by thanking the participant for their participation. The number of baseline sessions varied across participants (either three or four) and was preselected, consistent with the nonconcurrent MBL literature (Watson & Workman, 1981).

Frequency Building

Each frequency building session began by the researcher obtaining assent from the participant as described above. The session included presenting each of the three word lists (i.e., single demonstration, repeated demonstration, and control) in a randomized order (see Appendices K-M). Intervention sessions for all word sets were similar, apart from the

session improvement goal and criteria demonstration requirement. Frequency building consisted of brief interval sprints to build fluency with Dolch sight words, reinforcement (contingent praise and token economy), goal setting, and error correction procedures.

Frequency building sessions concluded once a participant met the associated word list mastery criteria. A terminal aim of 120 words read correct per min with no more than two errors (Kubina & Yurich, 2012) was modified for a 10 s sprint goal by dividing the suggested rate (120 words correct per min) by the number of 10 s intervals within a min (6). This resulted in a sprint terminal criteria aim of 20 correct words per min with no more than one error. When compared to oral reading fluency norms (Hasbrouck & Tindal, 2006), this terminal criteria would place the participants in the 75th percentile at the conclusion of the second grade. This terminal frequency level was selected as it is the upper boundary of correct performance, as suggested by Kubina and Yurich (2012), and supports the fluency literature in setting high, expert level aims (Haughton, 1972; White, 1985). Additionally, this criterion accommodates for a single error to be made during a 10 s sprint, while maintaining performance levels with no more than two errors when the sprint timing is converted to responses per minute on the Standard Celeration Chart (SCC).

Goal setting occurred during frequency building sessions for single and repeated demonstration criteria word lists. During the first frequency building session, a participant's performance goal was based on the average of their baseline performances for that word list. Improvement goals were increased by 10% from the highest frequency performance each time a goal was met (Fitzgerald & Garcia, 2006; Wood et al., 2002). Goal setting was not used within the control condition; however, the participant was notified of their frequency of words read and received praise. To aid in the discrimination of the criteria

conditions, each set of words was associated with a colored border (single, red; repeated, blue; control, yellow; Ledford & Gast, 2018). The border functioned as a discriminative stimulus indicating changes in experimental conditions. For purposes of data collection, the researcher had identical copies of the word lists to track words read correctly and incorrectly.

Participants received response-specific feedback for frequency of words read correctly (e.g., "You read 16 words."), and improved fluency (e.g., "You met your goal!"). If a participant met the improvement goal, that word list was terminated for the session and the participant moved onto the remaining word lists. If the improvement goal was not met, participants received specific feedback on strategies to try and improve performance; for example, using their finger on the screen to follow the words in order. In the case of words read incorrectly during either the single or repeated demonstration word lists, a hear-say error correction procedure was implemented. The error correction procedure consisted of the researcher identifying the word on the slide using the computer mouse. The researcher stated the word that was read incorrectly aloud to the participant (hear) and asked them to repeat it back (say). Participants received verbal praise for correctly completing the error correction procedure. Sessions ended once the participant completed all sprint lists.

Contingent upon meeting an improvement goal, the participant received specific praise (e.g., "You met your goal, great job!"). Contingent upon meeting two or more goals in a session, participants earned the option to select a preferred item/activity based on the preference inventory. At the end of the session, the researcher notified parents of the participants' performance towards their goals and if an activity had been earned. Parents assisted with ensuring delivery of the item/activity. Participants who did not meet two goals

were provided brief praise for participation in the session and reminded that another opportunity to beat their score would be provided in the next session.

Single demonstration criteria. During single demonstration criteria sessions, the participants were provided their highest frequency score (i.e., frequency of correctly read words) from the prior session. In the case of the first intervention session, participants were provided with their average score from baseline. The researcher shared the improvement goal with the participants during the instructional portion at the beginning of the session. Goals were stated in the form of number of words read correctly (e.g., "Today's goal is 18."). During single demonstration criteria sessions, participants needed to meet the goal for a single sprint in order to have met their goal for the session. Each session, participants had up to three opportunities to meet their goal. If the goal was met on the first sprint, this word list concluded for the day and would increase by 10% the following session. If the goal was not met during the three opportunities, the same goal was used for the next session. After completing a sprint, the researcher provided specific performance feedback (e.g., "Awesome, you met your goal!" or "You did not meet your goal."). Words read correctly and incorrectly were counted and shared with the participants after each sprint. Table 2 provides a breakdown of relevant information regarding the various frequency demonstration criteria manipulations.

Repeated demonstration criteria. Similar to the single demonstration criteria word lists, participants were provided their highest frequency from the prior session and their goal for the day. During repeated demonstration sprints, participants needed to meet and maintain the goal for three consecutive sprints. The repeated demonstration word list was presented for three sprints per session. Therefore, a participant had the possibility to meet the three consecutive

sprint requirement in a single session. Once a participant demonstrated the goal for three consecutive sprints, the goal was increased by 10% consistent with the single demonstration word list. The three consecutive sprints may have occurred within one session, or occurred across two sessions. Again, like the single demonstration criteria, the researcher provided performance-specific feedback, counted the total words read correctly and incorrectly, and implemented the error correction procedure as necessary.

Control. During the control word lists, the participants completed three sprints; however, there was no improvement goal associated for frequency of words read correctly. Furthermore, words read incorrectly did not receive the error correction procedure. Similar to baseline sessions, participants received praise for participation in the sprints, but not for their performance.

Table 2

Frequency Demonstration Criteria Information

	Frequency Demonstration Requirement	Improvement Goal	Border Color
Single	1 sprint	10% increase	Red
Repeated	3 consecutive sprints	10% increase	Blue
Control	N/A	N/A	Yellow

Retention

Retention probes occurred at an interval of 1-, 3-, 5- and 8-weeks postfrequency building. Retention probes were identical to baseline procedures with the exception that only the word

list(s) that met the terminal criteria were assessed. All other word lists remained in frequency building until mastery criteria was met.

Instructional Preference Assessment

Throughout the study, instructional preferences were assessed on approximately half of frequency-building sessions. To assess preferences, participants selected the order in which the word lists (single, repeated, control) were presented within the session. After a minimum of two sessions to ensure exposure to the word lists and various criteria, choice opportunities were offered using the same online random number generator as previously discussed. Instructional preferences were assessed at the start of a session, where the participant was presented with a PowerPoint slide containing three colored squares. Each color was associated with one of the conditions in the frequency building phase (red, single demonstration criteria; blue, repeated demonstration criteria; and yellow, control). The participant was instructed to verbally select the order of the word lists for the session by stating the associated colors. For example, a selection of "red, yellow, then blue" would indicate that the single demonstration word list would be presented first, followed by the control word list, and lastly the repeated demonstration word list. Once participants made their selections, the session began and the word lists were presented in the selected order.

Social Validity

To assess social validity of the procedures, parents received an online survey using the online program SurveyMonkey (see Appendix N). This survey sought to obtain the subjective opinions of the participants and parent on the effectiveness, acceptability, and impact of the intervention.

Chapter 4: Findings

Introduction

Results of the current study revealed that all four participants, who were between the ages of 6 and 8 years old, obtained expert-level frequency aims using brief interval sprints, goal setting, and use of contingent reinforcement. Examination of performance over the retention probes at 1-, 3-, 5- and 8-weeks post-frequency building suggested that performance retention is similar across criteria requirements, if not slightly in favor of the use of single demonstration criteria. Results of the concurrent choice in word list selection and the responses to the social validity survey demonstrated a positive correlation of within-session selection to poststudy feedback. Additionally, participants and parents both shared positive evaluations of the program, suggesting a high social acceptability of the procedures implemented.

Results

Figures 1 through 3 present data on participant "see-says" frequencies of sight words under single demonstration, repeated demonstration, and control criteria, respectively, on stacked celeration charts. As observed in Figure 1, all participants showed acceleration in correct performances (range: x1.28 – x1.88) and slight-to-moderate increases in errors (÷1 – x1.34) under baseline conditions, resulting in three of the four participants meeting the terminal criteria (one 10 s sprint at or above 20 words read correctly with one or fewer errors) without need for frequency building. Only Jack transitioned to the frequency building phase for the single demonstration criteria word list. Once frequency building procedures were implemented, an immediate jump-up was observed, resulting in Jack meeting the terminal criteria after two sprints.

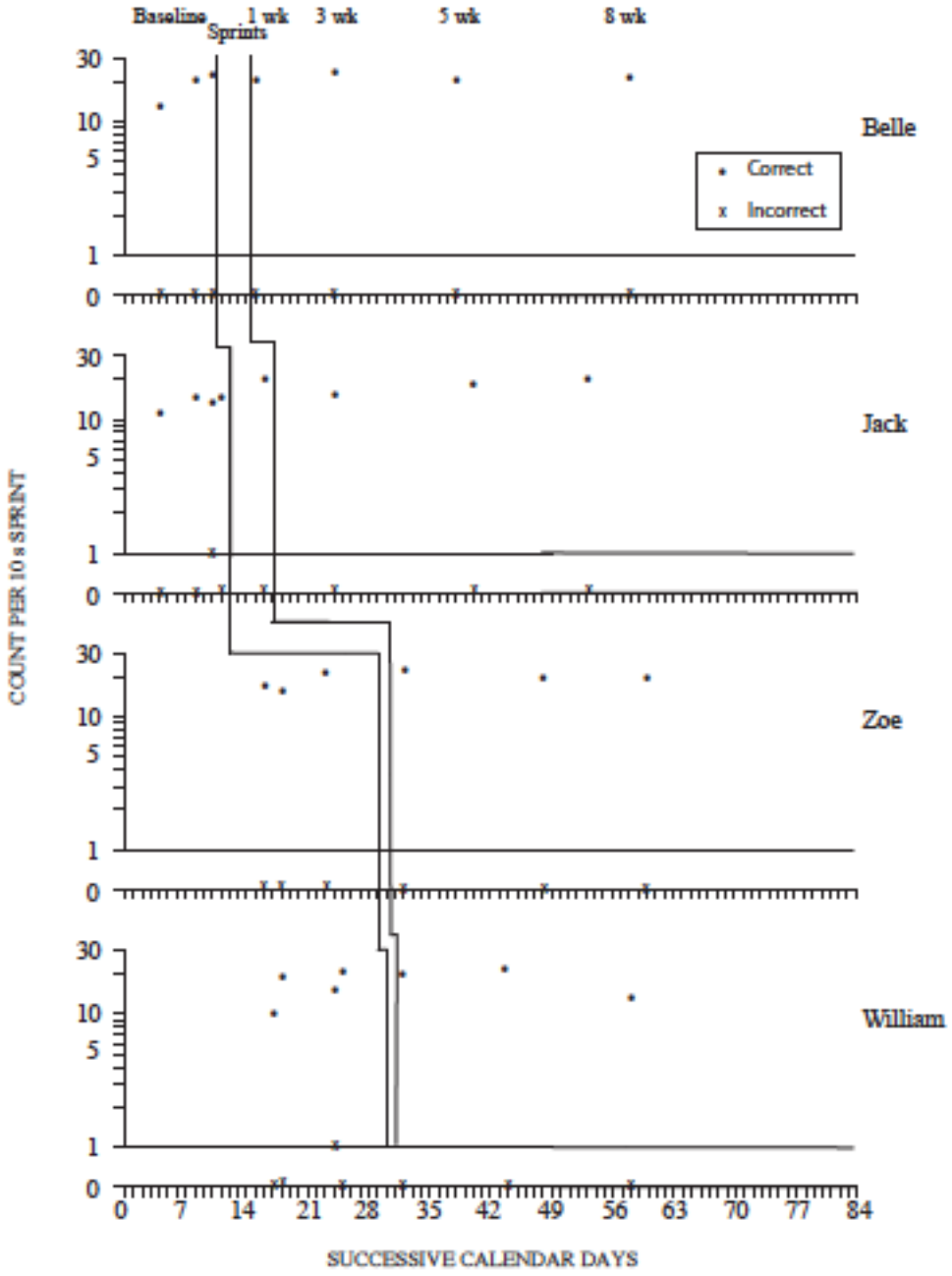


Figure 1. "See-says" sight word frequencies during 10 s sprints under single demonstration criteria conditions.

All participants completed retention probes at 1-, 3-, and 5-weeks postterminal performance with the single demonstration word lists. Belle also completed a retention probe at 8-weeks postterminal performance. Retention probes for performances measured under single demonstration criteria show slight to no decay (range: ÷1.1 – x1.06) in correct performances and retention rate of zero for errors, indicating that participants successfully retained the fluency gains observed over time without practice.

Figure 2 presents participant "see-says" sight word frequencies during 10 s sprints with the repeated demonstration criteria, as displayed on stacked celeration charts. Similar to baseline performances in the single demonstration criteria, all participants showed slight to small acceleration in correct frequencies (range: x1.05 – x1.28) and substantial to no deceleration in error performances (range: ÷2.38 – x1). Three of the participants transitioned to the frequency building condition following the preselected number of baseline sessions, in accordance with the nonconcurrent multiple baseline design. However, due to the observed pattern of the improving baseline performances across the other participants, William's baseline sessions were extended to evaluate whether or not the terminal criteria (three 10 s sprints at or above 20 words read correctly with no more than one error) would be met without the implementation of systematic frequency building and goal setting procedures. In his fifth baseline session, William met the terminal criteria.

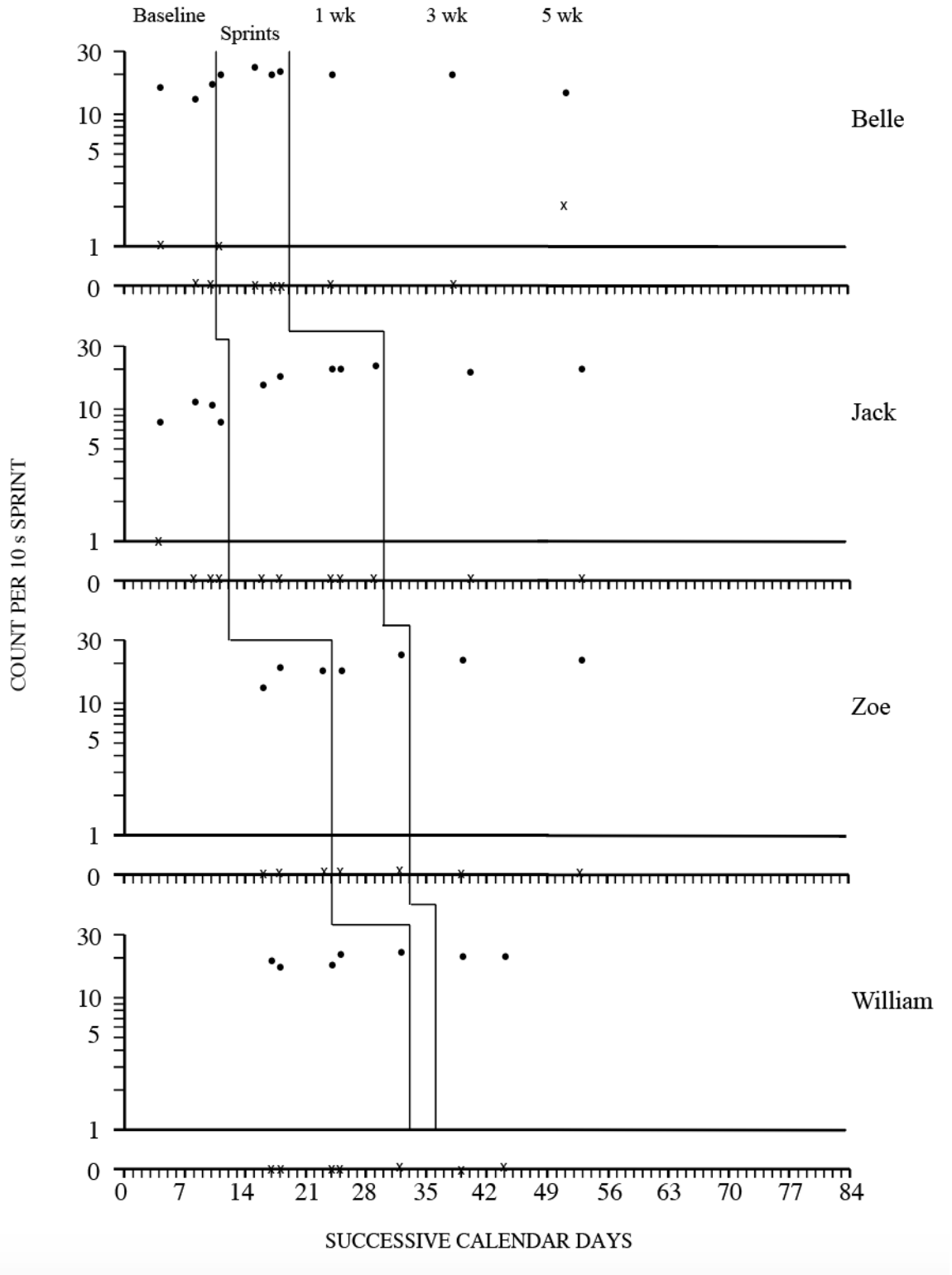


Figure 2. "See-says" sight word frequencies during 10 s sprints under repeated demonstration criteria conditions.

Under the repeated demonstration criteria, all participants completed retention probes at 1- and 3-weeks postintervention, with Belle being the only participant to complete a 5-week probe. Results of the retention probes again indicated slight to no decay (range: ÷ 1.08 –x1) in performances following time without exposure to frequency building procedures. Errors in performance showed slight to moderate increases during retention probes for the repeated demonstration criteria word lists.

As would be expected with the distance-learning format of the study, no covariation effects were observed as participants transitioned across conditions, meaning that the introduction of frequency building procedures or retention probes with one participant did not influence performances across other participants still in previous conditions (i.e., baseline, frequency building). However, two of the participants (Belle and Zoe) showed generalization of improved frequencies in the control condition upon the initiation of frequency building with the repeated demonstration word lists. Both Belle and Zoe's performance showed a jump-up on the control word lists on the fourth data point, following the initiation of frequency building on the repeated demonstration criteria word lists. The remaining two participants, Jack and William, did not show generalization effects in the control condition. Rather, Jack demonstrated a steady acceleration of correct responses (x1.7), while William remained stable (x1). Throughout all conditions of the study, all participants demonstrated smooth and consistent performance, as indicated by small bounce envelopes (range: x1- x2.6).

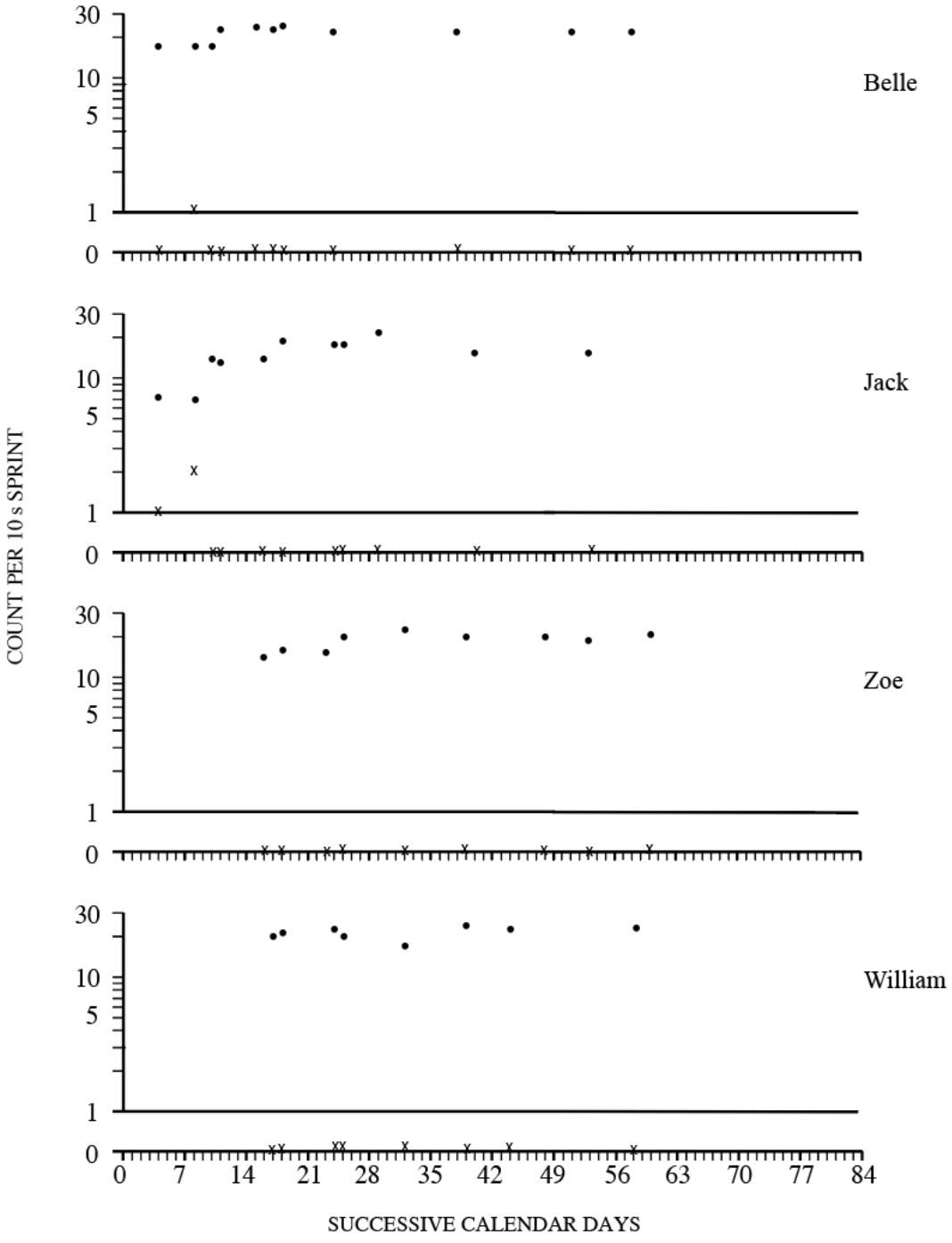


Figure 3. "See-says" sight word frequencies during 10 s sprints under control conditions.

Figure 4 presents the “see-says” sight word frequencies during 10 s sprints across participants Belle, Zoe, and Jack. During baseline, data indicated mixed patterns of responding, meaning there was no observed differentiation of performances across word lists. Upon the initiation of sprints, systematic increases towards the terminal aim were observed. Zoe displayed the most controlled increase to terminal aim across both single and repeated demonstration word lists. Terminal aims were met more quickly for the single demonstration word list as compared to the repeated demonstration word list. Retention probes showed slight drops in performance followed by stabilization. Similar, but more steep decreases were observed during Zoe’s retention probes. Jack displayed an immediate drop, followed by a slight increase in single and repeated demonstration word lists.

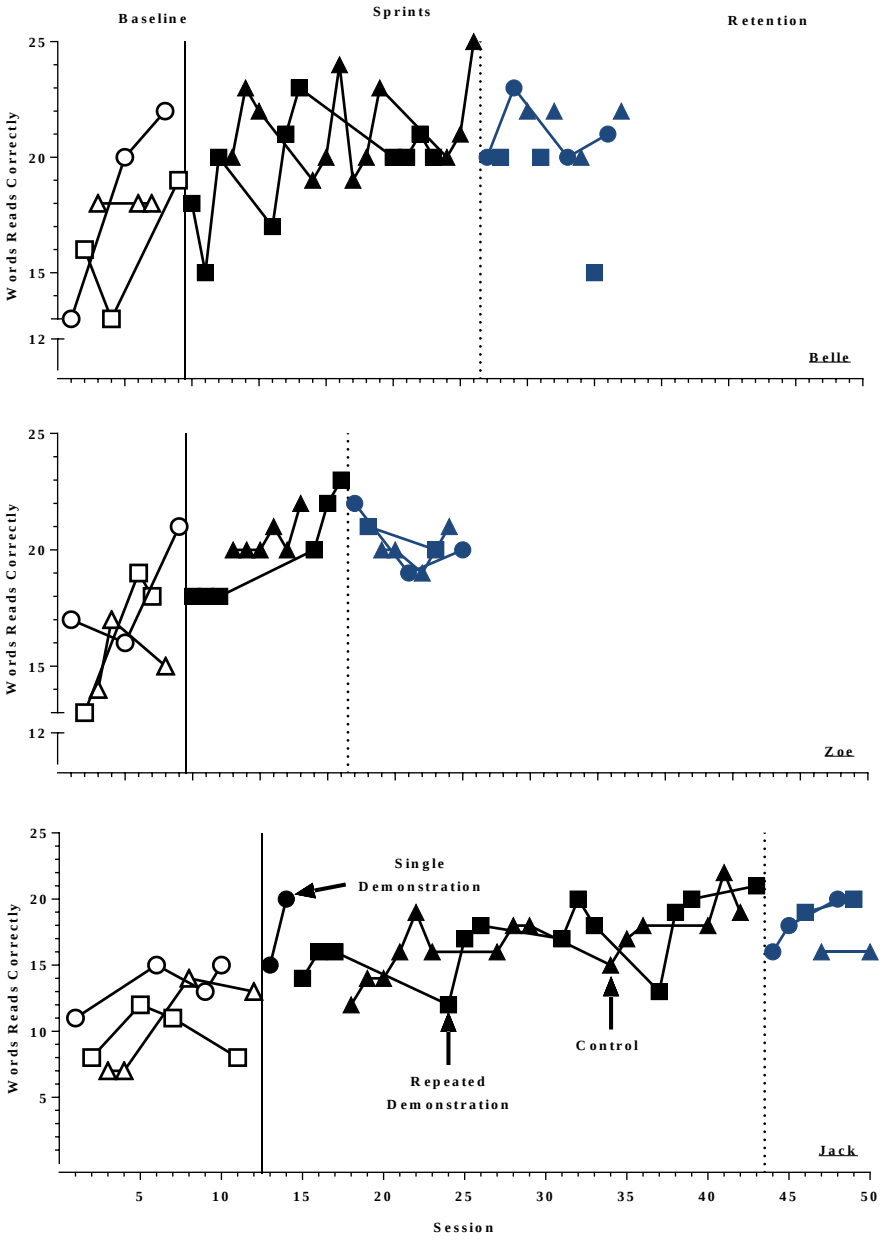


Figure 4. "See-says" sight word frequencies during 10 s sprints across conditions displayed on an equal-interval graph.

Belle

Belle's performance across the three criteria manipulations (single demonstration, repeated demonstration, and control) is presented in Figures 5 and 6 on the celeration chart and equal interval graph, respectively. Additionally, Table 3 provides celeration information across phases, as well as the number of sprints required before meeting mastery criteria. As mentioned earlier, correct performances in both the single and repeated demonstration criteria conditions showed an acceleration (x1.88 and x1.13, respectively) in "see-says" sight words read correctly and steady deceleration (x1 and ÷2.38, respectively) in "see-says" sight words read incorrectly. Examination of frequency multipliers across baseline and frequency building conditions showed an immediate jump up in correct performance level (x1.15) with the institution of frequency building procedures on the repeated demonstration criteria word list. Therefore, in comparing the two demonstration criteria requirements (single versus repeated), Belle met the single demonstration terminal criteria in fewer than half of the sessions (3), as compared to the repeated demonstration terminal criteria (7).

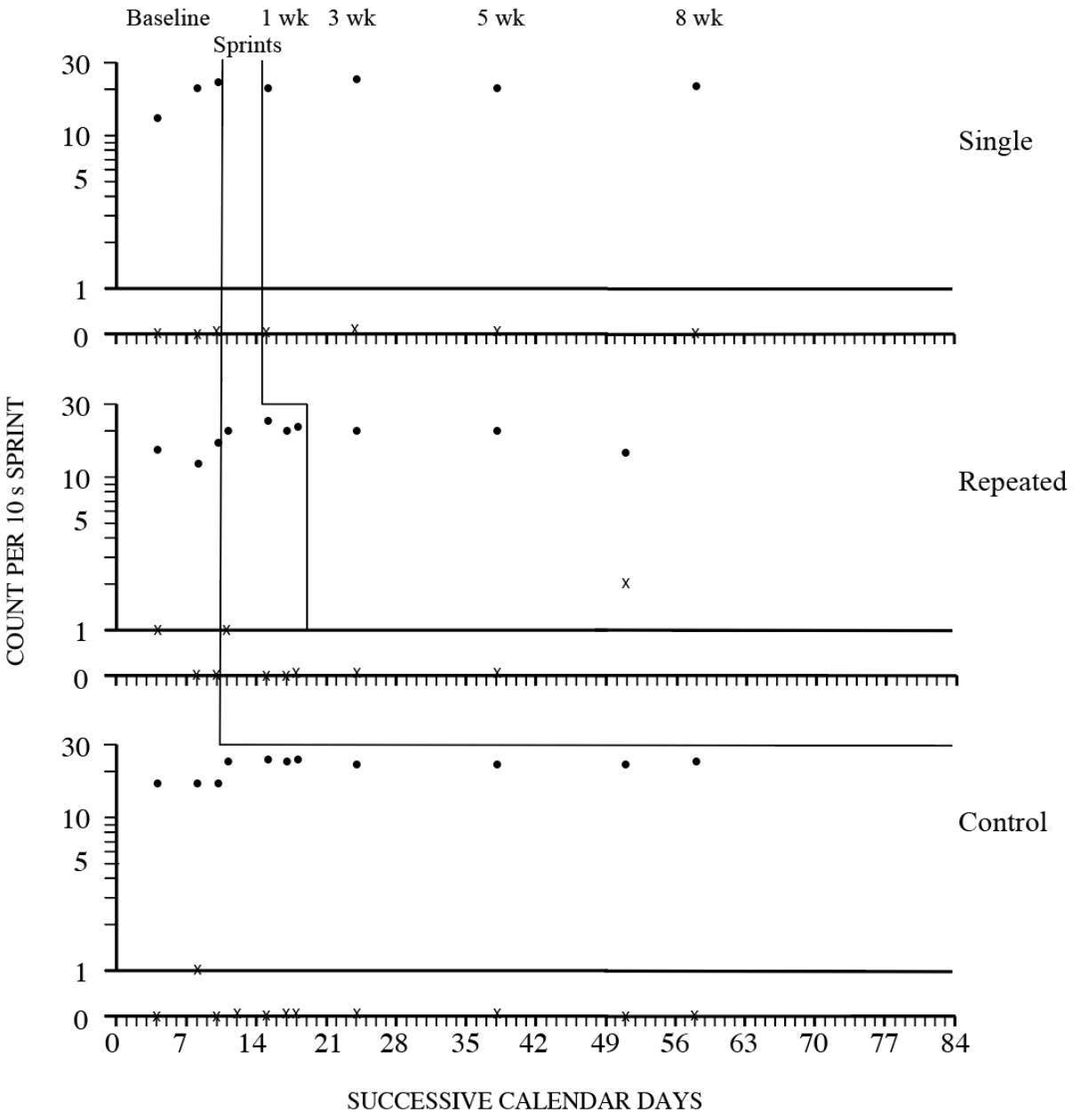


Figure 5. Belle's "see-says" sight word frequencies during 10 s sprints.

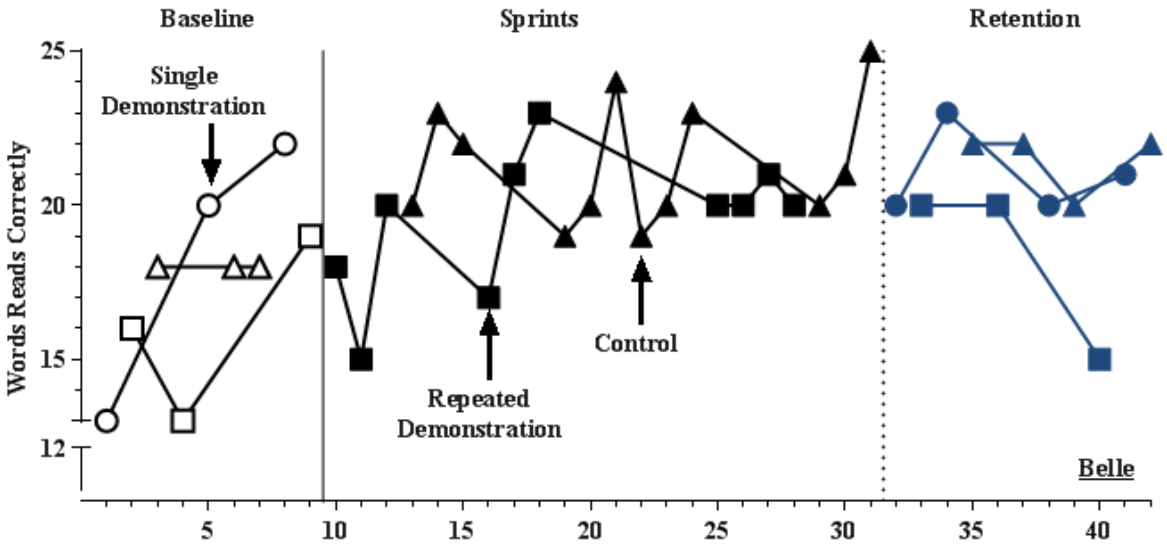


Figure 6. Belle's "see-says" sight word frequencies during 10 s sprints displayed on an equal-interval graph.

Table 3

Celerations and Sprints to Criteria for Belle

Criteria	Baseline	Accel Celeration Sprints	Retention	Sprints to Terminal Criteria
Single	x1.88	N/A	x1	2
Repeated	x1.13	x1.02	x1.01	10
Control	x1	N/A	N/A	4

Additionally, the single demonstration terminal criterion was met under baseline conditions (only two sprints required; see Table 3); therefore, required no systematic frequency building was necessary, as compared to the repeated demonstration criteria, which required four sessions of frequency building to be obtained. Generalization effects were observed in the control condition by a jump-up in correct performance level (x1.22) following the conclusion of

baseline with the other two word lists, suggesting that exposure to the frequency building procedures influenced responding under nontargeted conditions. Although Belle's retention probe performances showed slight immediate jumps down at the 1-week retention probe across both single and repeated demonstration criteria (÷1.1 and ÷1.04, respectively), overall, celerations across all retention probes (1, 3, 5 and 8 weeks) showed slight to no decay in the single demonstration (÷1) and repeated demonstration conditions (÷1.08), therefore indicating that both criteria demonstration requirements resulted in performances that retained following periods of time without frequency building.

Jack

Figures 7 and 8 present Jack's frequency of "see-says" sight words during 10 s sprints across single demonstration, repeated demonstration and control criteria on the celeration chart and equal interval graph, respectively. Table 4 contains Jack's celeration and number of sprints to criterion information. Jack's baseline performance accelerated at a much lower rate with both the single (x1.3) and repeated (x1.12) demonstration criteria as compared to Belle. Correct performance frequencies in the single demonstration criteria showed an immediate jump-up (x1.33) between baseline and frequency building conditions. An even more marked jump-up was observed in the transition from baseline to frequency building in the repeated demonstration criteria condition (x1.75), indicating that the implementation of frequency building procedures was responsible for the increased level of performance. Examining changes in celeration across baseline and frequency building showed a slight increase in celeration from x1.12 to x1.15, suggesting that the implementation of frequency building procedures resulted in not only an increase in level of performance, but also a trend.

In terms of timeliness to meeting criteria, Jack met terminal criteria in only one session, whereas the repeated demonstration criteria required five total sessions. No generalization was

observed within Jack's control word list, indicating a stronger demonstration of experimental control. Jack's retention probe data show a similar pattern to Belle's, in that correct performances in both the single and repeated demonstration criteria showed jumps down at the 1-week retention probe (÷1.25 and ÷1.10, respectively). Again, consistent with Belle's performance, Jack demonstrated consistency in performance across retention probes at 1-, 3-, and 5-weeks postfrequency building. Interestingly, Jack showed a very slight increase in performance in the single demonstration criteria (x1.06) and in the repeated demonstration criteria (x1.05) retention probes, suggesting that his skills were not only maintaining, but also still accelerating at a slight rate.

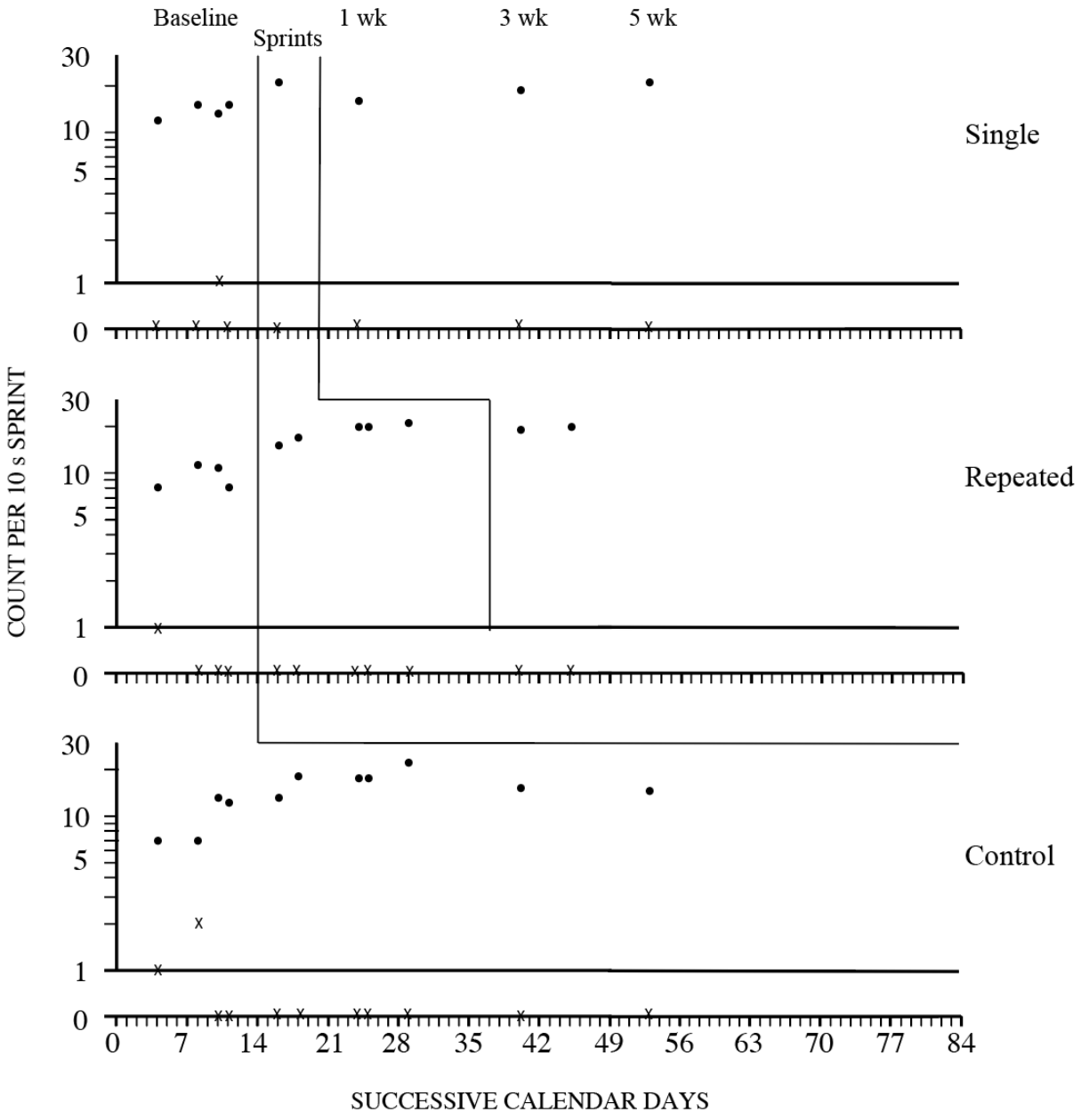


Figure 7. Jack's "see-says" sight word frequencies during 10 s sprints.

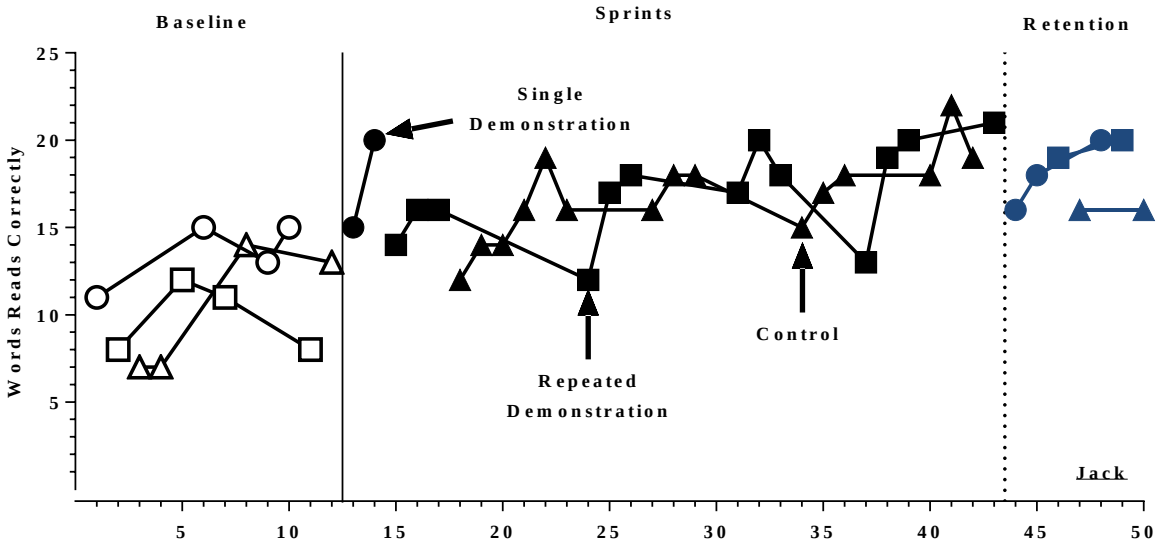


Figure 8. Jack's "see-says" sight word frequencies during 10 s sprints displayed on an equal-interval graph.

Table 4

Celerations and Sprints to Criteria for Jack

Criteria	Baseline	Accel Celeration Sprints	Retention	Sprints to Terminal Criteria
Single	x1.3	N/A	N/A	6
Repeated	x1.12	x1.24	N/A	17
Control	x1.11	N/A	N/A	18

Zoe

Zoe's "see-says" sight word performances across demonstration criteria are presented in Figures 9 and 10 on the celeration chart and equal interval graph, respectively. Similar to other participants, Zoe's baseline performance showed slight acceleration (x1.28; see Table 5) in correct responding under both single and repeated demonstration criteria requirements.

Performance frequencies ultimately met the single demonstration terminal criteria in baseline transitioning that word list directly to retention checks. Although the repeated demonstration criteria word list saw an accelerating trend in baseline, this word list transitioned to frequency building after the preselected three baseline sessions.

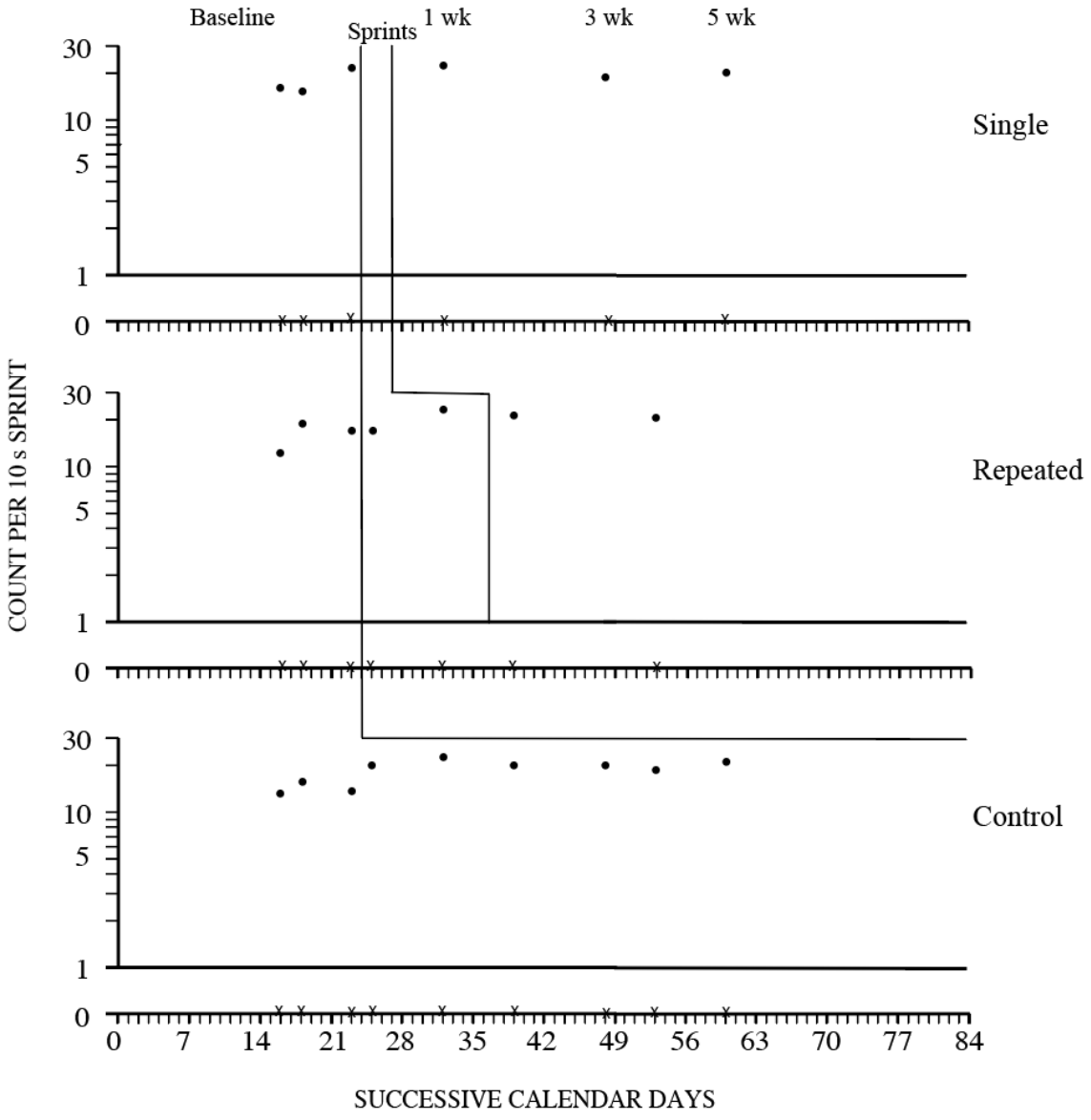


Figure 9. Zoe's "see-says" sight word frequencies during 10 s sprints.

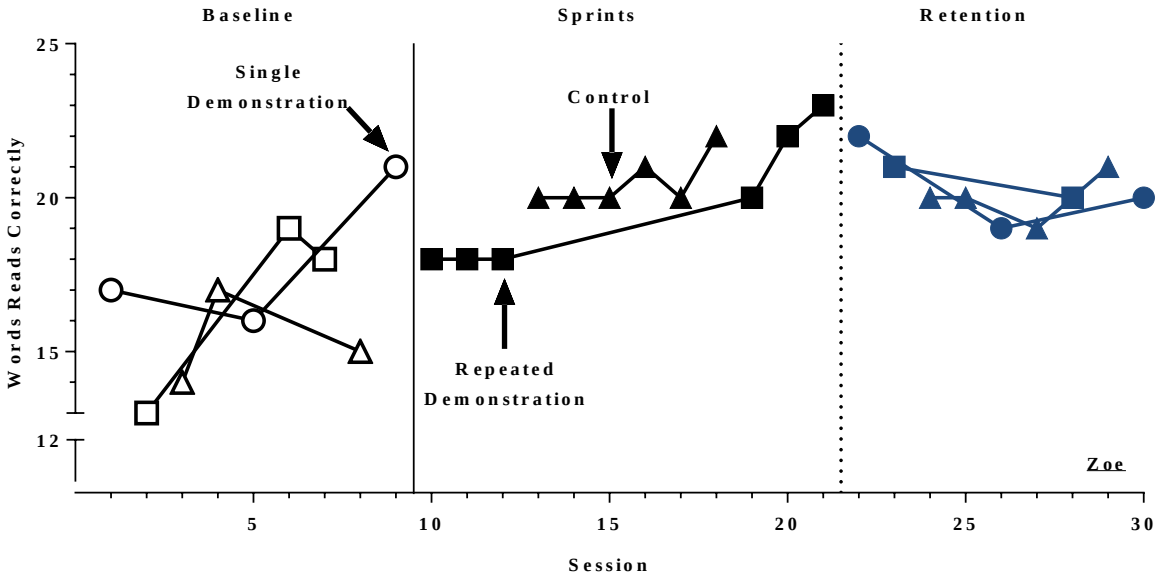


Figure 10. Zoe's "see-says" sight word frequencies during 10 s sprints displayed on an equal-interval graph.

Table 5

Celerations and Sprints to Criteria for Zoe

Criteria	Baseline	Accel Celeration Sprints	Retention	Sprints to Terminal Criteria
Single	x1.28	N/A	÷1.03	3
Repeated	x1.28	N/A	x1.03	9
Control	x1.02	N/A	N/A	4

Upon implementation of frequency building with goal setting and contingent reinforcement, Zoe's performance allocated to match the goal According to her baseline responding, Zoe's first frequency building goal was set at 18 words read correctly. Once notified of the goal, Zoe completed three consecutive sprints at 18. After successfully meeting the first

goal, the 10% improvement goal transitioned the subsequent goal to 20 words read correctly. At the following session, with the new goal of 20 words read correctly, Zoe successfully completed three consecutive sprints at 20 words read correctly. This indicated that the goal setting process exerted strong stimulus control over this participant's performance. This was most evident in comparing the performances of the repeated demonstration criteria list and the control list. With the repeated demonstration criteria, a specific goal was set (18), shared with Zoe, and reinforcement was made contingent upon meeting this goal. However, in the control condition, the only instructions that were provided regarding performance was to "try your best." Once frequency building procedures were implemented with the repeated demonstration criteria word list, there was an immediate jump-up in frequency of words read correctly in the control condition (x1.3), therefore suggesting that although she had the skill to perform higher (as indicated in the control condition), her responding allocated in accordance with the established contingencies of reinforcement. These findings further support the work of Haughton (1972) and Semb, Hopkins, and Hursh (1973), who maintained that students will allocate responding to lower levels, provided available reinforcement is provided even if they have demonstrated higher performance in the past.

William

Finally, Figures 11 and 12 show the "see-says" sight words frequencies for William across single demonstration, repeated demonstration and control criteria on the celeration chart and equal interval graph, respectively. As discussed above, due to the pattern of acceleration under baseline conditions, William's baseline was extended to evaluate whether or not the terminal criteria would be met without direct frequency building procedures (i.e., interval sprints). The word list for the single demonstration criteria was observed to have a baseline

celeration of x1.37, and the terminal criteria was met after four sessions. The word list for the repeated demonstration criteria had a more modest celeration of x1.05; however, William only required six sessions before the terminal criteria were met (see Table 6), therefore indicating that for this participant, mere exposure to the sprint procedure resulted in desired performance outcomes and did not require the addition of goal setting and contingent reinforcement. Examination of performance at a 1-week retention probe indicated no decay in either of the demonstration criteria performances (both were x1.0), thus indicating that William maintained the same levels of performance after reaching the terminal criteria with no additional practice. Throughout sessions, William's performance in the control condition remained high (average of 21 correct words/10 s) and stable (x1.01).

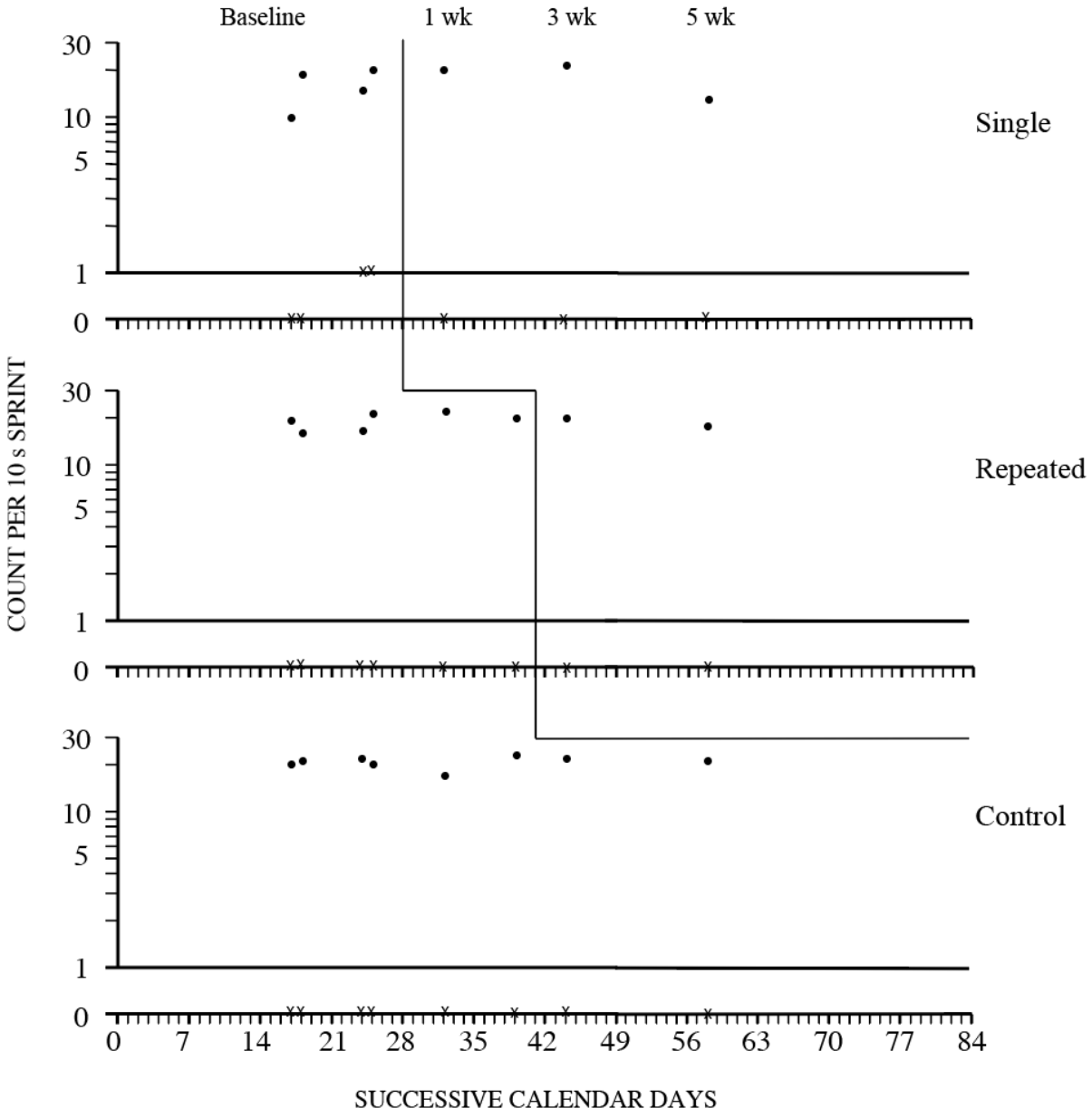


Figure 11. William's "see-says" sight word frequencies during 10 s sprints.

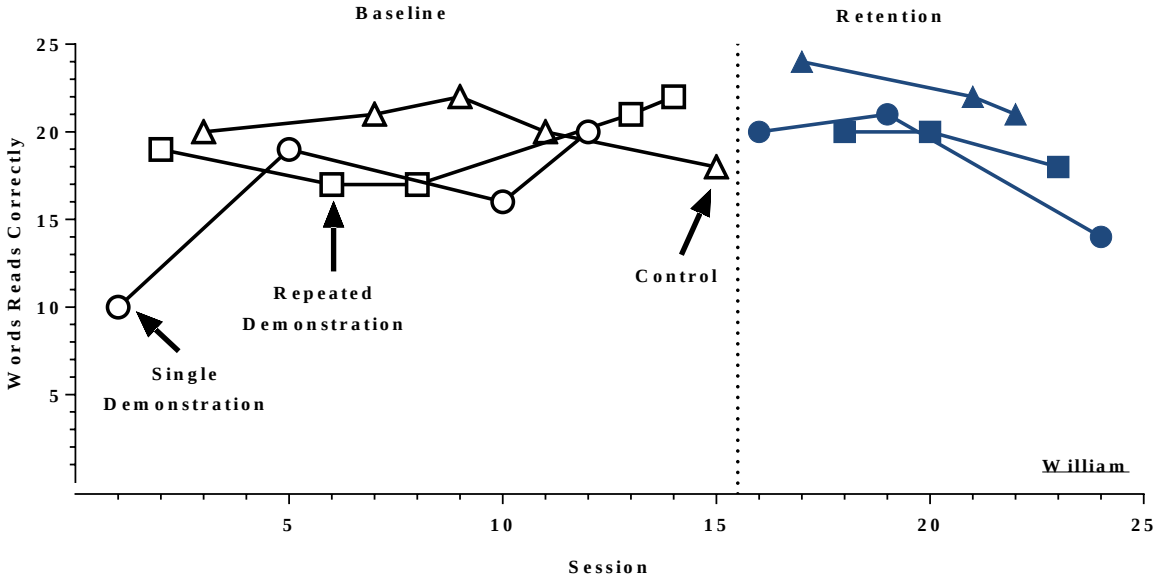


Figure 12. William's "see-says" sight word frequencies during 10 s sprints displayed on an equal-interval graph.

Table 6

Celerations and Sprints to Criteria for William

Criteria	Baseline	Accel Celeration Sprints	Retention	Sprints to Terminal Criteria
Single	x1.37	N/A	÷1.1	4
Repeated	x1.05	N/A	÷1.04	6
Control	x1.01	N/A	N/A	1

On approximately half of frequency building sessions, participants had the opportunity to select which order to complete the word lists for that session. A total of six choice sessions were held across participants (Belle, 2; Jack, 3; Zoe, 1; William, N/A). Table 7 contains the results of the participants' ranking of the word lists in order, from most preferred (first) to least preferred

(third), from the feedback survey. Comparing the within-session selection order to the survey ranking, Belle selected an inverse preference order within session. Although she indicated the control condition word list to be her most preferred in the survey, she placed this last in the within-session sequence. This inverse preference selection contrasted with the remaining participants. Both Jack and Zoe ranked their preferences as single demonstration (first), control (second), followed by repeated demonstration (third). These preferences were consistent with the order selection within session for both of these participants.

Table 7

Participant Survey Rankings of Word Lists from Most Preferred (First) to Least (Third)

	Survey Selection Order		
	First	Second	Third
Belle	Control	Repeated	Single
Jack	Single	Control	Repeated
Zoe	Single	Control	Repeated
William	N/A	N/A	N/A

Participants and parents provided outcome feedback using an online survey designed by the researcher through SurveyMonkey. Results of the survey can be found in Tables 8 and 9 for participants and parents, respectively. Overall, responses from both the participants and parents were favorable to the study outcomes. All participants stated that they agreed that attending sessions resulted in improvements in reading ability and reading speed; they also all expressed that they enjoyed the sessions. All parents shared that the sessions resulted in increases in their child's reading speed, that their child enjoyed participating in the sessions, and that they would

recommend a program like this to other parents. The majority of respondents (66.6%) found that the intervention also improved their child's reading ability and that the addition of the contingent reinforcement activities was easy to incorporate into their routine.

Table 8

Responses to the Participant Social Validity Survey

Question	No, I don't think so	Sometimes/Maybe	Yes, Definitely
After working with Lacy, I am a better reader.	0%	0%	100%
After working with Lacy, I can read faster.	0%	0%	100%
I enjoyed working with Lacy.	0%	0%	100%

Table 9

Responses to the Parent Social Validity Survey

Question	No, I don't think so	Sometimes/Maybe Unsure	Yes, Definitely
Having my child attend sessions with Lacy improved their reading ability.	0%	33.3%	66.6%
Having my child attend sessions with Lacy improved their reading speed.	0%	0%	100%
My child seemed to enjoy sessions with Lacy.	0%	0%	100%
The requirement of providing a fun activity for my child meeting their goals was easy to include into our routine.	0%	33.3%	66.6%
I would recommend this type of program to other parents.	0%	0%	100%

Chapter 5: Discussion and Conclusions

Introduction

The results of the current study provide valuable extensions to the current literature base on the role of defining mastery. All participants successfully achieved high-level frequency aims of "see-says" sight words during brief interval sprints. Furthermore, all participants retained fluency gains at retention probes conducted at 1-, 3-, 5-, and in one case, 8-weeks post-frequency building. These findings suggest that a single demonstration of performance at an expert level frequency aim is sufficient in building a foundation of learning that a student will retain despite periods of time without practice. However, these findings do not come without limitations and additional suggestions for future research. For example, all participants demonstrated at least slight accelerations in performance during baseline. Although beneficial from a therapeutic standpoint, the fact that there were accelerations prior to frequency building procedures limits the ability to draw conclusions regarding the effects of the procedures in isolation. Additionally, with updated oral reading frequencies (Hasbrouck & Tindal, 2017), researchers can explore the use of various levels of rate of performance on the effects of retention and better align them to current performance standards.

Interpretation of Findings

The current study examined the role of manipulating the frequency of demonstrations of performance levels on the development of reading fluency and retention. Results indicated that differences in celerations across frequency requirements were minimal. Furthermore, retention probes conducted at 1-, 3-, 5- and 8-weeks post-frequency building suggested that performances built under single demonstration criteria requirements were similar to performances built under a more rigorous repeated demonstration requirement. Finally, participants overall reported

preferences for single demonstration criteria conditions as compared to repeated demonstration and controlled conditions. These results provide interesting insights to the role of defining mastery and selecting instructional strategies.

First, this study supports the prior literature that suggested defining mastery using high criteria is achievable by all learners when structured in a manner that is specific to their learning needs (Bloom, 1968; Johnston & O'Neill, 1973; Keller, 1968). All four participants met the preselected expert level aims of 20 words read correctly with one or fewer errors during a 10 s sprint. Since celeration differences were minimal across criteria, this would suggest that beneficial outcomes can be met more quickly by setting high fluency aims and only requiring an individual to display this high level of performance on a single occasion. White (1985) stated that teachers are often untrusting, requiring repeated demonstrations to ensure that a student has developed the skill. However, this study suggests and supports White's (1985) assertion that a high performance aim demonstrated on a single occasion is sufficient to retain beneficial outcomes.

Second, although, the number of sessions to achieve the terminal criteria varied, final performances showed little variation. This further supports the mastery learning claims of Bloom (1968) and Carroll (1963), that instruction based on an assurance of prior learning gains (i.e., mastering one topic before moving on) decreases the distribution of scores, allowing all individuals to perform at mastery. Retention probes across participants showed minimal decay even after 5-weeks, and for Belle, 8-weeks post-frequency building. This consistency in performance across participants is ideal in supporting the mastery learning position that all individuals can obtain and retain mastery level skills (Bloom, 1968). Prior literature (Bloom,

1968; Johnston & Pennypacker, 1971) also suggested that the biggest gains were observed in those participants who would typically be considered poor performers.

The current research suggests that performance discrepancies were minimized across participants regardless of incoming baseline performance. However, in some instances baseline celerations were steeper (x1.13) as compared to during frequency building (x1.02), suggesting that initial performances that were farther away from the terminal aim band saw growth that was more proportionally significant than performances as rates neared the terminal criteria. Future research should continue to explore if this phenomena is observed across populations where larger discrepancies are present at the beginning of the learning process and evaluate if there are differences in celeration growth rates. For example, if a participant enters a study closer to the terminal aim band (e.g., baseline of 70; terminal aim band 80-120), are performance growth celerations more shallow (x1.2) compared to a participant whose initial performances are much farther away from the terminal aim band (e.g., baseline of 40)?

Additionally, this work provides support to the ascending mastery criteria described by Carlson and Minke (1975). Each participant proceeded through the goal setting process using a progression criteria individualized based on their performance. For example, improvement goals were individualized based on current rates of performance and increases of 10%. This contrasts with traditional approaches that state all students must obtain the same level of performance (i.e., 90% correct) to access reinforcement. With the individualized procedure, participants contacted reinforcement on a more consistent schedule. In fact, all participants always met the two goal requirement each session. There never was a session where the two out of three potential goals was not met, thus indicating a high success rate for both the procedure and the participant. Carlson and Minke (1975) found that participants in the ascending criteria condition were more

successful passing on the first attempt. This is relevant to the present study, as the individualized procedures used within also supported first attempt success. For example, Zoe required only two sessions in fluency building for the repeated demonstration word list. Each session, she met her goal on the first attempt. Additionally, aside from the very first sprint of the repeated demonstration list in frequency building, Belle also met each subsequent goal requirement on the first attempt. It is possible that the failure to meet the goal on the first sprint was more closely related to the ambiguity in the goal than in Belle's performance ability (discussed more below). This conclusion is further supported by the fact that her remaining sprints were performed at or above the goal performance of 20 correct words per sprint.

Finally, this study extends the mastery criteria literature base by taking into account participant choice and preference. During intervention preference assessments, participants overwhelmingly selected the single demonstration criteria word list first. This selection matched the results of the survey. These results suggest that incorporating participant preference in selecting a monitoring and evaluation measure may be both successful to learning outcomes and more engaging for the learner. A systematic review of literature conducted by Royer, Lane, Cantwell, and Messenger (2017) found that incorporating choice increased desired academic behavior, in addition to decreasing behaviors typically viewed as problematic. Interestingly, results from the current study also suggested that participants selected a word list order that would assist in building behavioral momentum; that is, they selected their favorite (according to the survey) list to complete first, followed by their next favorite, and lastly, their least favorite, essentially self-establishing a high-probability sequence of list completion. Belfiore, Lee, Vargas, and Skinner (1997) found that sequencing a single-digit math problems prior to 3-digit problems resulted in reduced latency to initiate the 3-digit problems. This finding is relevant to

the current literature, as it suggests that the student selection of the single demonstration word list prior to the repeated demonstration word list may have established a high-probability sequence, building momentum for success during the list with the more rigorous criteria requirement. Considering these findings, selection sequence, choice, and self-established high-probability sequence components of selecting and evaluating mastery criteria should be further explored.

Readers must take a few considerations into account when interpreting the results of this study, however. First, and possibly most concerning, were the inconsistent intervention effects and behavioral covariation observed across two of the participants (Belle and Zoe). Ledford and Gast (2018) identified behavioral covariation as a threat specific to the multiple-baseline design. With both participants, the onset of frequency building procedures on one set of words (i.e., repeated demonstration) resulted in an observable effect in a nonaltered condition (i.e., control). Behavioral covariation presents a threat to the internal validity of the study and limits the conclusions that can be drawn regarding the use of interval sprints in fluency building. A possible reason why these effects were observed in the present study lies in the functional similarity of the criteria requirements and the potential of the control instructions to encourage improved performance. For example, while both criteria under evaluation required demonstration of improved performance, it is possible that the differentiation of the contingency requirements was not clear to the participants. Therefore, participants demonstrated similar responding across all conditions. A colored border had been added to the word lists to aid in discrimination; however, this discriminative stimulus was not a salient enough stimulus to alter performances during frequency building procedures. An additional possibility as to why performance improvements were observed within the control condition is that the instructions participants

received suggested a functional contingency for displaying improved performance, meaning it is possible that as participants observed their increasing performances on the single and repeated demonstration word lists, the instruction to "try your best" was functionally altered to match their improving performances of the other lists.

A second observed threat to the internal validity of the study is the possibility of the Hawthorne effect (Ledford & Gast, 2018). The Hawthorne effect may occur when participants are sensitive to the desired outcomes of a study. Although not specifically discussed until the onset of frequency building, the format of presentation may have provided contextual cues to participants regarding the desired performance. For example, during baseline, participants were informed that the slide containing the words would only stay on their screen for a brief amount of time. Considering participant learning histories, it is possible that the temporal information regarding the presentation of stimuli may have been enough to result in fast-paced performance.

A third limitation was observed with regard to the goal setting process. What was thought to be a clear indication of desired performance (i.e., stating desired rate) became evident as being ambiguous. For example, with the exception of one participant (Zoe), the quantitative statement about desired rate (e.g., "Our goal this time is 18 words.") exerted little stimulus control over performance. Rather, a more concrete description of performance was necessary (e.g., "Our goal is 20. We will know we made it if we read all the words in the first two columns."). Therefore, future research should explore the role of within- and extra-stimulus prompts to guide improvement goal performance in a reading fluency context. Additionally, the use of specific versus vague instructions on the development of reading fluency and retention should also be explored. For example, LeFrancois, Chase, and Joyce (1988) found that the use of specific instructions resulted in responding that was insensitive to the changes in schedules of

reinforcement across conditions. Therefore, since a specific goal (e.g., “Your goal is 18.”) may have resulted in the participant becoming insensitive to the change in contingencies when the control list was presented.

Recommendations

An ideal outcome for instructional strategies is that they will be successful across a variety of contextual variables (Hains & Baer, 1989); however, to provide the most accurate instructional recommendations, it is important to identify the variables that may have an impact on performance. In the present study, the contextual variable of frequency of mastery level performance was examined. Results indicated that similar effects were observed across criteria requirements with regard to celeration and retention; however, it is worth noting that participants met the terminal criteria in less time under the single demonstration criteria, thus decreasing the amount of direct instructional time to that word list. This indicates that if similar effects of frequency building and retention effects are observed following a single demonstration of performance at a high frequency, the time formerly allocated to the requirement of the repeated demonstrations could be allocated to new or more complex material, thus expanding the learner’s academic repertoire more quickly.

Future researchers should seek to address the limitations discussed above. For example, to minimize the potential of participants meeting or exceeding the terminal criteria under baseline conditions, eligibility requirements may be set so that only learners presently performing within a particular range would be eligible for participation. Additionally, future researchers could seek to explore the role of frequency of criteria demonstrations on other academic and non-academic skills, and across various populations. The present study used an academic skill, as that is the setting most commonly tasked with the role of determining if a

student is competent in a skill or not. However, mastery with non-educational tasks, such as activities of daily living and vocational tasks, would be just as appropriate for research. This study sought performers within a specific age range and all participants attended local public schools. Only one had a diagnosed disability on record. With that said, the role of mastery is not confined to students in the general education setting. Students receiving special educational services are also in need of a thorough analysis of mastery with respect to the learning environment.

Additional areas of inquiry worth exploring include critical learning outcomes such as skill application and endurance. According to Fabrizio and Moors (2003), critical learning outcomes such as application and endurance should be examined prior to probing for retention. This is because to effectively monitor for retention of a skill, the participant must experience periods without intervention. Therefore, there is the possibility that valuable instructional time could be lost waiting to examine skill retention; whereas if skill application or endurance is measured first, and found to be insufficient, frequency building procedures can remain intact and the only instructional time lost is to complete the brief application or endurance probe (e.g., 30s – 1 min). Application measures could be especially relevant with regard to measuring performance gains in standardized academic tests. Schools and school districts are strongly influenced by student outcomes on these types of assessments. Therefore, it would be relevant to extend the literature on the use of mastery criteria within individual programs and its application to performances on standardized assessments.

Conclusions

The findings from the present study suggest that the concept of mastery has many facets yet to be explored and refined. In terms of designing instruction, this study supports earlier work

that academic discrepancies can be minimized with properly structured learning environments, thus allowing all individuals the possibility to obtain mastery level and opening up the academic world to classes filled with high performers. Although limiting conclusions from an experimental-control standpoint, in an educational environment, the observed behavioral covariation would be considered beneficial, meaning that the implementation of an effective intervention in one context may lead to beneficial outcomes in other contexts. The ability to pair a definition of mastery that leads to accelerated learning and retention, while minimizing direct instructional time would increase the amount of material that could be presented during a typical academic year, allowing increased access to educational gains for all.

References

Belfiore, P. J., Lee, D. L., Vargas, A. U., & Skinner, C. H. (1997). Effects of high-preference single-digit mathematics problem completion on multiple-digit mathematics problem performance. *Journal of Applied Behavior Analysis, 30*, 327-330.

Berens, K. N., Boyce, T. E., Berens, N. M., Doney, J. K., & Kenzer, A. L. (2003). A technology for evaluating relations between response frequency and academic performance outcomes. *Journal of Precision Teaching, 19*, 20-34.

Binder, C. (1993). Behavioral fluency: A new paradigm. *Educational Technology, 3*, 8-14.

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19*, 163-197.

Binder, C. (2011). Building fluent performance: Measuring response rate and multiplying response opportunities. *The Behavior Analyst Today, 11*, 214-225.

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment, 1*, 1-11.

Carlson, J. G., & Minke, K. A. (1975). Fixed and ascending criteria for unit mastery learning. *Journal of Educational Psychology, 67*(1), 96-101.

Carroll, J. A. (1963). A model of school learning. *Teachers College Record, 64*, 723-733.

Common Core State Standards Initiative. (2018). *English language arts standards.* Retrieved from: http://www.corestandards.org/ELA-Literacy/RF/K/

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed). Upper Saddle River, NJ: Pearson.

Fabrizio, M. A., & Moors, A. L. (2003). Evaluating mastery: Measuring instructional outcomes for children with autism. *European Journal of Behavior Analysis, 4*, 23-36.

Ferster, C. B. (1968). Individualized instruction in a large introductory psychology course. *Psychological Records, 18*, 521-532.

Fienup, D. M., & Brodsky, J. (2017). Effects of mastery criterion on the emergence of derived equivalence relations. *Journal of Applied Behavior Analysis, 50*, 843-848.

Fitzgerald, D. L., & Garcia, H. I. (2006). Precision teaching in developmental mathematics: Accelerating basic skills. *Journal of Precision Teaching and Celeration, 22*, 11-28.

Fuller, J. L., & Fienup, D. M. (2018). A preliminary analysis of mastery criterion levels: Effects on response maintenance. *Behavior Analysis in Practice, 11*, 1-8.

Graf, S., & Lindsley, O. (2002). *Standard celeration charting*. Poland, OH: Graf Implements.

Hains, A. H., & Baer, D. M. (1989). Interaction effects in multielement designs: Inevitable, desirable, and ignorable. *Journal of Applied Behavior Analysis, 22*, 57-69.

Hasbrouck, J., & Tindal, G. (2017). *An update compiled ORF norms* (Technical Report No. 1702). Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Haughton, E. (1972). Aims: Growing and sharing. In J. B. Jordan, J.B. & L. S. Robbins (Eds.), *Let's try doing something else kind of thing: Behavioral principles and the exceptional child. A report from the Invisible College Conference on Application of Behavioral Principles in Exceptional Child Education, March, 1971* (pp. 20-39). Arlington, VA: The Council for Exceptional Children.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve*. New York, NY: Free Press.

Johnson, K. R., & Layng, T. V. J. (1992). Breaking the structuralist barrier literacy and numeracy with fluency. *American Psychologist, 47*, 1475-1490.

Johnson, K. R., & Street, E. M. (2004). *The Morningside model of generative instruction*. Concord, MA: Cambridge Center for Behavioral Studies.

Johnson, K. R., & Street, E. M. (2013). *Response to intervention and precision teaching*. New York, NY: Guilford Publications, Inc.

Johnston, J. M., & O'Neill, G. (1973). The analysis of performance criteria defining course grades as a determinant of college student academic performance. *Journal of Applied Behavior Analysis, 6*, 261-268.

Johnston, J. M., & Pennypacker, H. S. (1971). Behavioral approaches to college teaching. *American Psychologist, 26*, 219-244.

Johnston, J. M., & Pennypacker, H. S. (2009). *Strategies and tactics of behavioral research* (3rd ed). New York, NY: Routledge.

Keller, F. S. (1968). "Good-bye teacher..." *Journal of Applied Behavior Analysis, 1*, 78-89.

Kostewicz, D. E., & Kubina, R. M. (2010). A comparison of two reading fluency methods: Repeated readings to a fluency criterion and interval sprinting. *Reading Improvement, 47*, 43-63.

Kubina., R. M., Jr., Amato, J., Schwilk, C. L., & Therrien, W. J. (2008). Comparing performance standards on the retention of words read correctly per minute. *Journal of Behavioral Education, 17,* 328-338.

Kubina, R. M., Jr., & Yurich, K. L. (2012). *The precision teaching book*. Lemont, PA: Greatness Achieved Publishing Company.

Ledford, J. R., & Gast, D. L. (Eds.). (2018). *Single case research methodology: Applications in special education and behavioral sciences*. New York, NY: Routledge.

LeFrancois, J. R., Chase, P. N., & Joyce, J. H. (1988). The effects of a variety of instructions on human fixed-interval performance. *Journal of the Experimental Analysis of Behavior, 49*, 383-393.

Lindsley, O. R. (1963). Experimental analysis of social reinforcement: Terms and methods. *Symposium: Social Reinforcement and Behavior Change,* 624-633.

Lindsley, O. R. (1964). Direct measurement and prosthesis of retarded behavior. *Journal of Education, 147,* 62-81.

Lindsley, O. R. (1991). Precision teaching's unique legacy from B. F. Skinner. *Journal of Behavioral Education, 1*, 253-266.

Lindsley, O. R. (1996). Is fluency free-operant response-response chaining? *The Behavior Analyst, 19*, 211-224.

McMichael, J. S., & Corey, J. R. (1969). Contingency management in an introductory psychology course produces better learning. *Journal of Applied Behavior Analysis, 2*, 79-83.

Motamedi, V., & Sumrall, W. J. (2000). Mastery learning and contemporary issues in education. *Action in Teacher Education, 22*, 32-42.

National Center for Education Statistics. (2011, April 7). How is grade average calculated? Washington, DC: U.S. Department of Education. Retrieved from: https://nces.ed.gov/nationsreportcard/hsts/howgpa.aspx

Pennypacker, H. S., Gutierrez, A. Jr., & Lindsley, O. R. (2003). *Handbook of the standard celeration chart*. Concord, MA: Cambridge Center for Behavioral Studies.

Royer, D. J., Lane, K. L., Cantwell, E. D., & Messenger, M. L. (2017). A systematic review of the evidence base for instructional choice in K-12 settings. *Behavioral Disorders, 42*, 87-107.

Semb, G. (1974) The effects of mastery criteria and assignment length on college-student test performance. *Journal of Applied Behavior Analysis, 7*, 61-69.

Semb, G., Hopkins, B. L., & Hursh, D. E., (1973). The effects of study questions and grades on student test performance in a college course. *Journal of Applied Behavior Analysis, 6*, 631-643.

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children*. New York, NY: Guilford.

Singh, N. N., Lancioni, G. E., Winton, A. S. W., Wahler, R. G., Singh, J., & Sage, M. (2004). Mindful caregiving increases happiness among individuals with profound multiple disabilities. *Research in Developmental Disabilities, 25*, 207-218.

Skinner, B. F. (1968). Teaching science in high school. *Science, 159*, 704-170.

Stevens, D. (2003). Single vs. multiple movement frequencies: How many times should we measure? *Journal of Precision Teaching, 19*, 45-47.

Ulman, J. D., & Sulzer-Azaroff, B. (1975). Multi-element baseline design in educational research. In E. Ramp & G. Semb (Eds.), *Behavior analysis: Areas of research and application* (pp. 371-391). Upper Saddle River, NJ: Prentice Hall.

Valencia, S. W., Smith, A. T., Reece, A. M., Li, M., Wixson, K., & Newman, H. (2010). Oral reading fluency assessment: Issues of construct, criterion, and consequential validity. *Reading Research Quarterly*, *45*(3), 270-291.

Watson, P. J., & Workman, E. A. (1981). The non-concurrent multiple baseline across-individuals design: An extension of the traditional multiple baseline design. *Journal of Behavioral Therapy and Experimental Psychology*, *12*, 257-259.

White, O. R. (1985). Decisions, decisions. *British Columbia Journal of Special Education, 9*, 305-320.

White, O. R., & Haring, N. G. (1980). *Exceptional teaching* (2nd Edition). Columbus, OH: Charles E. Merrill Publishing Co.

Wood, S. J., Murdock, J. Y., & Cronin, M. E. (2002). Self-monitoring and at-risk middle school students. *Behavior Modification, 26*, 605-626.

Zimmerman, B. J., & Dibenedetto, M. K. (2008). Mastery learning and assessment: Implications for students and teachers in an era of high-stakes testing. *Psychology in the Schools, 45*, 206-216.

Appendix A: Request for Participation

Request for Participation

TheChicagoSchool of Professional Psychology

Purpose: The purpose of this research project is to evaluate the development of basic reading skills and the procedures that lead to lasting change

Who is eligible?

- Children 6 - 8 years of age
- Children must have basic reading skills (don't worry, we will help if you are not sure)

What will your child be asked to do?

- Read brief lists of common words
- Help chart their performance
- Set goals and work to achieve them
- Enjoy interactive games/activities
- Sessions are anticipated to last between 15-20 minutes

Potential benefits from participation?

- Increases in reading speed & accuracy
- Reading improvements that last

Anticipated involvement

- 10-12 weeks in total
  - Instruction is anticipated to last 2-4 weeks with sessions occurring 3-7 times per week
  - 4 follow-up sessions will occur at 1, 3, 5 and 8 weeks after the conclusion of instruction

Where will it occur?

- At your home through your personal computer

Participation in this research project is entirely voluntary.
All sessions will be audio and video recorded for data collection purposes.
If you have any questions or are interested in participating, please contact:
Lacy Knutson at (651)-247-1841 or Email: lmk4193@tcps.thechicagoschool.edu

3.1.2018v.2, Page 1 of 1

Appendix B: Informed Consent

Informed Consent

Investigators: Lacy Knutson

Study Title: Single vs. Repeated Criterion Demonstrations on Reading Fluency and Retention

I am a student at The Chicago School of Professional Psychology and this study is being completed as my dissertation requirement for the PhD in Applied Behavior Analysis.

I am asking consent for your child to participate in a research study. Please take your time to read the information below and feel free to ask any questions before signing this document.

Purpose: This study is interested in using short learning opportunities to build children's reading abilities and to evaluate two monitoring strategies that may help children maintain learning gains. This is especially important for beginning readers to build a solid reading foundation and maintain those gains over typical breaks (e.g., summer/winter break).

Procedures: All activities will occur using the video conferencing program Adobe® Connect™. If you agree to have your child participate by signing this document, I will ask you a few questions about your child which will include information about their current language exposure and preferred activities, as well as they type of computer/technology that will be used in the home to access sessions.

In the first session, your child will become familiar with the online learning format. During this session, your child will complete up to five 10 second "sprints" identifying shapes presented to them on the computer screen. They will be asked to state the name of as many shapes as they

can prior to the slide changing. The slide will only remain on their screen for 10 seconds. Once your child knows what to expect with the sprints, your child and I will complete a baseline session where we will identify 45 known words from the Dolch word lists. The Dolch word lists are frequently used English words and were developed by Edward William Dolch. This will be done by showing a single word on their screen and asking them to read or try to read the word. The first 45 words that can be read correctly will be used for the intervention sessions.

Intervention will involve the use of short 10 second sprints and a 10% improvement goal to build fluency (speed and accuracy). Based on your child's performance during baseline, three 15-item sprint word sets will be developed. Two of the word sets will represent a unique monitoring strategy (single demonstration or repeated demonstrations). A third condition (control) will be completed to monitor learning that is developed with a neutral monitoring strategy. The single demonstration condition will seek your child to demonstrate performance at or above his or her 10% improvement goal on one occasion. The repeated demonstration condition will seek that your child demonstrates performance at or above his or her 10% improvement goal on three consecutive occasions.

Children will seek to improve their performance using a 10% improvement goal which will be shared with them at the beginning of each sprint. After each sprint, I will provide the children with feedback as to whether they made their improvement goal or how close they were. If children meet their 10% improvement goal for a word set, I will deliver a token towards an end of session preferred activity. Each word set will be conducted 1-3 times per session. At the end of each session, children will assist me in documenting their performance using an online data collection system. If the children collect a minimum of 2 out of 3 tokens in a session, they will be provided access to a preferred activity or game, based on their completed preference assessment and with collaboration from you as their parent/guardian.

On approximately half of the sessions, children will be asked to select the order they would like to complete the word sets (single, repeated or control). In the remaining sessions, the order will be selected by myself. This choice opportunity is included to evaluate your child's preference across the monitoring strategies.

Sessions are anticipated to occur three to seven times per week for approximately 4 weeks. All are expected to last 15-20 minutes in duration. Once your child has met the end goal, there will be a break in sessions to complete four post-intervention assessments to check for skill maintenance.

At the end of the study, you will be provided an opportunity for a debriefing meeting. This meeting will be held via Adobe® Connect™. Additionally, a link to a social validity measure will be provided via email to you.

All sessions will be audio and video recorded for data collection purposes. These measurements will be collected by members of the research team, including my Dissertation Chair.

Risks to Participation: Perceived risks include the possibility of children becoming frustrated if they make errors or do not meet their improvement goal. Since sessions will be broadcast live between your child and myself, an additional risk may be invasion of privacy if other home-related activities are occurring during the session and may be captured either on video or through audio. Also, if sessions do not occur at home or with a secure internet source, there is risk of breaches of confidentiality this include personally identifying information that will be collected as well.

To minimize these potential risks, your child will be provided with frequent feedback focusing on individual improvements and growth. To encourage children to try their best, they will be provided with frequent praise for participating throughout the session. Additionally, the structure of the brief 10 second sprints will also minimize the duration that your child may feel frustrated. To minimize the risk of breaches of confidentiality, it is recommended that all sessions occur from a quiet, distraction free area of your home and use of a secure, password protected internet connection. The Adobe® Connect™ room for your child's sessions will be protected through a unique access code that only you and I will have access to. Furthermore, the use of the four-letter identifier rather than the use of your child's names on the datasheet will assist in preventing breaches of confidentiality.

Benefits to Participants: Possible benefits to your child include increases in his or her reading ability. Your child will also participate in performance monitoring and goal setting which may support him or her in future self-directed goal development. The online model will allow for the benefit of your child's participation in sessions regardless of his or her physical proximity to me. Possible benefits to the field include increases to the effectiveness of teaching procedures regarding reading fluency, procedural structure that enhances skill retention, and evaluating the online format as a means of increasing reading fluency. Additional potential benefits include extending the literature base for fluency-based educational practices, improving systems for reading instruction, and establishing a more rich literacy community.

Alternatives to Participation: Participation in this study is voluntary. You may withdraw from study participation at any time without any penalty. The session will terminate if your child declines to participate any further in the session.

Confidentiality: During this study, information will be collected about you and your child for this research. This includes your child's name, your name, home phone and/or cell phone number, email address, medical information, age and gender. Data regarding your child's performance on the reading tasks will also be collected throughout the study.

All personally identifying information will be kept in a locked file. To protect confidentiality, all data will be kept in my care and be coded with a four-letter identifier to separate name from performance data. Research files will be kept in a secure electronic cloud drive. Only pseudonyms (fake names selected by your child) will be used in data reporting.

Data will be utilized for the proposed research project. If the research project is submitted for publication or presentation, all identifying information will be removed to protect your and your child's confidentiality. If you approve of the audio and video recording being used for presentation on the study, please initial the box below. If not, please place an X through the box.

Initial providing support for presentation of video and audio content: ☐

Research files and session videos will be kept in a secure electronic cloud drive for a minimum of 5 years prior to destruction. During this time, myself and my dissertation chair will have access to the electronic files via secure password. After a minimum of 5 years, electronic data files will be permanently deleted by wiping them from the hard-dive they were stored and permanently deleting the files from the electronic cloud drive. Hard copies of data sheets will be stored in a locked cabinet at my home office and after a minimum of 5 years, will be shredded.

Your research records may be reviewed by federal agencies whose responsibility is to protect human subjects participating in research, including the Office of Human Research Protections (OHRP) and by representatives from The Chicago School of

Professional Psychology Institutional Review Board, a committee that oversees research.

Questions/Concerns: If you have questions related to the procedures described in this document, please contact Lacy Knutson (Principal Investigator) at lmk4193@ego.thechicagoschool.edu or by phone at (651-247-1841) Dr. Julie Ackerlund-Brandt (Dissertation Chair) at jbrandt@thechicagoschool.edu or by phone at (715) 456-1707.

If you have questions concerning your rights in this research study, you may contact the Institutional Review Board (IRB), which is concerned with the protection of subjects in research project. You may reach the IRB office Monday-Friday by calling 312.467.2343 or writing: Institutional Review Board, The Chicago School of Professional Psychology, 325 N. Wells, Chicago, Illinois, 60654.

Consent to Participate in Research

Parent/Guardian/Legally Authorized Representative:

I have read the above information and have received satisfactory answers to my questions. I understand the research project and the procedures involved have been explained to me. I give my permission for my child/relative/conservatee to participate in this research project. My child/relative/conservatee's participation is voluntary and I do not have to sign this form if I do not want him/her to be part of this research project.

I will receive a copy of this consent form for my records.

_____

Name of Child/Relative/Conservatee Participant (print)

_____

Name of Parent/Guardian/Legally Authorized Representative (print)

_____

Signature of Parent/Guardian/Legally Authorized Representative

Date: _____

_____

Name of the Person Obtaining Consent (print)

_____

Signature of the Person Obtaining Consent

Date: _____

Appendix C: Pre-Eligibility Screening

Participant Pre-Screening (Inclusion/Exclusion Criteria)

Demographics	
Participant Age	
Can your child read basic words aloud to you? If you pointed out a common word such as 'the' in a book, could your child read it to you?	
Can your child sit in front of a computer and attend to the screen for brief amounts (30s – 1 min) without your assistance or support?	
Can your child access electronic devices with minimal assistance?	
To the best of your knowledge, does your child currently have or have they ever had a diagnosis of photosensitive epilepsy?	
To the best of your knowledge, is there any medical reason why your child should not participate in this study?	

Eligible Script

To be read if individual meets initial eligibility requirements:

"Mr./Mrs. (name), based on your responses to the pre-screening questions, (child's name) is eligible for participation in the study. If you'd like, we can proceed to the pre-test phase where we will identify a set of words your child knows how to read. Would you like for your child to continue?"

Ineligibility Script

To be read if the individual does not meet initial eligibility requirements:

"Mr./Mrs. (name), based on your responses to the pre-screening questions, (child's name) is not eligible for continued participation in the study. If you'd like, I'd be happy to answer any questions you might have. If not, I greatly appreciate you taking time to answer my questions."

Appendix D: Sample Pretest Slide

the	up	one	here
to	look	my	help
and	is	me	make
a	go	big	yellow
I	we	come	two
you	little	blue	play
it	down	red	run
in	can	where	find
said	see	jump	three
for	not	away	funny

Appendix E: Demographic Questionnaire

Participant Demographic Information

Demographics	
Parent/Guardian Name	
Parent/Guardian Email	
Parent/Guardian Phone	
Participant Location (State)	
Participant Name	
Participant Age	
Grade Level	
Race/Ethnicity	
Participant's Primary Language	
Participant's Secondary Language (if applicable)	
Primary Language Spoken in the Home	
Is Your Child Exposed to a Different Language at School? (Y/N) If yes, what language?	

Technology & Computer Specifications	
Do You Have Access to a Personal Computer? (Y/N)	
Is it a Laptop or a Desktop?	
What Brand? (Windows / Mac)	
Do You Know the Screen Size? (Y/N) If yes, please provide.	
Do You Know the Screen Resolution? (Y/N) If yes, please provide.	
Do You Have Access to the Internet? (Y/N)	
Is Your Internet Connection Wireless or Direct Connection (Hard-lined)?	
Do You Have a Webcam?	

Is the Webcam Internal or External?	
Do You Have Access to Speakers?	
Are the Speakers Internal or External?	
What Level of Volume is Most Appropriate? (Provide Number)	

Documents – To be completed by PI as occurs based on participant progress	
Consent Provided for Participation?	
Preference Inventory Returned?	
Retention Check Completed? (1 wk)	
Retention Check Completed? (3 wk)	
Retention Check Completed? (5 wk)	
Retention Check Completed? (8 wk)	
Social Validity Questionnaire Completed?	

Appendix F: Preference Inventory

Preference Inventory

Purpose: As your child participates in the study and works to improve his or her reading fluency, I want to ensure I can provide a meaningful "thank you." Completing the table below will help me to identify a variety of items/activities that can be offered to your child as a, "Great job! Thank you!" for working hard and meeting their improvement goals.

Instructions: Please identify 3-5 items/activities in each presented category that you would be willing to assist your child in accessing if they meet 2 out of 3 of their session goals (note: I will communicate this with you at the end of each session). Keep in mind, these should be special items/activities that they may not get on an everyday basis. Some examples may include: 10 extra minutes of TV, getting to watch a special episode of a favorite show, 10 minutes on the trampoline/video gaming system, special snack, etc.

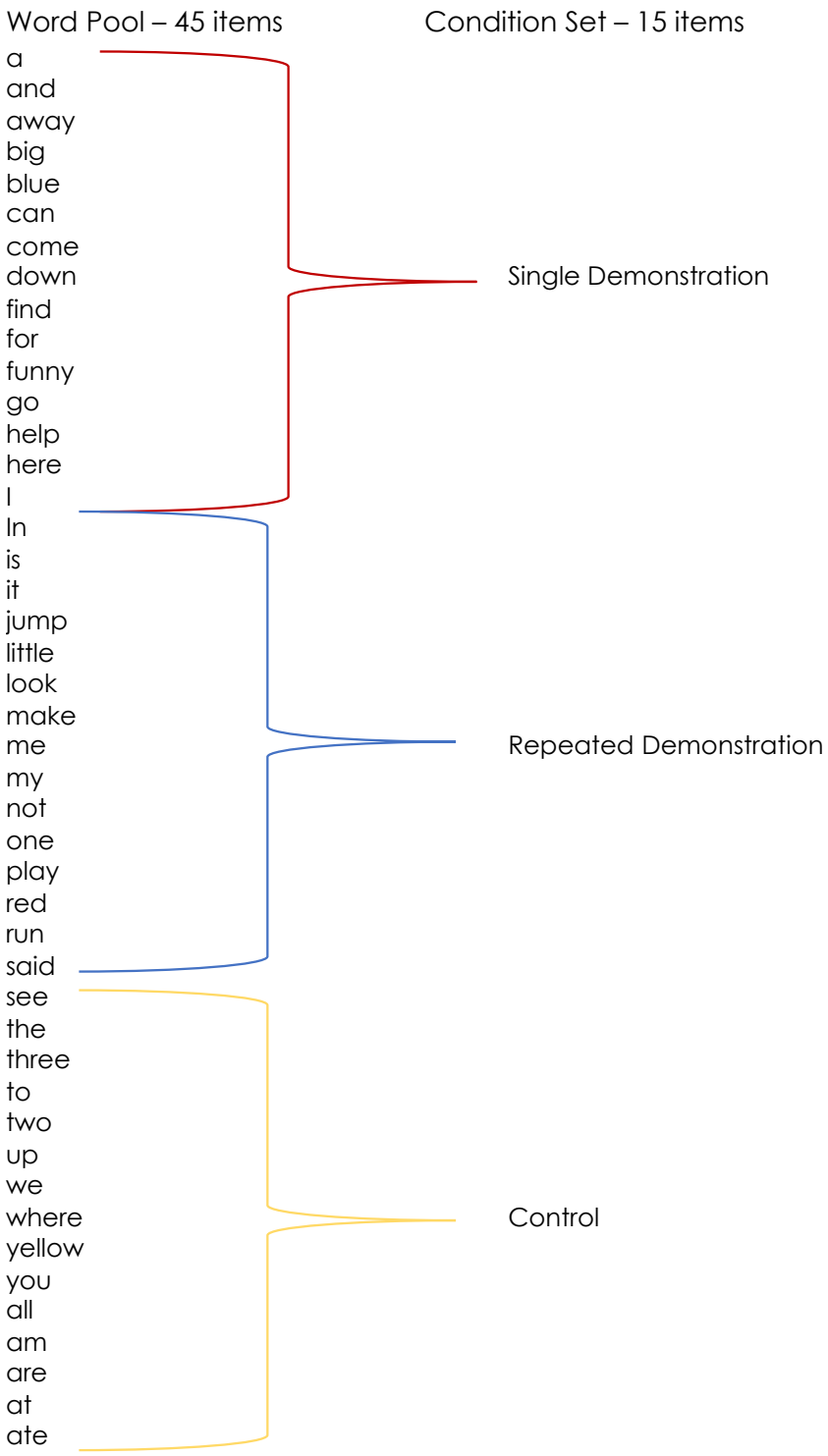
Menu		
	Activities	**How much or how long would you allow?**
	Example: iPad time	*10 minutes*
1		
2		
3		
4		
5		
	Social	**How much or how long would you allow?**
	Example: 1:1 time with mom	*20 minutes*
1		
2		
3		
4		
5		
	Snacks	**How much or how long would you allow?**
	Example: fruit snacks	*1 personal package*
1		
2		
3		
4		
5		

If you have any difficulty completing this, or questions on possible options, please let me know. You can reach me at lmk4193@ego.thechicagoschool.edu or 651-247-1841.

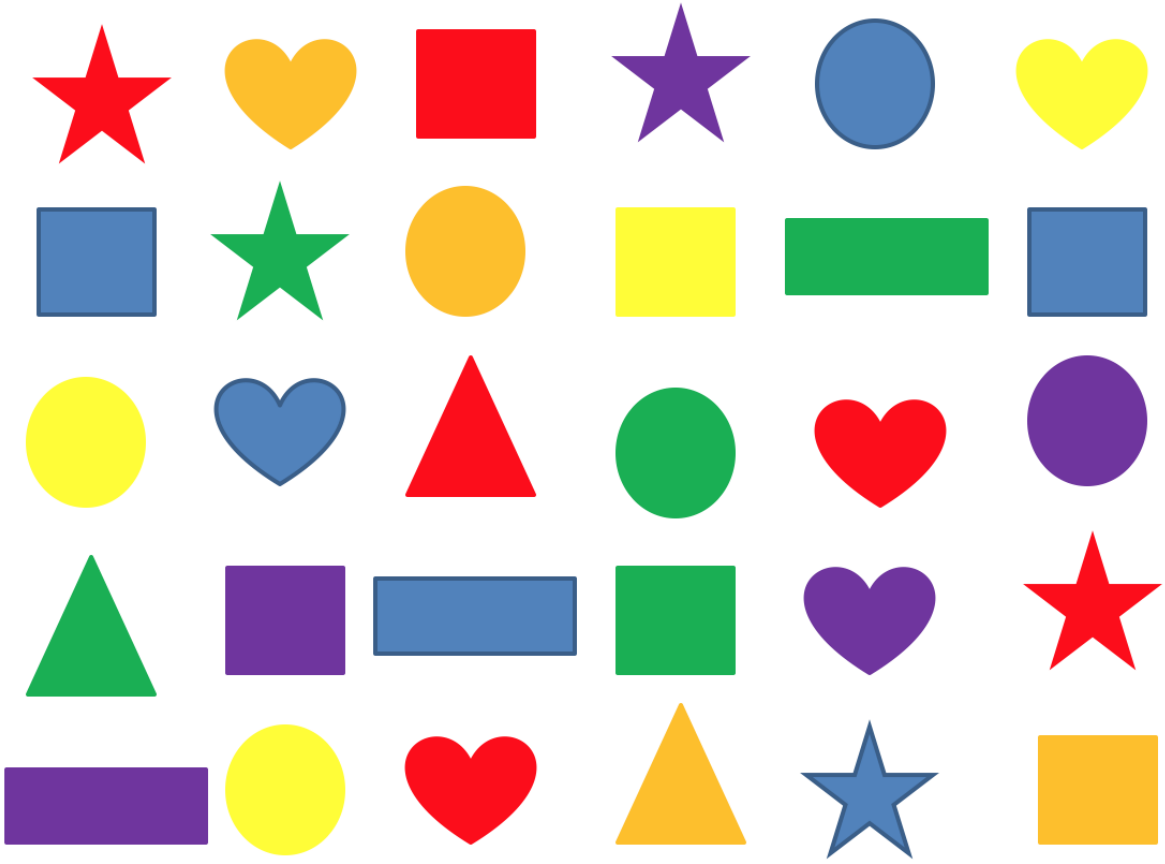
Appendix G: Dolch Words

Complete Dolch Word List Divided by Level										
Pre-primer	Primer			Grade One		Grade Two			Grade Three	
a	all	under		after		always	why		about	
and	am	want		again		around	wish		better	
away	are	was		an		because	work		bring	
big	at	well		any		been	would		carry	
blue	ate	went		ask		before	write		clean	
can	be	what		as		best	your		cut	
come	black	white		by		both			done	
down	brown	who		could		buy			draw	
find	but	will		every		call			drink	
for	came	with		fly		cold			eight	
funny	did	yes		from		does			fall	
go	do			give		don't			far	
help	eat			going		fast			full	
here	four			had		first			got	
I	get			has		five			grow	
in	good			her		found			hold	
is	have			him		gave			hot	
it	he			his		goes			hurt	
jump	into			how		green			if	
little	like			just		its			keep	
look	must			know		made			kind	
make	new			let		many			laugh	
me	no			live		off			light	
my	now			may		or			long	
not	on			of		pull			much	
one	our			old		read			myself	
play	out			once		right			never	
red	please			open		sing			only	
run	pretty			over		sit			own	
said	ran			put		sleep			pick	
see	ride			round		tell			seven	
the	saw			some		their			shall*	
three	say			stop		these			show	
to	she			take		those			six	
two	so			thank		upon			small	
up	soon			them		us			start	
we	that			then		use			ten	
where	there			think		very			today	
yellow	they			walk		wash			together	
you	this			were		which			try	
	too			when					warm	

Appendix H: Word Pool to Experimental Word Sets



Appendix I: Preteaching Slide



Appendix J: Assent Script

Assent Script

At the beginning of each session, the researcher will obtain assent from the participant to participate in the day's session by stating the following:

"Hi (name), your (mom/dad) said that you could help me with a project on reading. Would you like to do that today?"

Examples of assenting to participate include:

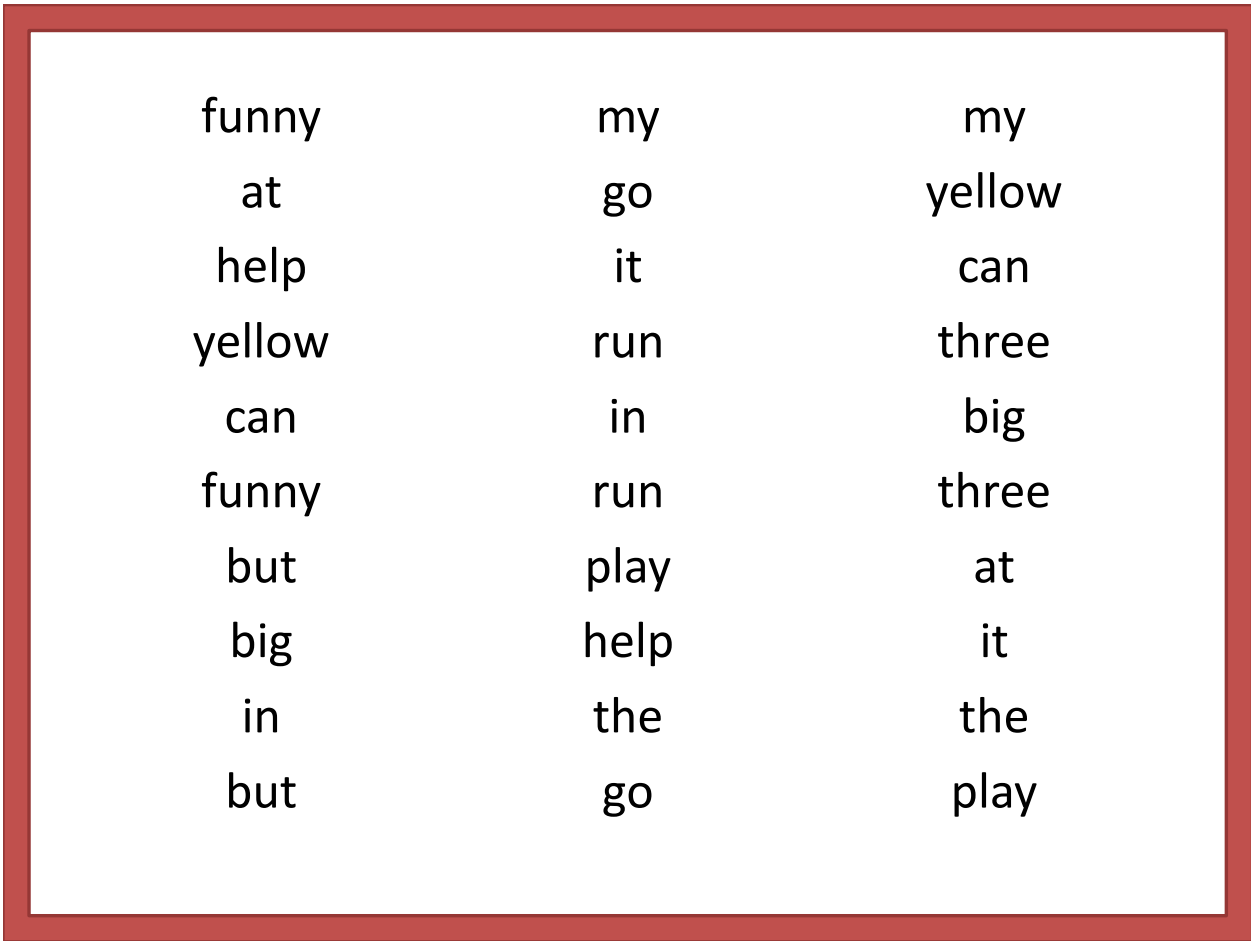
- Participants vocally stating, "yes," "yup," "yeah," "sure," "uh huh," or an equivalent agreement response
- Participants nodding his/her head in agreement
- Giving a thumbs up or "okay" symbol with their hand/hands

Examples of declining to participate include:

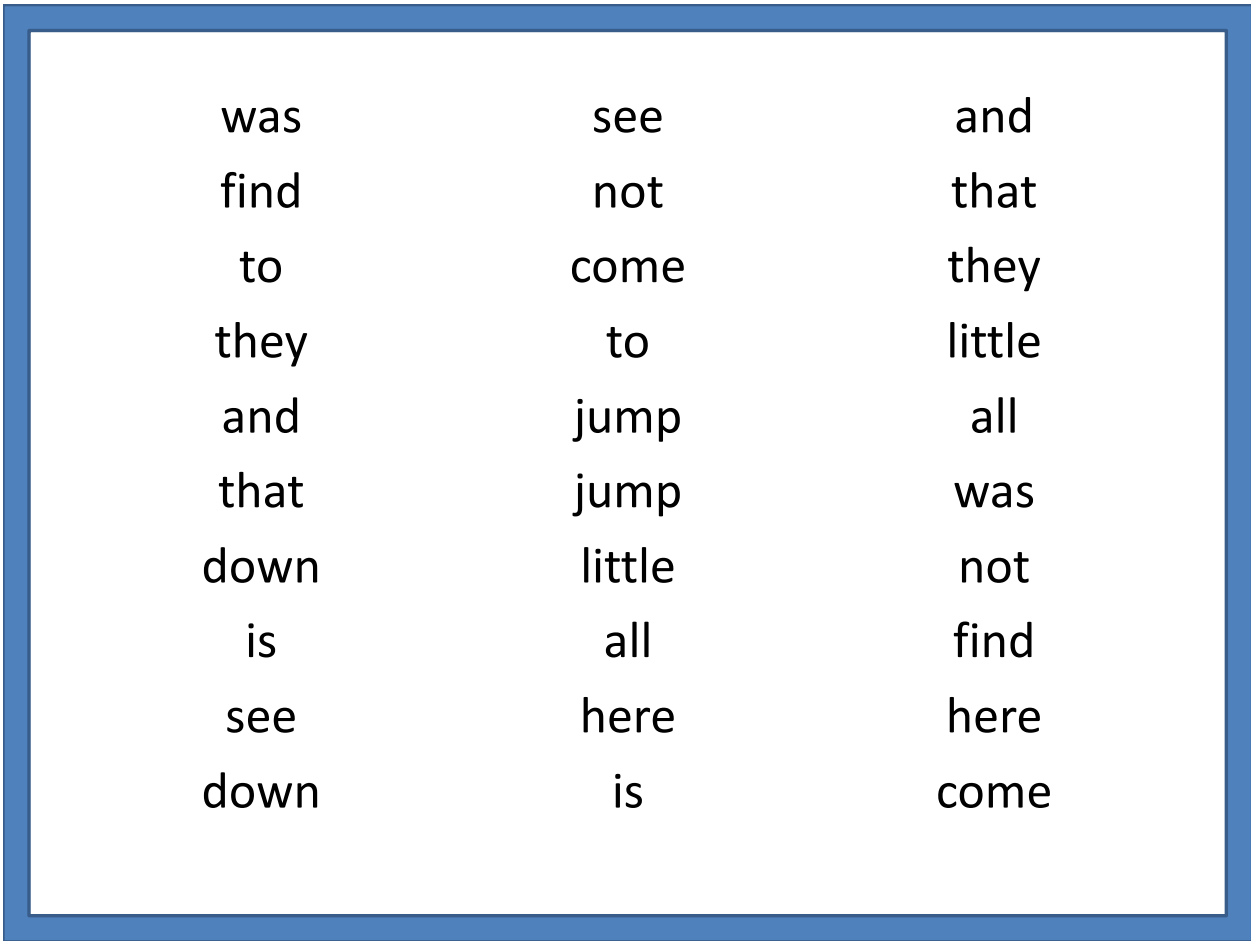
- Participants vocally stating, "no," "I don't want to," "no, thanks," "that's ok," "maybe another time," or equivalent response indicating a decline to participate
- Participants shaking his/her head in disagreement
- Giving a thumbs down symbol with their hand/hands

If the participant declines, the participant will be told, "That's okay, thanks for letting me know," and the session will terminate.

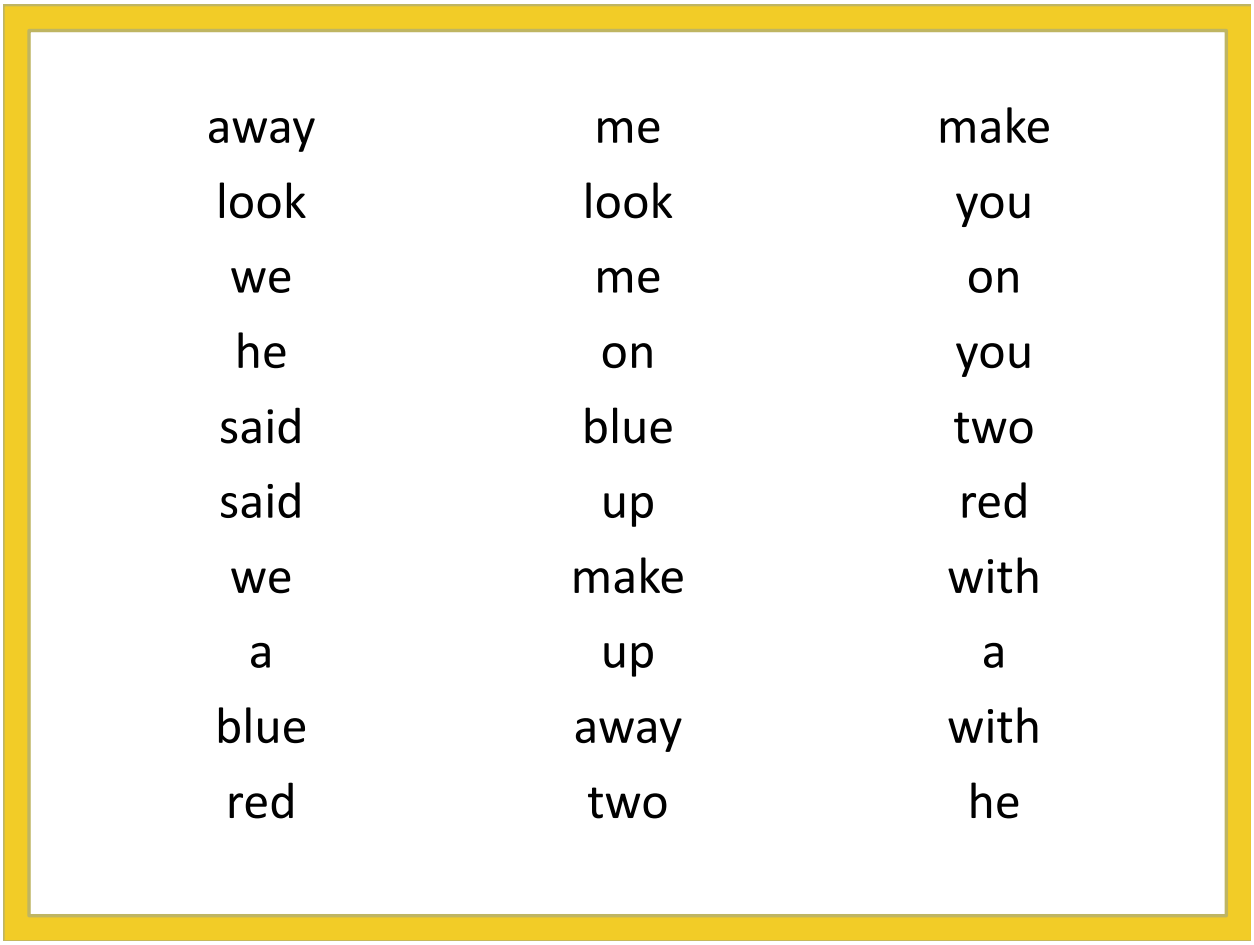
Appendix K: Single Demonstration Slide Example



Appendix L: Repeated Demonstration Slide Example



Appendix M: Control Condition Slide Example



Appendix N: Social Validity Questionnaire

Online Reading Program Feedback

Welcome & Thank You

Hello,
I first want to thank you for joining me on this exciting part of my PhD journey! I could not have done this without the support and dedicated time that you and your family kindly shared with me for this project.

As the research project comes to a close, I would like to gather feedback from both yourself and your child (the participant). The first few questions are for your child to respond. If you need to assist them in better understanding the question, you are welcome to do so. The remaining questions are for you to respond.

If you have any questions, please feel free to reach out to me:
lmk4193@ego.thechicagoschool.edu
651-247-1841

Sincerely,
Lacy

1

Online Reading Program Feedback

Please have your child respond to the following questions

1. After working with Lacy, I am a better reader

- ◯ Yes, definitely
- ◯ Maybe, I'm not sure
- ◯ No, I don't think so

2. After working with Lacy, I can read faster

- ◯ Yes, definitely
- ◯ Maybe, I'm not sure
- ◯ No, I don't think so

3. Please put the following word lists in order of your favorite (1) to least favorite (3)

- ☐ Red Word List
- ☐ Blue Word List
- ☐ Yellow Word List

4. I enjoyed working with Lacy

- ◯ Yes, definitely
- ◯ Maybe, a little
- ◯ No, not really

Online Reading Program Feedback

Please complete the following questions as the participant's parent

5. Having my child attend sessions with Lacy improved their reading ability

- ◯ Yes, definitely
- ◯ Maybe, I'm not sure
- ◯ No, I don't think so

6. Having my child attend sessions with Lacy improved their reading speed

- ◯ Yes, definitely
- ◯ Maybe, I'm not sure
- ◯ No, I don't think so

7. My child seemed to enjoy attending sessions with Lacy

- ◯ Yes, Definitely
- ◯ Sometimes
- ◯ No, I don't think so

8. The requirement of providing a fun activity for my child meeting their goals was easy to include into our routine.

- ◯ Yes, definitely
- ◯ Sometimes
- ◯ No, it disrupted our typical schedule

9. I would recommend this type of program to other parents

- ◯ Yes, definitely
- ◯ Maybe, I'm not sure
- ◯ No, I don't think so

10. Please feel free to provide any additional feedback/comments as you would like.

3

Appendix O: Procedural Integrity Checklist

Pre-Test Checklist

Step Number	Description	Y	N	N/A	Comment
1	**Assent** Hi _____, your (mom/dad) said that you could help me with a project on reading. Would you like to do that today?				
2	**Provides Instructions** I'm going to show you some words. You might know some of the words, and others you might not. That's okay. Just try your best. When one comes up on the screen, I want you to read it aloud for me. If you want to skip a word, you can say "skip" and go onto the next word. If you need a break, let me know. Ready to start?				
3	Cues stimuli				
4	Provides consistent feedback throughout pre-test				
5	Transitions to next word after 5s – approximated for delay				
6	Offers break if participant begins to fatigue				
7	Pauses session after approximately 45 words have been read correctly; counts and either ends session or has participant complete remaining word count				
8	Provides completion feedback – all positive				
		Y	N		Total
	Score				

Pre-Teaching Checklist

Step Number	Description	Y	N	N/A	Comment
	If Pre-Teaching starts the session, include step 0.5 otherwise mark N/A				
0.5	***Assent*** *Hi _____, your (mom/dad) said that you could help me with a project on reading. Would you like to do that today?*				
1	**Provides Instructions** We are going to practice doing sprints. Sprints are really short practice opportunities to show how fast and accurate we can be. This will just be to practice starting right away and stopping when the slide disappears. What is going to happen is, the screen is going to change and you will see a whole bunch of shapes. When the shapes show up on the screen, I want you to start saying the shape name right away. You will start at the top left corner (show with cursor) and read down the first column. You will name as many shapes as you can before the slide disappears. When it disappears, you will stop. I will practice one first to show you. Are you ready?				
2	Cues stimuli				
3	Demonstrates model				
4	Asks & answers any questions				
	Asks if participant is ready				
5	Cues stimuli				
6	Provides feedback regarding performance (emphasis on starting within 3 seconds of slide change)				
7	Provides up to 5 total trials if not starting within 3 seconds				
8	If initiating within 3 seconds, transitions to baseline session.				
9	If not initiating within 3 s after 5 total trials, ends the session.				
		Y	N		Total
	Score				

Baseline Sight Word Fluency & Retention Checks

Step Number	Description	Y	N	N/A	Comment
	If Pre-Teaching does not occur before baseline session, include step 0.5 otherwise mark N/A				
0.5	***Assent*** *Hi _____, your (mom/dad) said that you could help me with a project on reading. Would you like to do that today?*				
1	**Provides Instructions** I'm going to show you a list of words. You should know most, if not all, of them as I picked them from the words you read correctly for me the other time we met. If you don't know one, that's okay. Just try your best. When the list comes up on the screen, I want you to start reading aloud for me as fast as you can. The list will stay on your screen for just a short time so be quick. If you make it through all the words and they're still on your screen, start back at the beginning. We will read top to bottom and then go to the next column. If you want to skip a word, you can tell me "skip" and go onto the next word. Ready to start?				
2	Cues stimuli				
3	Prompts to continue if hesitates for 3sec – approximated for delay				
4	Provides participation feedback				
5	Provides brief 30 s - 1 min break (if conducting more than 1 baseline/retention sprint)				
		Y	N		Total
	Score				

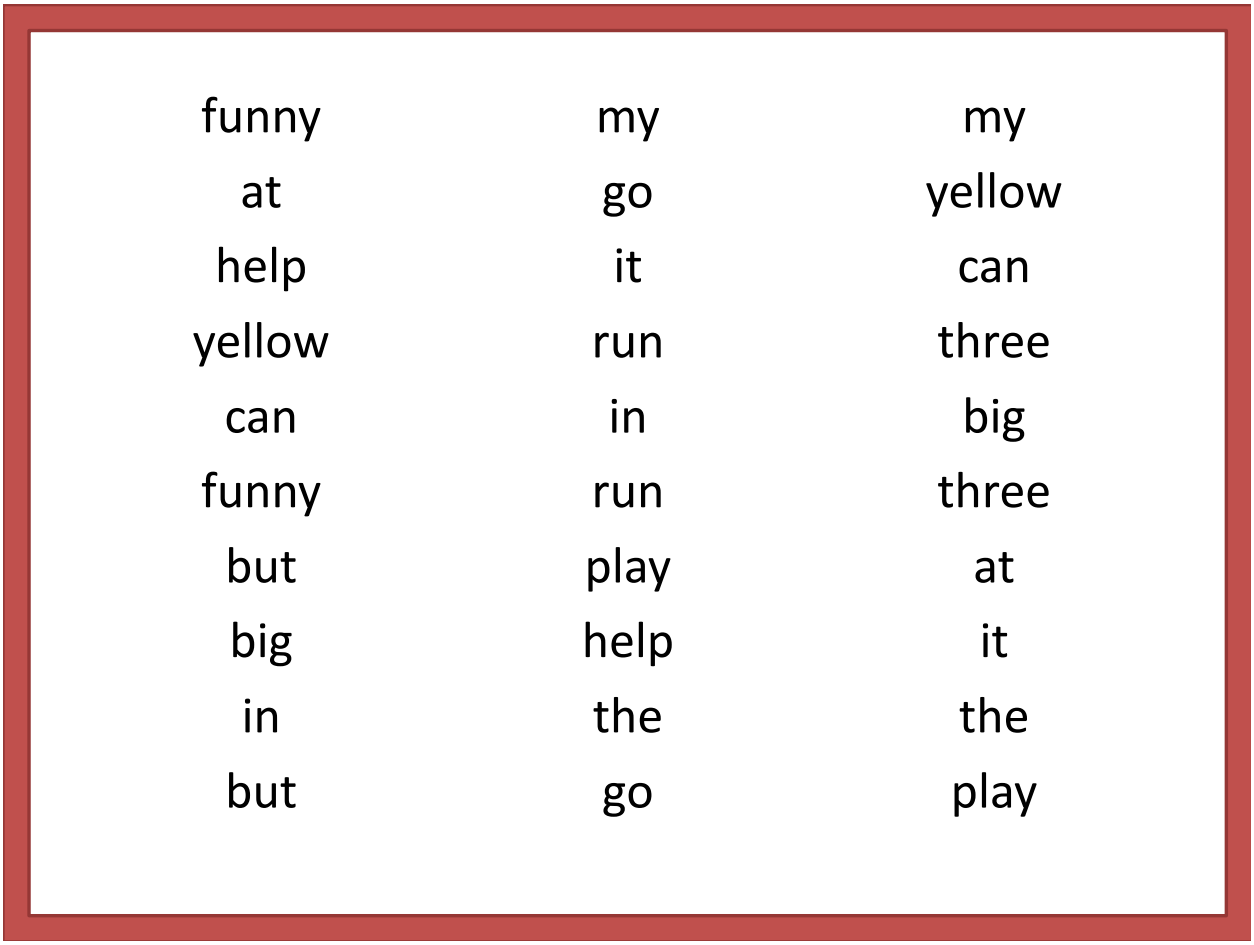
Interval Sprints

Step Number	Description	Y	N	N/A	Comment
1	**Assent** Hi _____, your (mom/dad) said that you could help me with a project on reading. Would you like to do that today?				
2	**Change in Contingencies – *shared 1st intervention session only*** Today, we are going to do things just a little different than before. First, if you remember, each list of words we have read, have a different color around them. One is red, one is blue and one is yellow. These colors help me keep track of which words we are reading. I might reference them by their color as we go ok? Each colored list is going to have its own special goal and we are going to try make each time we meet. The goal is going to be based on how you did the last time we met and I'll let you know what it is before we do the list. Each time you make a goal, I'm going to let you know and put a * in the box on our screen. You'll have 3 chances to earn a star each time we meet. If you get 2 or 3 *'s you get to pick something fun from the list you helped your mom fill out. How does that sound? If we don't get 2 *'s, then we won't get to pick that day, but the next time we meet you can get another chance to make your goals and pick something. Deal? **Token Check –** Okay, so how many *'s are we trying to get? –Answer: 2 or 3 Also, on some days I'll pick which order the lists go in, and on other days you'll get to pick, okay? After we do our sprints, you'll help me put our scores into a program online that shows how we are doing. Once that is done, we can let your mom/dad know if we met our goals or not. Ready?				

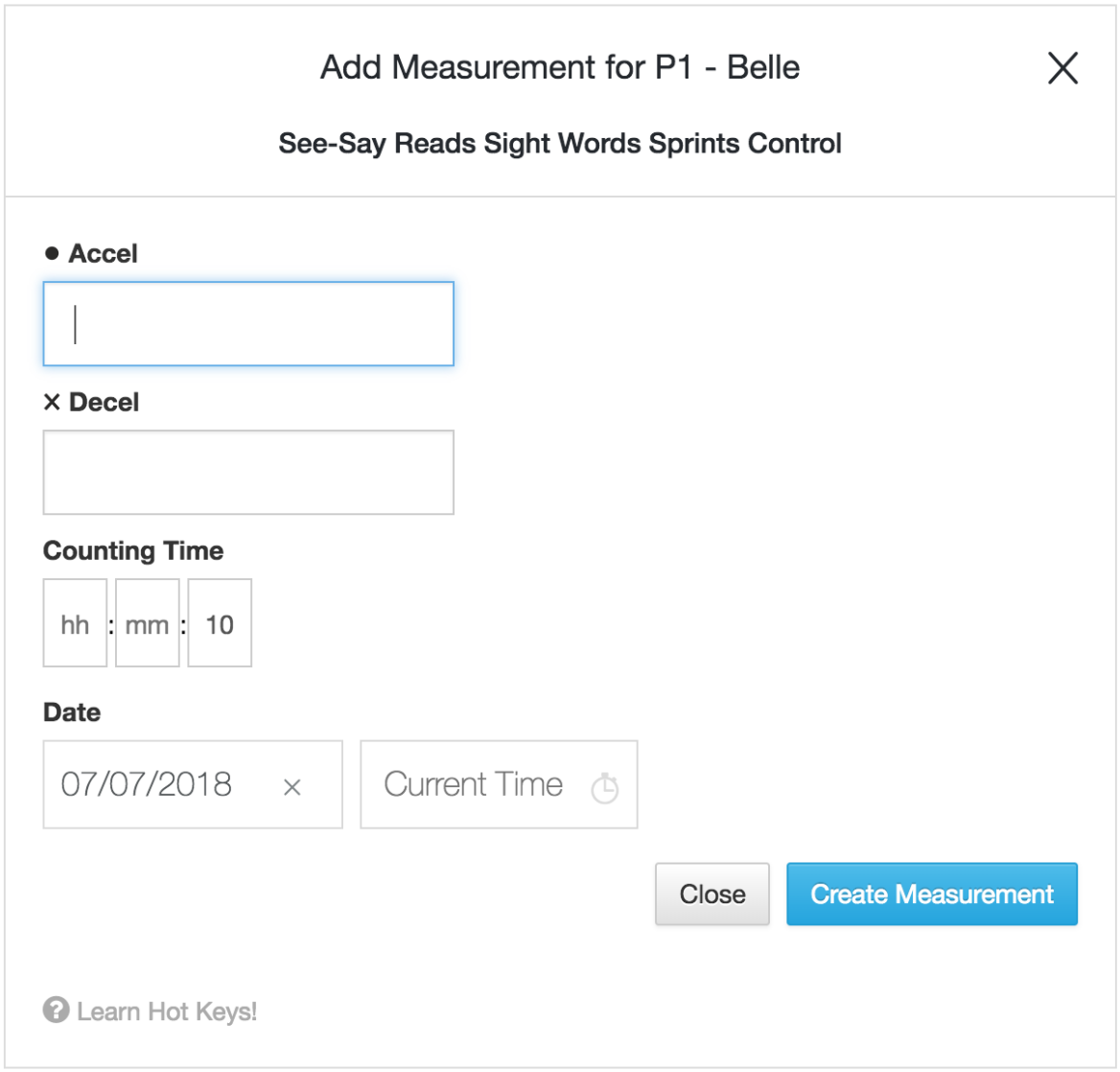
3	**Provides Instructions** I'm going to show you a list of words. When it comes up on the screen, I want you to start reading aloud for me as fast as you can. This list will stay on your screen for just a really short time so be quick. If you make it through all the words and they're still on your screen, start back at the beginning. We will read top to bottom and then go to the next column. If you want to skip a word, you can tell me "skip" and go onto the next word.				
4	**Intervention Preference Assessment** Today, I am going to pick the order **OR** You can tell me the order that we do the lists				
a	Cues choice slide if participant choice day				
5	Ready to start?				
	Improvement goal – before each sprint Last time, our highest score was _____ today we are trying to beat _____.				
6	Cues stimuli				
a	Prompts to continue if pauses for 3s				
b	Provides performance feedback after sprint (C/I)				
c	Provides Hear-Say Error Correction for I				
d	Provides feedback for completing Error Correction				
	Order of conditions should match step 5				
7	Conducts 1-3 sprints for Single Demonstration Condition (Red)				
8	Conducts 3 sprints for Repeated Demonstration Condition (Blue)				
9	Conducts 3 sprints for Control Condition (Yellow)				
10	Provides token for meeting improvement goal for a condition (single, repeated) in chat box. Hint: Single – 1 10% improvement sprint Repeated – 3 consecutive 10% improvement sprints (may cross sessions) Control – always provided a token				
11	Participant assists data entry into Chartlytics				
12	Reviews token accumulation at end of session				
13	If earns a minimum of 2 tokens, parent is notified of earned activity				

14	Receives verbal confirmation from parent regarding the earning of the activity				
		Y	N		Total
	Score				

Appendix P: Session PowerPoint Example Slide



Appendix Q: Chartlytics Online Worksheet



Appendix R: Chartlytics Standard Celeration Chart Display