# PILOTING A TOKEN ECONOMY INTERVENTION DURING VIRTUAL READING TUTORING IN AN APPLIED SETTING: A PROOF-OF-CONCEPT STUDY

By

Sage E. Pickren

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Special Education

May 12, 2023

Approved:

Laurie Cutting, Ph.D.

Marcia Barnes, Ph.D.

Andrea Capizzi, Ph.D.

Jason Chow, Ph.D.

Joseph Wehby, Ph.D.

## ACKNOWLEDGEMENTS

This research was funded in part by a grant from the Association for Positive Behavior Support.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

# CHAPTER 1

## Introduction

Reading is an important academic skill that influences life outcomes like employment and incarceration (Christle & Yell, 2008; Kern & Friedman, 2009) and is required to complete functional tasks like reading instructions, filling out tax forms, and paying bills. Nonetheless, 67% of 4th graders in the United States are not performing at a proficient or advanced level in reading [National Center for Education Statistics (NCES), 2022]. Minority groups, in particular, and students from families with low socio-economic status (SES) are at high risk for reading failure (NCES, 2022).

In spring of 2020, the COVID-19 pandemic created a need for social distancing across the world. Many school districts responded by implementing virtual learning. Though students have been able to return to school as the pandemic has slowed, options for hybrid and virtual learning remain in place. Students who were receiving in-person, one-on-one reading tutoring may now be receiving their instruction virtually. Many of these students who have been identified as needing intensive reading instruction likely also engage in inattentive and challenging behavior that impedes their access to virtual instruction (Aaron et al., 2002; Oakes et al., 2010; Vaughn et al., 2002). In fact, Lamb et al. (2020) explains that there are several risk factors that predict virtual learning failure including poor reading, low SES, and disability diagnosis. Critically, these are the children that are most at risk for being causalities of what has been coined as the "COVID-19 slide," or academic setbacks due to extended school closure (Kuhfeld & Tarasawa, 2020). In fact, data from the 2022 National Assessment of Educational Progress (NAEP) report revealed that on average student reading scores in 3rd grade dropped five

points during the pandemic. For children who were in the 10th percentile in reading, their scores had a more dramatic slide, dropping 10 points (NCES, 2022). The focus of the current project is therefore to examine ways to facilitate virtual engagement in reading instruction for populations that may be especially vulnerable to the COVID-19 slide: those with behavioral concerns and reading deficits.

**Comorbidity of Reading and Behavior Problems**

It is well documented in the literature that many students with behavior problems are poor readers (Arnold et al., 2005; Nelson et al., 2004) and that many students with learning disabilities have more behavior problems than their typical peers (Horbach et al., 2020). In fact, comorbidity rates between emotional and behavior disorder (EBD) and learning disability (LD) have been reported to be as high as 50% (Glassberg et al., 1999). Comorbidity between reading disability and behavior disorders has been studied for almost 50 years, beginning when Rutter and colleagues (1976) discovered that the children with reading problems on the Isle of Wight were four times as likely than their typically developing peers to exhibit antisocial behavior. Recent meta-analyses have shown that students with reading disabilities have higher levels of internalizing and externalizing behavior when compared to typical peers (Donolato et al., 2022; Francis et al., 2019) and that there is a significant, negative relation between reading and externalizing behavior (Pickren, Chow, & Cutting, in prep).

There are several theories that strive to explain why children with reading difficulties have cooccurring behavior problems. One supposition stems from the psychology literature and explains comorbidity as a function of associated or correlated liabilities (Krueger & Markon, 2006), meaning that a third variable explains deficits or symptoms in the two cooccurring disabilities. Under this model, the relation between reading and behavior might be accounted for

by shared cognitive factors that correlate with deficits in both domains (Hinshaw, 1992). For example, language deficits may explain why children struggle with reading and engage in problem behavior. Chow and Wehby (2018) conducted a correlational meta-analysis and found that there was a significant, negative concurrent and longitudinal correlation between oral language deficits and problem behavior, which persisted over time and across ages. Additionally, oral language is a critical skill for both word reading (NICHD, 2005) and reading comprehension (Foorman et al., 2015). Given the influence that language has on both problem behavior and reading, it is possible that language deficits are a correlated liability and can underlie why students struggle in both domains.

Correlated liability has also been examined through behavior genetics. Researchers have investigated the heritability of reading and behavior disorders and identify inattention as being a factor that links reading and behavior. Willcutt (2014) describes behavioral genetic methods used to explain the comorbidity between reading and attention-deficit hyperactivity disorder (ADHD), including twin studies that tease apart common genetic influences (e.g., Willcutt et al., 2007). In a study that examined comorbidity between reading disability and math disability, Willcutt and colleagues (2013) found that children with reading disability exhibited more internalizing and externalizing behavior symptoms and increased ADHD symptomology than those without disabilities. Within the same sample, groups who met criteria for oppositional defiant disorder (ODD) and conduct disorder (CD) were restricted to those that also had ADHD, which suggests that attention deficits may link reading disability and behavior disorders. This pattern was also found in Willcutt and Pennington (2000), where the presence of an ADHD diagnosis fully mediated the relation between reading disability and aggression, CD, and ODD. Importantly, liability models can be singular with one liability causing deficits in both disabilities or, more

likely, the relation is complex, and there are multiple liabilities causing deficits in both domains that have bidirectional influence (Krueger & Markon, 2006).

In contrast to the correlated liability model, under the causation model (also taken from the psychology literature) one disability directly causes another disability (Krueger & Markon, 2006). This influence can be unidirectional or bidirectional, though studies suggest that reading and behavior likely have a bidirectional relationship (Cook et al., 2012; Hinshaw, 1992; Morgan et al., 2008), with each skill deficit simultaneously worsening each other. The Transactional Model describes both how a child's behavior is shaped by their environment, and how the environment is shaped by the child's presence (Sutherland & Oswald, 2005). This theory can include interactions between teachers and students and explain how children engage with (or avoid) academic instruction. Specifically, The Cycle of Negative Reinforcement (Gunter et al., 1993; Gunter et al., 1994) describes a pattern that can explain why children with problem behavior have poor academic performance and tend to continue to engage in problem behavior over time. Children may engage in problem behavior to escape academic instruction. When the teacher removes the academic instruction in response to the child's problem behavior, the child's problem behavior is negatively reinforced and typically stops. Consequently, the teacher's behavior of removing instruction is also negatively reinforced by the cessation of the problem behavior. Teachers then engage in behavior patterns to avoid problem behavior from the child by consequently presenting less academic instruction. Since there are fewer opportunities presented to the child to participate in academic instruction and practice, learning loss occurs, and makes presentations of academic instruction (when they do happen) more aversive to the child, thus strengthening the cycle (see Appendix A). This cycle mirrors The Matthew Effect, which is a popular reading theory that describes why poor readers become worse readers over time

(Stanovich, 1986). When students struggle with reading, they tend to avoid reading. Avoiding reading prevents children from having exposure to good language models and new vocabulary (McNamara et al., 2011). Without this practice and exposure, children become worse readers and become frustrated by their failure and tend to continue to avoid reading (Echols et al., 1996).

**Token Economies**

Fortunately, there is a growing body of literature that suggests behavior interventions can support improvement in academic outcomes for students who struggle with both reading and behavior (Warmbold-Brann et al., 2017). There is much evidence indicating that Token Economies (TEs) are an effective intervention for in-person classroom settings for reducing challenging behavior and increasing academic skills (Maggin, et al., 2011; Soares et al., 2016). The theory behind why TEs work is based on B.F Skinner's operant learning theory, which describes behavior as being maintained by consequences, namely reinforcement and punishment (Skinner, 1948). If students' good behavior is reinforced (or rewarded), then that increases the likelihood that they will demonstrate that behavior again in the future. Briefly, a TE includes six components: a target behavior, tokens that function as conditioned reinforcers, backup reinforcers, a token production schedule, an exchange-production schedule, and a token-exchange schedule (Ivy et al., 2017). TEs are individualizable because there are options when implementing each component, namely when tokens are delivered and what the back-up reinforcement includes.

TEs are generally most effective for children ages 6-15 (Soares et al., 2016) and have been shown to improve reading outcomes (Gable & Shores, 1980; Solis et al., 2016) and increase attention for students with various disabilities (Aziz & Yasin, 2018). Kim et al. (2022) reviewed the TE literature from 2000-2019 and conducted a meta-analysis to describe the overall effect

size and important components of TEs for children in kindergarten through 5th grade. They found that TEs were highly effective in increasing engagement and decreasing disruptive behavior in both general education and special education classrooms, with a mean improved rate difference (IRD) of 0.83 and 0.87, respectively. Soares et al. (2016) also conducted a meta-analysis of single-case research on TEs and found a medium effect size of 0.82 using Tau-U. On the other hand, Maggin et al. (2011) reviewed the TE literature, and though they also found effect sizes that indicated TEs were effective, they concluded that there was not enough evidence to claim TEs were an evidence-based practice according to the What Works Clearninghouse standards (WWC, Kratochwill et al., 2010). This conclusion was mainly due to methodlogical flaws in the study designs. Importantly, lack of reporting on treatment fidelity was one of the most pervasive issues that harmed the methodological rigor of the studies. Furthermore, not collecting and reporting social validity data, through not directly related to internal validity of the studies, was another weakness across the TE literature, reported by Maggin et al. (2017).

**Virtual Learning**

Virtual learning, as compared to in-person learning, has the advantages of being able to reach students in remote locations, provide rich multimedia resources, eliminate costs related to travel and workplace rent, and be made available outside of typical school hours (Chen et al., 2005; Vasquez et al., 2011). Students with disabilities and those from low SES are choosing to participate in virtual learning at higher rates than other types of students (Thompson et al., 2012). However, most of the research that has been conducted surrounding virtual learning has been implemented with typically developing, college-aged students (Kara & Yildirim, 2020; Vasquez & Straub, 2012). Even with these adult samples, studies have found that students have poor engagement, lack of self-discipline, and exhibit poor interactions between students and

instructors (Regmi & Jones, 2020). Bernard et al. (2004) conducted a meta-analysis on virtual learning and compared it to in-person classroom learning for all educational levels, including post-secondary and professional adult training. After reviewing 232 studies, they found no difference between virtual learning and classroom learning in terms of overall effect size, but they found a tremendous amount of variability in effect sizes across studies. Similarly, Cavanaugh et al. (2004) meta-analyzed 116 effect sizes from 14 virtual learning programs in grades K-12 and found no statistical difference in performance between virtual and in-person learning. Critically, these reviews did not include students with disabilities and had very few studies that focused on reading outcomes for young children. Yet, these reviews suggest that virtual instruction can have similar effectiveness to in-person instruction, especially when considerations have been made to strengthen student engagement.

Virtual instruction can be implemented synchronously through live interactions (often through video conferencing) or asynchronously with teachers and students accessing material and communicating without real-time interactions. Chen et al. (2005) explain that synchronous instruction has advantages over asynchronous instruction, namely delivery of immediate feedback and increases in student motivation and involvement. When working with young children who need immediate academic and behavioral feedback and individualized instruction to acquire a skill like reading, synchronous instruction is crucial (Chen et al., 2005; Hastie et al., 2007).

There are unique challenges presented by virtual instruction, particularly related to behavior management, because of the physical distance between students and instructors. Kara and Yildirim (2020) studied virtual learning of college-aged students and reiterated the importance of positive and warm faculty-student interactions with more dialogue to mitigate the

psychological and communicative distance of virtual learning, something which is known to be especially important for children. In fact, Baker et al. (2008) found that, specifically for elementary school children with behavior problems, having a warm and trusting relationship with a teacher predicted positive school outcomes. Notably, Bernard et al. (2004) explains that the effectiveness of virtual instruction "depends on the provision of pedological excellence," meaning that teacher actions during instruction, like the use of behavior management strategies, can influence how much learning occurs during virtual instruction (p. 37). Cavanaugh et al. (2009) claims that virtual instruction for young children requires an "extensive reinforcement system" and "rewards for learning such as multimedia praise and printable stickers or certificates" (p. 7). Therefore, there is a need to improve synchronous, virtual, one-on-one reading tutoring to make it more effective and engaging for young, struggling readers (Coy, 2013; Vasquez & Serianni, 2012).

**Current Research on Virtual Instruction with Children**

As previously noted, very few studies examine virtual learning for elementary-aged students, especially from disability groups. In fact, Vasquez and Straub (2012) were only able to identify six empirical studies that were based on virtual instruction for students with disabilities in grades K-12. Only one of the six studies examined synchronous instruction (Yong & Ping, 2008), only one addressed reading outcomes (Vreeburg et al., 2010), and all employed a quasi-experimental group design. One exploratory study published during the pandemic (Beach et al., 2021), evaluated the virtual implementation of a one-on-one foundational reading intervention with rising second and third grade children from low-income families who were struggling with reading. The authors adapted the curriculum to a virtual format and found that students made significant improvements on curriculum-based reading mastery tests and that the tutors and

caregivers reported the tutoring program was feasible to implement. Critically, this exploratory study did not measure engagement or student disruptive behavior.

## Current Research on Virtual Behavior Interventions

There is a paucity of research which evaluates behavioral interventions embedded into virtual instruction, especially reading tutoring. Vasquez and Slocum (2012) conducted a multiple baseline across participants design to evaluate the effects of virtual Direct Instruction (DI; Engelmann, 1999) on reading fluency in a sample of struggling readers in 4$^{th}$ grade. Although they used a modified TE to manage behavior, behavior was not a measured outcome of the study and authors cited that it was still an obstacle that impeded reading instruction. Recently, LeJeune et al. (2022) evaluated the use of CW-FIT (a behavior intervention based on group contingencies) on full-class engagement during virtual reading instruction and found therapeutic effects. This study provides preliminary evidence that behavioral interventions that have a strong evidence base for in-person use may work in virtual settings with a full class of students. To my knowledge, the current study is the first study to evaluate the effects of a TE on engagement and disruptive behavior when embedded in virtual academic instruction.

## Study Goals and Theory of Change

The purpose of this study was to create a proof of concept for evaluating the effects of a TE on engagement and disruptive behavior when embedded into virtual reading tutoring in an applied setting. Additionally, this study aimed to evaluate the social validity and feasibility of intervention of a TE intervention in a virtual setting. A secondary purpose of the study was to determine if there were transfer effects of the behavioral intervention to reading fluency

performance. Specifically, I examined whether reading fluency performance increased as a function of embedding the TE into virtual reading instruction.

I hypothesized that a TE would improve engagement because behavior that was aligned with academic engagement was being reinforced. The Matching Law describes the tendency of humans to choose behavior that produces more reinforcement relative to other available responses (Herrnstein, 1961). Therefore, when the TE is in place, students should choose to engage in academic instruction rather than disruptive behavior. Under this theory of change, academic instruction cannot occur if it is being continually interrupted by disruptive behavior (Wehby et al, 2003). If disruptive behavior is reduced, then time for academic engagement will be increased; tutors will be able to focus their efforts on delivering academic instruction and in turn, students will spend a larger percentage of time engaged in the lesson.

In reference to reading performance, I hypothesized that oral reading fluency would gradually increase across tutoring sessions because students were receiving an evidence-based reading intervention. I also hypothesized that the average oral reading fluency performance would be higher in intervention sessions than it would be in baseline sessions. Notably, oral reading fluency performance is distinct from reading achievement or learning (Soderstrom & Bjork, 2015) and is particularly susceptible to attentional or behavioral disruptions because the tasks are timed. If the TE effectively reduced disruptive behavior and increased engagement during intervention sessions, then students would have more time to focus on reading during the one-minute timing and thus perform better during intervention. On the other hand, during baseline sessions the one-minute timings of oral reading fluency may be impeded by disruptive behavior and disengagement and cause students to perform worse when the token economy intervention was not being implemented.

This theory of change follows the causal model of comorbidity, in which environmental factors influence deficits in both reading and behavior, rather than genetic or correlated liabilities. Embedding a behavioral intervention into academic instruction may interrupt the cycle of negative reinforcement and create opportunities for students to engage in instruction and rewards them for doing so.

**Vanderbilt Kennedy Center Reading Clinic**

To accomplish the study goals, I used a resource within Vanderbilt University: The Vanderbilt Kennedy Center's Reading Clinic (VKC-RC). This clinic offers virtual reading tutoring to students who have demonstrated or are at risk for reading failure. The students have a wide range of disabilities including dyslexia, ADHD, down syndrome (DS) and autism spectrum disorder (ASD). While the VKC-RC has historically provided invaluable services to the community, the rapid shift to virtual learning due to COVID-19 highlighted the extensive lack of tools available to address challenging behavior in this new learning environment. This made it difficult for tutors to manage engagement and challenging behavior from their tutees. Tutors have expressed concerns over managing student behavior and have cited issues like students not paying attention, leaving the computer, playing with toys, refusing to participate in lessons, crying and complaining. Therefore, this project aimed to produce results that can inform virtual reading tutoring instruction and serve as a model for behavioral supports that can be used with virtual school. Additionally, this format of instruction has the capacity to reach populations of students that do not normally have access to high-quality reading instruction with behavior support, which made this endeavor worthwhile even now that social distancing regulations have been lifted and the reading clinic has switched to a hybrid model, serving students both in person and virtually.

Importantly, there are limitations to conducting research in an applied setting like the VKC-RC. The clinic setting had practical constraints that limit the experimental control over the study. First, the tutors who provided instruction to the students were pre-service teachers who have varying levels of training and experience with teaching. This clinic setting served the dual purpose of creating learning opportunities for both the students receiving tutoring and the tutors conducting tutoring. Second, the tutor-student pairings were predominately dictated by student and tutor schedules, both of whom have outside class responsibilities. Uneven numbers of tutors and students also created pairings that were not always one-to-one formats. Sometimes two tutors shared one student, or one tutor instructed two students at a time. Third, although participating tutors were trained on study procedures, they were not paid research assistants, and therefore they had to balance responsibilities of serving their clinic students and participating in the research study. The tutors' priority was making choices that best meet the instructional and practical needs of their students. In certain cases, experimental rigor was sacrificed to preserve the tutoring experience for the students. Considering these limitations, this study was situated as a proof of concept for conducting single-case design research in an applied setting.

**Research Questions**

1. Does a TE embedded into virtual reading tutoring increase engagement and decrease disruptive behavior for struggling readers?
2. How feasible is implementing a TE through virtual instruction and what do relevant stakeholders report about the social validity of this intervention?
3. Does oral reading fluency performance improve during tutoring and increase as a function of implementing a TE during virtual reading tutoring?

# CHAPTER 2

## Method

First, I wrote an application to Vanderbilt's Institutional Review Board to gain approval to run the study. Notably, this was the first study conducted within the context of the VKC-RC. Once approved, I recruited participants by contacting the VKC-RC director. She referred students who met the eligibility criteria listed below and put me in contact with their caregivers. I met with each family to obtain informed consent and student assent, prior to data collection. Once participants were consented, I met with the tutors assigned to the participants to obtain their informed consent and train them on study procedures.

### Participants and Tutors

#### *Participant Inclusion & Exclusion Criteria*

The participant criteria to be eligible for the study were: (1) Students had to be between the ages of 6-12 because that is within the age range has evidence for being most responsive to TEs (Soares et al., 2016). (2) Students had to be behind in reading, as that is a criterion to be eligible for tutoring at the VKC-RC. For the purposes of this study, that was determined by having reading achievement scores that were at least one grade level below the mean and/or <40 percentile score on a standardized reading assessment (i.e., Test of Word Reading Efficiency [TOWRE]; Torgesen et al., 1999; Qualitative Reading Inventory [QRI]; Leslie & Caldwell, 1995). These assessments were completed upon intake into the VKC-RC and were already completed at the start of the study. (3) Tutor or parent had expressed concerns to the VKC-RC director during appointment debriefs about behavior and/or inattention during reading instruction. This could be reported as issues with virtual schooling outside of the VKC-RC.

Students were *excluded* from the study if they (1) had a hearing or visual impairment, (2) had an intellectual disability, (3) or if they were on medication for behavior that had been newly introduced or would be changing throughout the course of the study.

Once it was determined that students were eligible for the study, and after I collected consent, I collected initial baseline data during reading tutoring for all consented students. To be considered eligible to receive the TE intervention, participants had to demonstrate low baseline engagement. If participants had average initial baseline engagement lower than 80%, they were eligible for intervention. If participants had average initial baseline engagement higher than 80%, I determined that behavioral intervention was unnecessary, and they were discontinued from the study.

***Spring 2022 Participants***

Three participants participated in the study during Spring 2022. Melinda was a twelve-year-old, Black girl, with diagnoses of autism, dyslexia, and anxiety. She attended the reading clinic for several years prior to participating in the study and the VKC-RC director reported that her behavior had improved over time. The director also reported that her anxiety seemed to be highest when she felt overwhelmed with tasks that she perceived to be above her ability level. She showed avoidant behaviors when tasks got too complex. Recently she was placed in a private school for her general schooling and had responded well to intense behavior plans that were set in place at school. At the start of the intervention, the VKC-RC director reported that Melinda was reading on a 1$^{st}$ grade level and working on decoding vowel teams and multisyllabic words.

Cody was an eight-year-old, Black boy with ADHD and Ryan was an eight-year-old White boy without any diagnoses. Notably, both Cody and Ryan were discontinued from the

study due to high baseline engagement (see Appendix B) and I did not collect any further descriptive data on them (e.g., topography of behavior and current reading level).

### *Summer 2022 Participants*

Three additional participants participated in the study during Summer 2022. John was a nine-year-old, White boy with ADHD, ODD, auditory sensory disorder, and anxiety diagnoses. When he was younger, he received occupational therapy and had an expressive receptive speech delay that was resolved by speech therapy. During the study, he was not medicated. In the past, he was homeschooled, but his caregiver decided to start him in public school after the conclusion of the study. His caregiver reported that he was shy about receiving praise and was easily embarrassed. She also reported that he did not like to be told what to do but responded well to being given choices. The caregiver reported that John was reading on an early 1$^{st}$ grade reading level and was working on decoding consonant-vowel-consonant words (e.g., dog).

Alexander is John's younger brother. Alexander was a seven-year-old white male without any diagnoses. He too had an early speech delay that was resolved when he was younger. His caregiver described him as being "bossy, rough, and tough," who often got out of work by saying flattering statements to others. At the start of the study, the caregiver reported that he was reading on a Kindergarten reading level and learning letter names and sounds.

Ed is the younger brother of Ryan. He was a seven-year-old boy with no diagnoses. He had no prior reading clinic tutoring and no special services at his public school. Notably, Ed was discontinued from the study due to high levels of engagement during baseline (see Appendix B) and I did not collect any further descriptive data on him.

### *Tutor Information*

The only tutor inclusion criteria were to be (a) hired at the reading clinic as a paid tutor or practicum student and (2) be paired with a student who was eligible to participate in the study. During Spring 2022, six tutors participated in the study (two per student). During Summer 2022, two tutors participated in the study (one per student). In the spring, two tutors were paired with each student due to over-enrollment in a master's level course that completes field experience at the VKC-RC.

Marnie and Evan were paired with Melinda. Marnie and Evan were both master's students enrolled in a reading course that does field experience at the VKC-RC. They shared tutoring responsibilities to meet their course requirements. They both were in the first year of a two-year high-incidence special education program at Vanderbilt University. Neither tutor was in the applied behavior analysis program nor had a behavior certificate. Marnie had no prior teaching experience before starting the master's program. Evan taught students with disabilities for one year prior to the program. Marnie and Evan tutored Melinda together for the first four sessions. After that, they rotated sessions, with Marnie tutoring sessions 5, 7, 10, 12, 13, 15 and Evan tutoring sessions 6, 8, 9, 11, 14, 16.

The other four spring tutors were also in the master's program and tutored at the VKC-RC as part of their field experience. I did not obtain further information about their training or teaching experience, since their paired participants were discontinued from the study.

In the summer, Kiley served as both John and Alexander's tutor. She tutored them in back-to-back, separate sessions. Kiley was a paid VKC-RC tutor who had recently graduated from school with her undergraduate degree in special education from the severe disabilities program at Vanderbilt University. She had completed student teaching and held a teaching

license. She did not complete the applied behavior analysis program and held no behavior certificates.

Another student tutored Ed during the summer of 2022. She had just finished the first year of the two-year master's degree program in severe disabilities special education at Vanderbilt University. I did not obtain further information about her training and teaching experience because her paired participant was discontinued from the study.

## Setting

Each participant received a total of 20 hours of tutoring from the VKC-RC. Spring participants attended tutoring twice a week for 50 minutes each session. Data collection had a delayed start in the spring while waiting for IRB approval, which is why data are only reported for 16 sessions for Melinda. Summer participants attended tutoring five times a week for 50 minutes each session. Data collection started at the beginning of tutoring for summer participants. All tutoring was conducted virtually via Zoom. Participants and tutors attended the Zoom meetings from their homes. Caregivers were instructed to find a quiet place for their children to attend tutoring with stable internet and headphones with a microphone; however, the environment varied among participants and among sessions. Some participants attended tutoring from an iPad, though for the participants in which full behavioral data are reported, their sessions were done on a laptop or computer with a monitor.

## Materials

### *Reading Curricula*

The VKC-RC director decided which reading curriculum to use to best meet the instructional needs of the students. In the fall, tutors used The Road to Reading (Tangel &

Blachman, 2008) curriculum to deliver reading intervention. Road to Reading is a curriculum that has a strong phonics focus, which is a deficit of many struggling readers (Al Otaiba & Fuchs, 2002). This curriculum is flexible and adjustments to the curriculum were made concerning pacing, depth, and breadth to individualize the program, with help from the VKC-RC director (Fuchs et al., 2014; Lemons et al., 2014). As part of the VKC-RC procedures, prior to beginning the curriculum, students were given a placement test, which is a phonics inventory that aligns to the Road to Reading levels. Students focused on one phonics skill at a time (e.g., reading short vowel, one-syllable words with digraphs). To practice this phonics skill, students went through a five-step lesson including sound review, building words, reading words, sentence dictation, and reading in context. Tutors created their own PowerPoints to administer the Road to Reading Curriculum virtually.

At the start of tutoring, Melinda placed into the green level of Road to Reading. This level teaches students how to read one-syllable words with vowel digraphs and vowel diphthongs (e.g., green, flew). Melinda had already mastered reading one-syllable words with consonant digraphs and consonant blends and the CVCe long vowel pattern (e.g., gate) in prior levels. In addition to her instruction with the green level, she was also learning how to read very simple two-syllable words with short vowel sounds (i.e., CVC/CVC, e.g., picnic, bedrock). At the end of tutoring, Melinda was still working in the green level on vowel digraphs and vowel diphthongs, but she had advanced to reading multisyllabic words with more complex spelling patterns (e.g., baseball, outrun).

In the summer, tutors used Friends on the Block (Allor et al., 2018), an evidence-based curriculum that uses explicit modeling, cumulative review and practice, and feedback when teaching decoding skills (Allor et al., 2020). In addition to a focus on decoding, Friends on the

Block has a focus on learning sight words and reading connected text. Most of the lesson is conducted with a shared text. When conducting the lesson, the tutor reads the "helper text," which is at a higher reading level and supports vocabulary and comprehension growth. Then the student reads a simplified text that contains only decodable words and sight words. The curriculum also comes with learning games that helps tutors review letter sounds and sight words. Friends on the Block has electronic versions of their curriculum that were accessible to tutors to use for virtual administration of the curriculum.

At the beginning of tutoring, John placed into the "Preparing to Decode" range of Friends on the Block and began reading Level 1 books. In this level, students are taught letter names and sounds and practice phonemic awareness. They also memorize a small number of high-frequency words (e.g., a, do, not, like, want). At the end of tutoring, John had progressed to Level 3 books, which are still in the "Preparing to Decode" range. By this point, he had mastered most of his consonant sounds and memorized approximately 15 high-frequency words.

Alexander also began tutoring with Level 1 books. He had more trouble learning letter names and sounds and some lessons were supplemented with YouTube letter-sound videos. By the end of tutoring, Alexander had only progressed to Level 2 books and knew approximately 8 high-frequency words.

***Token Economy Materials***

Generally, tutors at the VKC-RC do not implement individualized behavior plans; therefore, the token economy procedures described below are specific to the study and have been added as a proof of concept to determine if this kind of intervention is feasible in this applied setting. The TE had several components, including a rule review, token delivery, and reward delivering. (More description about the TE intervention procedures can be found in the

independent variable section below.) I helped each tutor create tutoring rules for their participant with descriptions of what it means to follow each rule. These rules were reviewed at the start of each intervention session along with a reminder to students that when they followed the rules, they earned points that could be exchanged for a reward at the end of the session. Each tutor also interviewed their participant to find out what rewards they wanted to earn, and which was their favorite of the rewards. Then I helped the tutors create a menu of individualized rewards that had corresponding points associated with each reward. Though I did not conduct a preference assessment to determine reinforcement preference, rewards students expressed were their favorite were assigned a higher point value than lesser preferred rewards. The rule reminder and reward menu were inserted as the first slides into the tutoring PowerPoints that were used on intervention sessions and then the reward menu was made available at the end of tutoring for students to select their reward (see Appendix C).

Tutors used an automatically restarting, vibrating timer (i.e., MotivAider) to remind them when to award or withhold points during the intervention sessions. The vibration was intended to just be felt by the tutors, but on some recordings, the sound of the vibration was audible to students. Points were displayed as tallies using a notes application on the screen. This allowed points to remain visible to the participants on the top right-hand corner of the screen during intervention sessions.

**Tutor Training**

Prior to the start of the study, the tutors were trained by the VKC-RC director on virtual reading instruction delivery through Zoom. Therefore, I only trained tutors on the TE procedures, CBM-R administration, and general study guidelines. I met with each tutor individually on Zoom for training, which lasted about 45 min. During each training session, I went through a

PowerPoint that described the study procedures including informed consent, basic withdrawal design, study research questions, baseline procedures, TE intervention procedures, and data storage. For baseline, I explained to tutors that they were to deliver instruction in a "business as usual" way. They implemented reading tutoring procedures as the VKC-RC trained them to do. This included phonemic awareness activities, phonics instruction, reading in context, high-frequency word practice, and progress monitoring. They were instructed to not do anything for behavior management outside what they normally would do. For intervention, tutors were told to implement the TE and to not include any other behavior management procedures aside from the ones described in the TE. The TE procedures were explained with reference to the procedural fidelity form, which showed the tutor each step of the TE intervention that needed to be performed. As mentioned earlier, the steps included a rule review where participants reminded participants of the tutoring rules with examples using a PowerPoint slide. Next, participants were reminded that they were earning points during the session for following rules and that those points could be used to purchase a reward at the end of the session. Then, tutors were trained to use the vibrating timer with a 3-minute fixed interval and deliver tokens with praise if the students were following the tutoring rules. If the timer vibrated while the student was reading, the tutor was instructed to wait until the student finished the sentence and then interrupt with praise and a token. During training, I gave an example of behavior-specific praise and explained that behavior-specific praise referenced which rule students were correctly following. Last, tutors were trained to allow students to purchase a reward from the customized reward menu at the end of intervention sessions. Tokens could not be banked for a subsequent session, but rather had to be spent during the current session. Procedural fidelity was monitored throughout the study on

100% of sessions to ensure tutors were implementing the intervention as they were trained (see section on Procedural Fidelity).

## Data Collector Training

I hired and trained seven undergraduate-level data collectors on procedural fidelity observation and dependent variable measurement. I intentionally did not tell the undergraduate coders the purpose of the study or the research questions, to attempt to reduce bias in coding. However, it is possible they were able to guess the purpose of the study after watching intervention sessions and using procedural fidelity forms. For dependent variable measurement, I trained coders how to use the *Behavior Buddy* app and reviewed the coding manual with them. We also went over safe data storage procedures (i.e., using Vanderbilt Box to download videos and upload data). Prior to coding participant videos, coders had to code one training video and reach above 85% reliability on each variable.

## Measurement

### *Independent Variable*

The TE is the independent variable in this study, which was introduced and withdrawn from reading instruction. As mentioned earlier, a TE has six components that are detailed below (Ivy et al., 2017). The general target behaviors were: be respectful, responsible, and safe. However, the target behaviors were individualized to meet the needs of each student in the study (see Appendix C). The tutor reminded the student of the target behavior at the start of each lesson and used a behavior matrix (see Appendix C) to give examples of what each of those behaviors manifest as (e.g., sitting in seat, answering questions, using kind words). This was done with the aid of a PowerPoint slide. Notably for Alexander, examples were not given

because Kiley felt the detailed rule review was frustrating him. At the start of the lesson, the tutor explained that the tokens could be used to purchase rewards to verbally pair the tokens with the back-up reinforcers and to remind students they had the opportunity to earn points during the session. Tokens were delivered on a 3-minute fixed interval (FI) schedule of reinforcement. This interval was chosen to limit interruption of tutoring while still providing frequent reinforcement. Though fixed interval schedules of reinforcement can result in scalloped responding (Dews, 1978), a fixed interval, rather than a variable interval schedule of reinforcement, was chosen for feasibility. Tutors were instructed to deliver tokens only when students were following the tutoring rules at the end of each 3-minute interval when their vibrating timer alerted them. I suggested during tutor training that tutors use behavior-specific praise when delivering the tokens to remind the students why they earned the points; however, praising was not a required part of the TE intervention or recorded on the procedural fidelity form. If a student was not following the rules when the timer vibrated, then the tutor withheld the token and reminded students of the rule they needed to follow to earn a token next time. For example, tutors were taught to say, "Whoops- I cannot give you a point right now. If you want to earn a point next time my timer goes off, you need to be sitting safely in your chair." Students formally exchanged tokens for back-up reinforcers at the end of each session. As stated earlier, students chose from a reinforcer menu that was individualized based on a student preference interview (e.g., Quick, 2014). Some reward examples included getting to show the tutor their family's pet or watching a favorite YouTube video.

***Dependent Variables***

Engagement and disruptive behavior were the two behavioral dependent variables that were graphed and used in visual analysis to determine the effectiveness of the TE intervention.

*Engagement* was the primary independent variable and was defined as any time the student demonstrates engagement in the activity at hand. Engagement could be active, such as responding to the tutor's directions and prompts, sounding out words, or reading aloud. Engagement could also be passive, such as orienting towards the computer screen, nodding, and staying in the seat. For more examples and non-examples of engagement and disengagement, see the coding manual in Appendix D. *Disruptive behavior*, which was a secondary dependent variable, was defined as a behavior that actively disrupts the flow of the lesson. Examples included getting up from the seat without permission, touching toys, and playing with things on the computer that were not part of the lesson. If disruptive behavior was persistent, coders counted it once for each interval or each time it changed in topography.

Because the ultimate goal of increasing engagement during reading tutoring is to improve reading achievement, I monitored reading performance throughout all phases of the study. I also collected oral reading fluency data to determine whether oral reading fluency performance increased during sessions in which the TE was implemented, compared to baseline. Tutors administered a 1-min oral reading fluency curriculum-based measurement (CBM-R) probe (i.e., DIBELS Next ORF; Good & Kaminski, 2011) during sessions, when time permitted. Though collecting various assessment data during tutoring is typical for tutors at the VKC-RC, using CBM-R probes was an added procedure specific to the study. Tutors administered probes that corresponded to the student's current reading level (i.e., all first grade) in number order. Importantly, words read correctly per minute is a metric that predicts future reading performance and is a validated method for screening and identifying students who need more support (Burns et al., 2016; Fuchs et al., 2001; Kim et al., 2010). Though oral reading fluency performance does not measure reading comprehension or reading achievement, it is an indicator of overall reading

competence and predicts reading comprehension performance (Fuchs et al., 2001; Wise et al., 2010). Following tutoring, tutors scored oral reading fluency passages by counting the number of words read correctly in the passage and uploaded their data to Vanderbilt Box. At the conclusion of the study, I graphed the oral reading fluency data to analyze it for trends that emerged through the duration of tutoring and whether there was a functional relation with the implementation of the TE.

### *Tutor Behaviors*

I collected data on tutor behavior to monitor changes in instruction delivery throughout the phases of the study. Tutor behaviors were not classified as dependent variables, because they were not expected to have a functional relation with the TE intervention. Instead, tutor behaviors were observed in the background to detect patterns about how tutors responded when implementing the TE intervention. Coders observed instances of general praise, behavior-specific praise, and refocus/reprimands. Tutor behaviors were continuously recorded by coders using event recording within intervals. General praise was defined as positive feedback to the student that does not specifically mention the activity or behavior at hand, whereas behavior-specific praise went beyond generic praise and explicitly mentioned the student's action that was attempting to be reinforced (e.g., "good job answering my question correctly"). Both types of praise had to be delivered in verbal statements, rather than changes in body language or signaling good job with the tutors' hands. Refocus was defined as instances when the tutor redirected the student towards the activity at hand. Examples of refocus could be gentle, like simply changing the conversation topic or repeating the instructions or instances could be more aggressive, which sounded like a reprimand. For more information about tutor behaviors, including onset and offsets of definitions, see the coding manual in Appendix D.

### *Data Collection System and Interobserver Agreement (IOA)*

Tutors video recorded all sessions on Zoom and uploaded them to Vanderbilt Box for secure storage. Video recording is advantageous because it allows for discrepancy discussions (Ledford & Gast, 2018). Data collectors used the lab-created app, *Behavior Buddy,* to code videos and calculate interobserver agreement (IOA). Coders used momentary time sampling to estimate engagement because it is the most accurate interval-based measurement system for estimating duration of behavior when used with short intervals (Ledford et al., 2015). Specifically, coders marked whether the participant was engaged or disengaged, based on definitions found in the coding manual, at the exact moment the interval was over (i.e., every 15 s). To measure disruptive behavior, I used a frequency count (i.e., event recording) to continuously record behavior. Coders counted every instance of disruptive behavior that occurred during the entire tutoring session and marked which 15s interval during which the disruptive behavior occurred. A portion of randomly selected sessions (across conditions and participants) were double scored to collect IOA on behavioral dependent variable measurement. Coders double scored six sessions (33%) for Melinda, eight sessions (36%) for John, and six sessions (28%) for Alexander. I calculated total agreement within each interval and then averaged within and across observed variables.

Behavioral dependent variable IOA (i.e., engagement, disruptive behavior) and tutor variable IOA (i.e., refocus, praise) is reported in Table 1. Results are reported for each participant by variable. All participants had adequate IOA (average above 85%) for disruptive behavior. Engagement had lower IOA compared to disruptive behavior with an average of 79.6% across participants. John had the lowest engagement IOA (51.1%) during session four, which

**Table 1**

*Interobserver Agreement Results*

| Participant | | Dependent Variables | | Tutor Variables | | |
|---|---|---|---|---|---|---|
| | Overall | Engagement | Disruptive Behavior | General Praise | Behavior-Specific Praise | Refocus |
| Melinda | 91.3 (89.0-94.1) | 87.2 (82.1-94.4) | 87.31 (83.9-95.3) | 88.7 (81.6-93.4) | 98.8 (90.6-100) | 94.6 (90.8-98.7) |
| John | 88.9 (84.6-95.0) | 73.1 (51.1-88.1) | 91.5 (79.5-98.6) | 90.4 (82.0-95.4) | 95.9 (92.5-97.6) | 93.6 (88.8-97.6) |
| Alexander | 89.2 (83.8-92.6) | 78.6 (71.3-90.5) | 87.7 (76.4-91.5) | 89.3 (81.2-95.6) | 97.7 (95.1-100) | 92.6 (89.7-95.4) |
| Average | 89.83 | 79.65 | 88.86 | 89.52 | 97.53 | 93.64 |

*Note.* Average interobserver agreement is reported for each participant by variable. Ranges are reported in parentheses. Overall = average interobserver agreement across variables.

occurred during the initial baseline phase. Tutor variables had higher IOA than behavioral dependent variables, with all tutor variables averaging higher than 85% across participants.

## Experimental Design

A Withdrawal (A-B-A-B) design was used to evaluate the effects of the TE on engagement and disruptive behavior for each tutee participant because it is an appropriate design for evaluating multiple, reversible dependent variables (Ledford & Gast, 2018). This design allows for within participant replication (i.e., direct intra-person replication) and across-participant replication. Although experimental control can be established in a withdrawal design with only one participant, multiple participants were included in the study to improve its external validity (Ledford & Gast, 2018).

### *Study Phases*

Continuous data collection started in baseline. During baseline, the tutor implemented virtual reading instruction procedures only. I did not provide tutors with instructions on any behavior management strategies during baseline. After stable and predictable baseline responding (i.e., at least three sessions with engagement lower than 85% or countertherapeutic trends), the TE was added to reading instruction (intervention). Then, the TE intervention was

withdrawn once there were at least three intervention sessions with high engagement (i.e., greater than 85%) or therapeutic trends. After the return to baseline phase, the TE intervention was reintroduced to complete the withdrawal design. I graphed the data following the coding of each session and used formative visual analysis to dictate when each participant progressed through the phases of the design. After the baseline and intervention phases concluded, I planned to collect maintenance data to see if tutors were still voluntarily using the TE procedures, however time restrictions prevented me from collecting maintenance data.

***Data Analysis***

First, I analyzed the dependent variable data (i.e., engagement, disruptive behavior, oral reading fluency) to determine if the TE intervention was effective for each participant. I used summative visual analysis to examine immediacy, levels, trends, variability, consistency, and overlap in the data to determine whether there were enough demonstrations (i.e., 3 demonstrations) to conclude that there was evidence of a functional relation (Barton et al., 2018). Specifically, I looked to see if the change in the dependent variable was immediate and abrupt when the independent variable was introduced/withdrawn and determined if levels observed in the first baseline condition were retrieved in the second baseline condition (Ledford & Gast, 2018). After assessing whether there was a functional relation between the TE and the dependent variables for each participant, I looked across participants to see if there was evidence of replication of effects across participants.

Then, I analyzed the tutor variables to draw conclusions about patterns observed across the duration of tutoring. I looked for increasing or decreasing trends in the data for each tutor and looked for dramatic shifts in behavior during specific tutoring sessions. These data informed what tutor behaviors were occurring during baseline and intervention.

**Procedural Fidelity**

Data collectors used a checklist (with dichotomous options and areas for tallies of direct systematic observation) to measure procedural fidelity that includes the important components of the reading intervention (control variables) and all six components of the TE procedures (independent variables). See Appendix E for baseline and TE intervention procedural fidelity forms. It is important to collect procedural fidelity data on control variables in addition to the independent variables to ensure changes across conditions occurred for the independent variable only (Ledford & Gast, 2018). I collected procedural fidelity during 100% of sessions. I randomly selected 25% of sessions across all participants and conditions to double score them for IOA. The average IOA on procedural fidelity was 89.7% and ranged from 63.6% to 100%.

Procedural fidelity results can be found in Table 2. Baseline fidelity was 0% for each

**Table 2**

*Procedural Fidelity Results*

| Participant | Study Phase | | | |
|---|---|---|---|---|
| | Baseline | Intervention | Baseline' | Intervention' |
| Melinda | 0 | 85.9 (62.5-100) | 0 | 92.3 (87.5-100) |
| John | 0 | 93.6 (88.2-100) | 0 | 97.6 (92.8-100) |
| Alexander | 0 | 84.0 (33.3-100) | 0 | 88.0 (76.9-100) |
| Average | 0 | 87.8 | 0 | 92.6 |

*Note.* Procedural Fidelity is averaged for each participant within each study phase and then averaged across participant on the bottom line. Ranges are reported in parentheses. Baseline'= return to baseline phase, intervention'= return to intervention phase.

participant, indicating that TE intervention components were absent during baseline sessions. Intervention procedural fidelity was more variable, but acceptable across participants (mean= 89.6%, range 33.3%-100%). Kiley was able to implement the TE intervention with higher

fidelity for John than for Alexander (i.e., mean was 97.6% verses 88.0% during the second intervention phase). Notably, errors in procedural fidelity for all tutors occurred mostly from incorrect token delivery (e.g, failing to deliver a token when the timer went off and the students were following the rules), which is the crux of a TE. Session 16 was particularly bad for Alexander (mean= 33.3%). Kiley abandoned the TE during this session after Alexander left the session crying. She did not reimplement the TE once he returned to the session because they just played games once he returned. With occasional errors and disruption to the TE intervention, these data show that the intervention was delivered with variable fidelity.

**Social Validity**

As social validity was the focus of one of the research questions, I measured it in multiple ways using subjective and objective methods. As a subjective measure, I asked tutors, students, and caretakers to fill out social validity questionnaires that covered topics related to the acceptability of the intervention goals, procedures, and effects (Wolf, 1978). The tutor questionnaire included items that assessed the feasibility of implementing the TE intervention, whereas the student questionnaire included items about whether they enjoyed earning points and rewards (see Appendix F).

To measure social validity objectively, I compared student levels of engagement and disruptive behavior in the TE intervention phases of the study with students in the VKC-RC who were discontinued from the study after baseline because they did not demonstrate a need for behavioral intervention (i.e., normative comparison; Ledford & Gast, 2018). This technique helps to evaluate if the effects of the TE intervention reached a level that is socially acceptable. I also planned to collect data two to six weeks after the end of the study to see if tutors were voluntarily using the TE procedures with other students (i.e., maintenance or sustained use;

Ledford & Gast 2018); however, time constraints prevented me from collecting maintenance data. Lastly, I examined procedural fidelity data to objectively measure how feasible implementing this TE intervention was for the tutors. For example, it was possible that tutors reported on their questionnaire that using the TE was easy and feasible, yet the procedural fidelity data revealed they were missing several implementation components during several sessions, thus revealing a discrepancy in subjective and objective measurement.

# CHAPTER 3

## Results

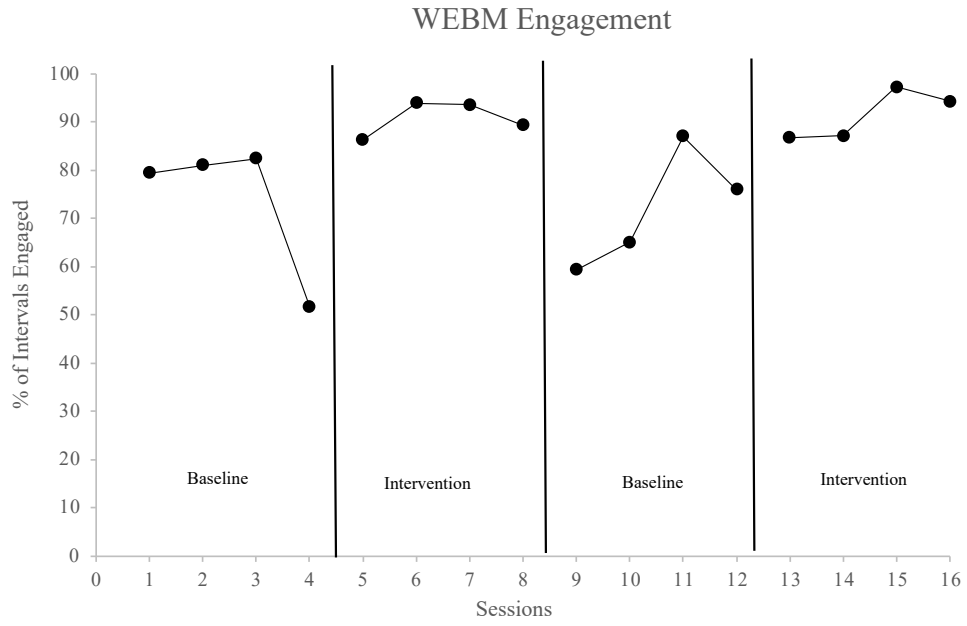
### Intervention Results

#### *Effects on Engagement and Disruptive Behavior*

Figure 1 depicts the TE intervention results on Melinda's engagement during tutoring.

**Figure 1**

*Intervention Effects on Melinda's Engagement*



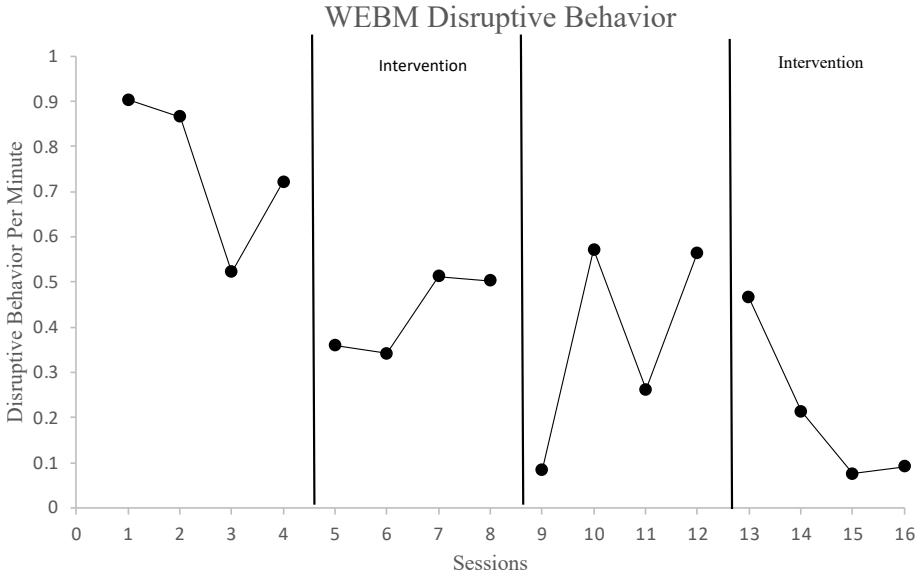

During baseline, Melinda was engaged between 51%-82% of intervals. There was an immediate large decrease in engagement during session four, which signaled the need for behavior intervention. During the first intervention phase, Melinda's engagement was stable and slightly higher than levels of engagement during baseline (intervention engagement ranged from 86%-93%). On session nine, immediately following a return to baseline, there was a large decrease in

engagement. During the return to baseline phase, there was an increasing trend in engagement. This increasing trend continued through to the second phase of intervention. There was very little overlap between baseline and intervention phases, though the level of engagement between phases was not substantial. I concluded there was not a functional relation between the TE intervention and engagement for Melinda because of the increasing trend in engagement in later sessions (including baseline) and the small difference in level of engagement between baseline and intervention sessions. These data do not provide evidence that the TE was effective in increasing Melinda's engagement during reading tutoring; however, on average, Melinda's engagement during intervention sessions was high (mean =84.65).

Figure 2 depicts the TE intervention results on Melinda's disruptive behavior during

**Figure 2**

*Intervention Effects on Melinda's Disruptive Behavior*



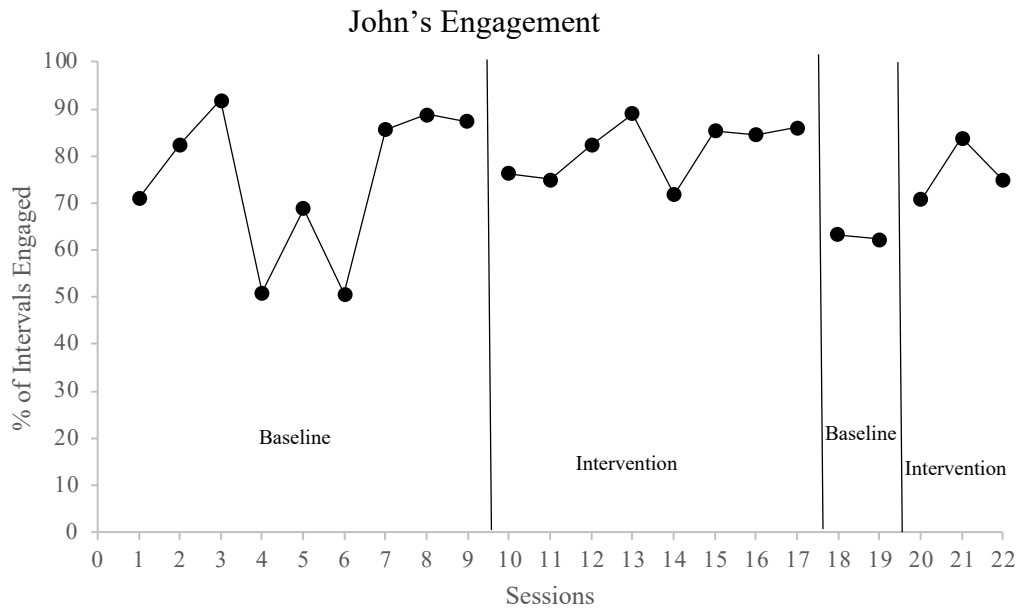

tutoring. During the first two tutoring sessions, Melinda's disruptive behavior was high (almost one instance per minute). There was a large decrease in disruptive behavior during session 3 and

overall, there was a decreasing trend in disruptive behavior during baseline. Once intervention was introduced, disruptive behavior decreased to levels lower than baseline (ranging from 0.3 to 0.5 instances per minute). During the return to baseline phase, disruptive behavior was variable, but did not return to levels observed during the first baseline session. During the last intervention phase, there was a decreasing trend in disruptive behavior with rates reaching as low as 0.07 instances per minute. Due to the decreasing trend in disruptive behavior across all phases, I concluded there was not a functional relation between the TE intervention and disruptive behavior. These data suggest that the TE did not decrease disruptive behavior for Melinda; however, the therapeutic trend of disruptive behavior means that as tutoring progressed, Melinda disrupted tutoring less often.

Figure 3 depicts the TE intervention results on John's engagement. The baseline phase

**Figure 3**

*Intervention Effects on John's Engagement*

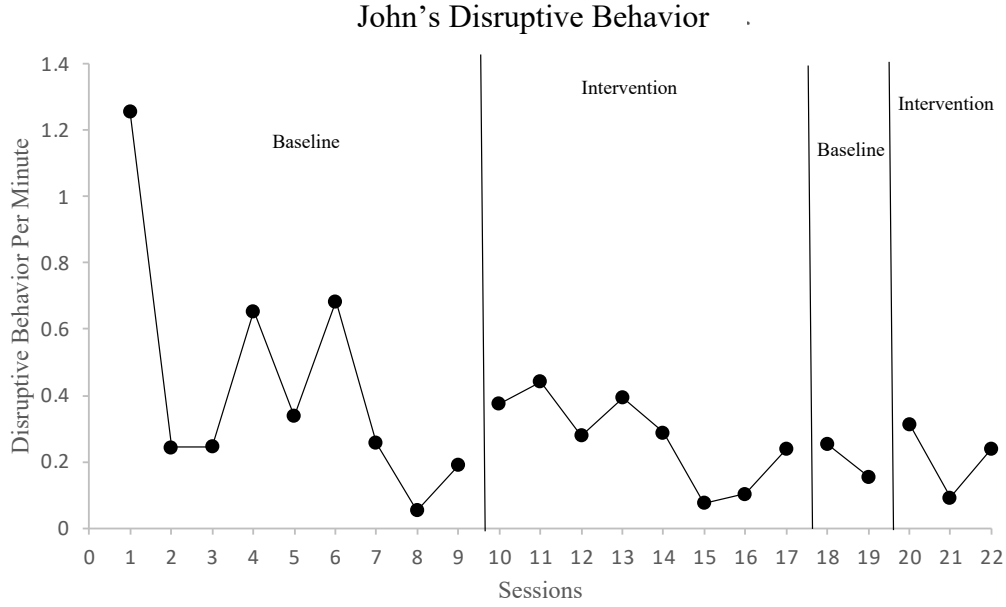


ran for longer than planned due to a fidelity error. Baseline data were very variable and had an increasing trend towards the end of baseline with very high levels of engagement during the last three sessions (mean= 87.25). When intervention began, there was a slight decrease in engagement that then returned to close to the same high levels of engagement found during baseline. I ran intervention for additional sessions, hoping to see higher levels of engagement return. Then, the return to baseline occurred for only two sessions (due to time constraints), but there was a significant decrease in engagement during the return to baseline, with sessions ranging from 62%-63% of engaged intervals. When the TE intervention was reintroduced during the last three sessions of tutoring, engagement increased, but did not reach higher levels of engagement than were found in the initial baseline phase. Due to substantial overlap across conditions and increasing trends in engagement in the first baseline phase, I concluded there was not a functional relation between the TE intervention and engagement. These data suggest that the TE did not improve John's engagement during tutoring.

Figure 4 depicts the TE intervention results on John's disruptive behavior. John's disruptive behavior during baseline was very variable with instances per minute ranging from 0.05 to 1.2. Similar to the way engagement improved prior to intervention, disruptive behavior for John decreased before intervention began. At the start of intervention, there was an uptick in disruptive behavior that decreased across intervention sessions and maintained at low levels. During the return to baseline, there was not a noticeable difference in the level of disruptive behavior compared to intervention. And during the return to intervention, disruptive behavior maintained the same amount of variability and level that was observed in previous phases. I concluded there was not a functional relation between the TE intervention and John's disruptive behavior. While these data do not show that the TE improved John's disruptive behavior, they do

**Figure 4**

*Intervention Effects on John's Disruptive Behavior*

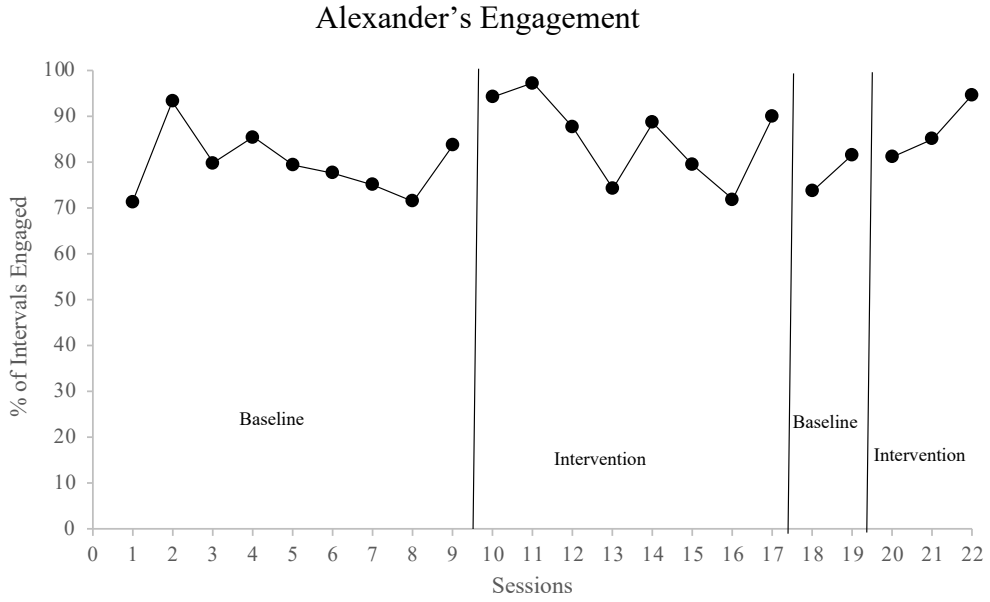



indicate that John's disruptive behavior decreased across the duration of tutoring (therapeutic trend) and remained low and stable.

Figure 5 depicts the TE intervention results on Alexander's engagement. Alexander's engagement during baseline was high and then had a decreasing trend until session 9, where it sharply increased right before intervention. Again, the extended baseline phase was due to a fidelity error by the tutor who was responsible for both John and Alexander's tutoring. When the TE intervention was introduced, there was a slight increase in engagement that became more variable as sessions continued. Levels of engagement during intervention were similar to levels of engagement during baseline. During the return to baseline, engagement decreased slightly, but still overlapped with levels of engagement during intervention. An increasing trend in engagement started during the return to baseline and then continued through the last intervention phase. I concluded there was no functional relation between the TE intervention and Alexander's

**Figure 5**

*Intervention Effects on Alexander's Engagement*

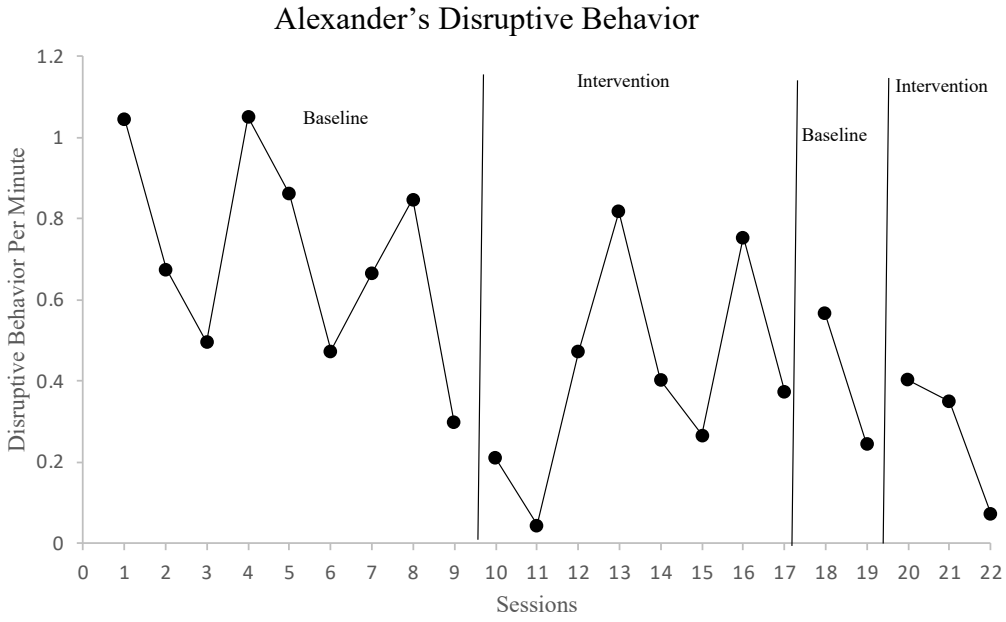



engagement. These data suggest the TE did not improve Alexander's engagement; however engagement was high throughout all phases of the study (above 70% of intervals engaged).

Figure 6 depicts the TE intervention results on Alexander's disruptive behavior. Alexander's disruptive behavior was variable during baseline, ranging from 0.29 to 1.05 instances per minute. When intervention was introduced, there was a slight decrease in disruptive behavior that then returned to similar levels found during baseline. During the return to baseline and when the intervention was reimplemented, levels of disruptive behavior remained close to the same rate and were similar in terms of their large variability. Across conditions, there was significant overlap in the level of disruptive behavior. Though I concluded there was no functional relation between the TE intervention and disruptive behavior, these data suggest that as tutoring progressed, Alexander's disruptive behavior decreased.

**Figure 6**

*Intervention Effects on Alexander's Disruptive Behavior*



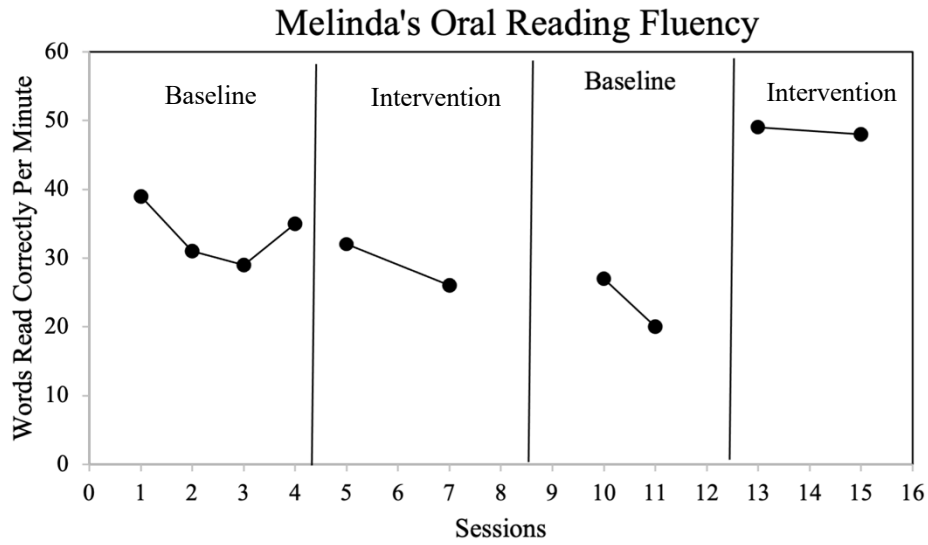

There were no functional relations detected across the three participants that completed the study. This suggests that the TE was not effective in increasing engagement or decreasing disruptive behavior during reading tutoring.

***Effects on Oral Reading Fluency***

Figure 7 depicts Melinda's passage oral reading fluency across tutoring sessions. Oral reading fluency performance appeared to be relatively stable with a slight decreasing trend until session 13, when oral reading fluency performance had a large increase and remained high for two sessions. Notably, Melinda also had high engagement during sessions 13 and 15 (86.7% and 97.2%, respectively). Though oral reading performance is higher in the last intervention phase than it is in baseline sessions, there was not an increase in oral reading fluency performance during the first intervention phase. Therefore, I concluded there was no functional relation

**Figure 7**

*Melinda's Oral Reading Fluency*




*Note.* Data are from 1st grade DIBELS passage oral reading fluency probes.

between the TE and oral reading fluency performance. Typically, first graders read 29 words correct per minute in the winter and improve to 60 words correct per minute by the spring (Hasbrouck & Tindal, 2017). Melinda was in 7th grade when we administered these probes, and we administered 1st grade probes to match her reading level. These data show that she is still reading more slowly than an end-of-year first grader. Her progress, or rate of increase in words read correctly per minute, is also slower than typical peers, as the average rate of weekly improvement is 1.9 words per minute (Hasbrouck & Tindal, 2017).

The tutor initially administered John and Alexander 1st grade passage oral reading fluency probes. The tutor expressed concern that these probes were too hard and causing the participants frustration. In response, I switched both Alexander and John to DIBELS nonsense word fluency probes. The tutor reported that these probes worked better during tutoring. I

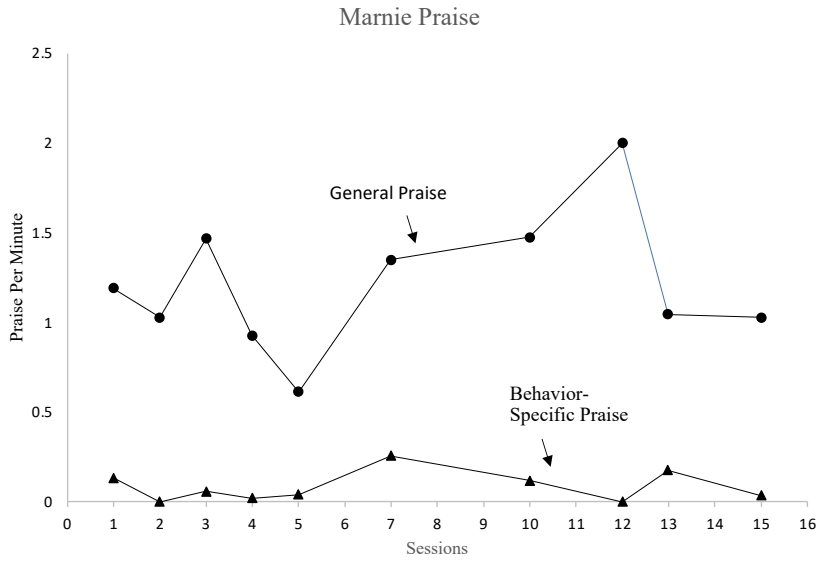
emailed the tutor twice to request she return the data; however, she did not return the completed probes for data analysis.

### *Tutor Variables Results*

Because Melinda had two tutors, general praise, behavior-specific praise, and refocus were graphed and analyzed separately for each tutor. Figure 8 shows Marnie's general praise and

**Figure 8**

*Marnie's Praise While Tutoring Melinda*

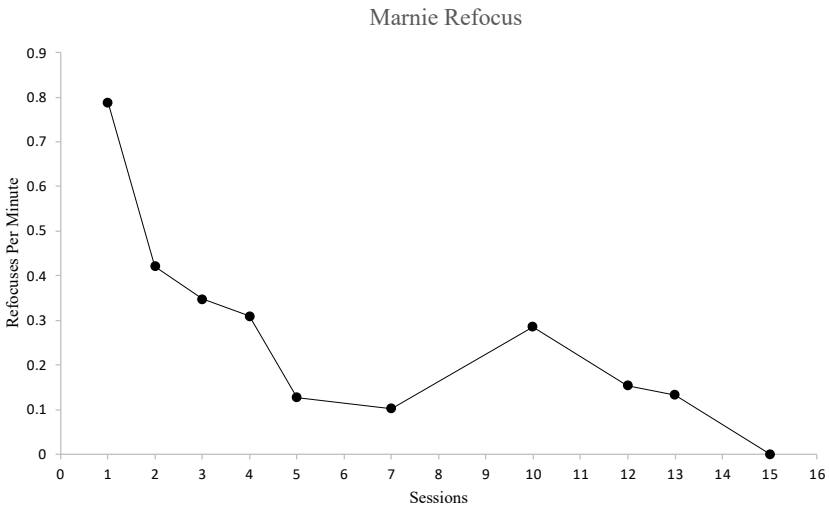



*Note.* Marnie tutored with Evan for sessions 1-4. The data on this graph for sessions 1-4 also contain Evan's praises.

behavior-specific praise graphed together across tutoring sessions. Notably, sessions 1-4 have data for Marine and Evan combined, since they tutored those sessions together. Across all sessions, general praise occurred at a higher rate than behavior-specific praise. General praise was consistent and averaged at 1.2 instances per minute, whereas behavior-specific praise was very low, averaging at 0.08 instances per minute. Figure 9 shows Marine's refocuses per minute across tutoring sessions. Though there appears to be a decreasing trend in refocusing across

**Figure 9**

*Marnie's Refocus While Tutoring Melinda*

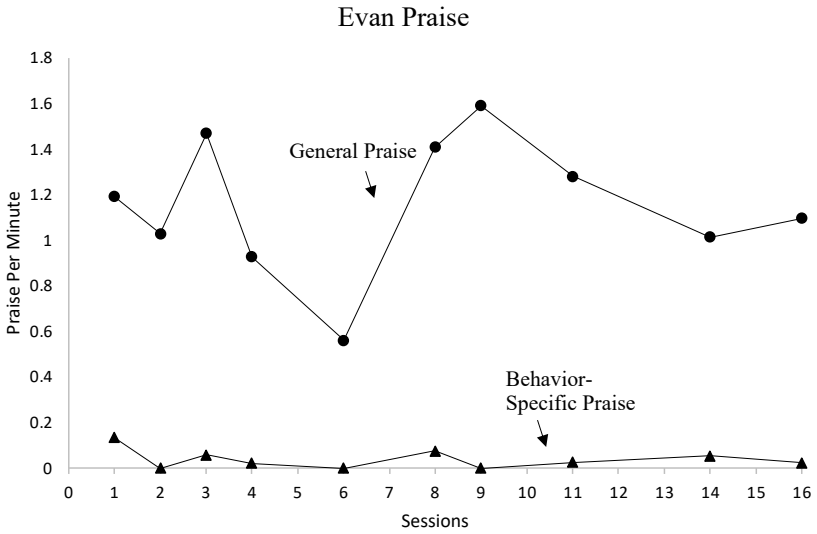



*Note.* Marnie tutored with Evan for sessions 1-4. The data on this graph for sessions 1-4 also contain Evan's refocuses.

sessions, if you ignore sessions 1-4, where refocuses were combined with Evan's refocuses, instances of refocus appear stable and low, averaging at 0.13 instances per minute. Figure 10 depicts Evan's general praise and behavior-specific praise graphed together across tutoring sessions. Like Marnie, Evan's used general praise at higher rates than behavior-specific praise across sessions. Evan's behavior-specific praise was very stable and low throughout the course of tutoring, with three sessions containing no instances of behavior-specific praise. Figure 11 depicts Evan's refocuses per minute across tutoring sessions. There is an overall decreasing trend in Evan's refocusing across tutoring sessions (which is in a therapeutic direction), even when you remove sessions 1-4, which have combined data with Marnie. Most of Evan's tutoring sessions have very few instances of refocusing (mean=0.09).

**Figure 10**

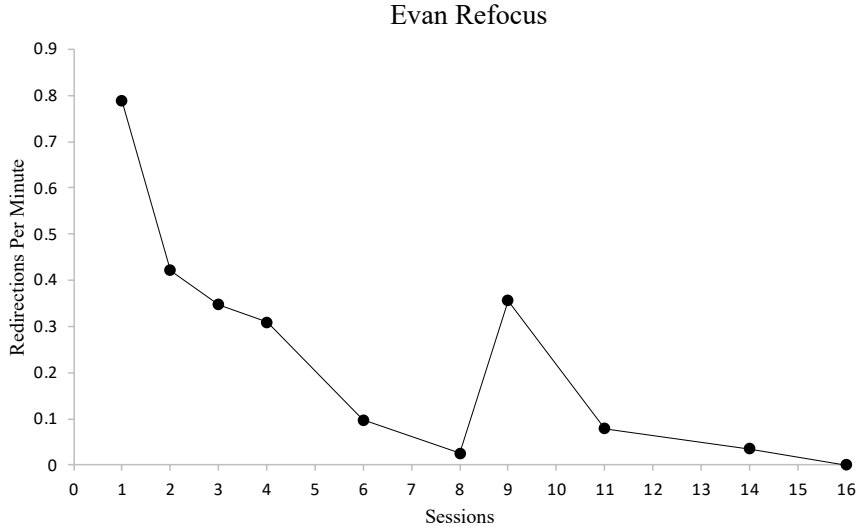
*Evan's Praise While Tutoring Melinda*




*Note.* Evan tutored with Marnie for sessions 1-4. The data on this graph for sessions 1-4 also contain Marnie's refocuses.

**Figure 11**

*Evan's Refocus While Tutoring Melinda*

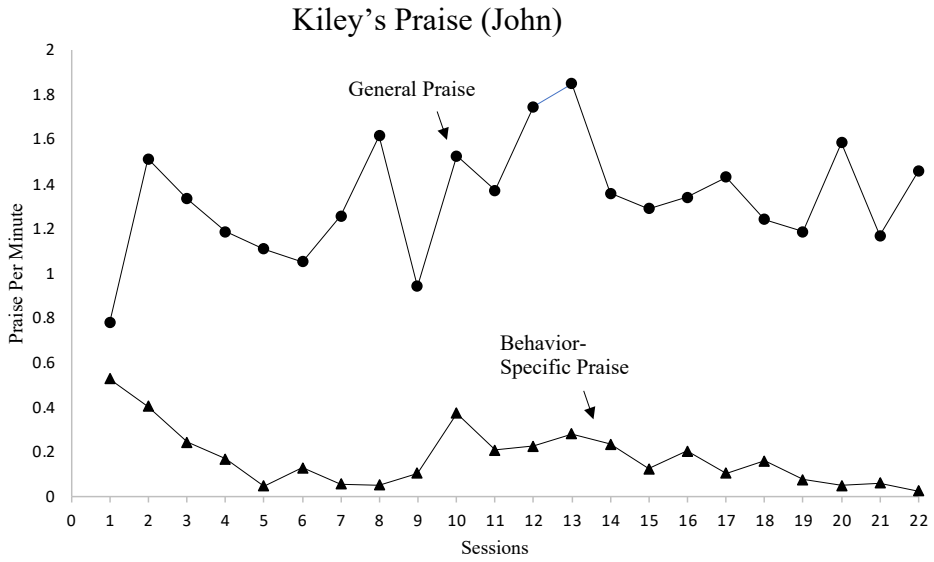



*Note.* Evan tutored with Marnie for sessions 1-4. The data on this graph for sessions 1-4 also contain Marnie's refocuses.

Figure 12 shows Kiley's praise rate while tutoring John. As seen across all tutors, her

**Figure 12**

*Kiley's Praise While Tutoring John*




general praise rate was higher than her behavior-specific praise rate. Rates of behavior-specific praise had a slight increase in level during sessions 10-14, which was when intervention was introduced for the first time. Across time, rates of behavior-specific praise decreased to very low levels, while general praise stayed higher, averaging 1.3 instances per minute. Kiley's refocusing while tutoring John was very variable (see Figure 13). Although certain sessions had high rates of refocusing, rates of refocusing only averaged at 0.29 instances per minute and no session was higher than the average rates of general praise for Kiley.

Kiley's praise during Alexander's sessions is depicted in Figure 14. There is an overall decreasing trend in general praise across sessions, with the last three sessions averaging at a rate of 0.34 instances per minute. Behavior-specific praise remained stable and very low, with 9 sessions having no instances of behavior-specific praise. Kiley's refocusing while tutoring

**Figure 13**

*Kiley's Refocus While Tutoring John*

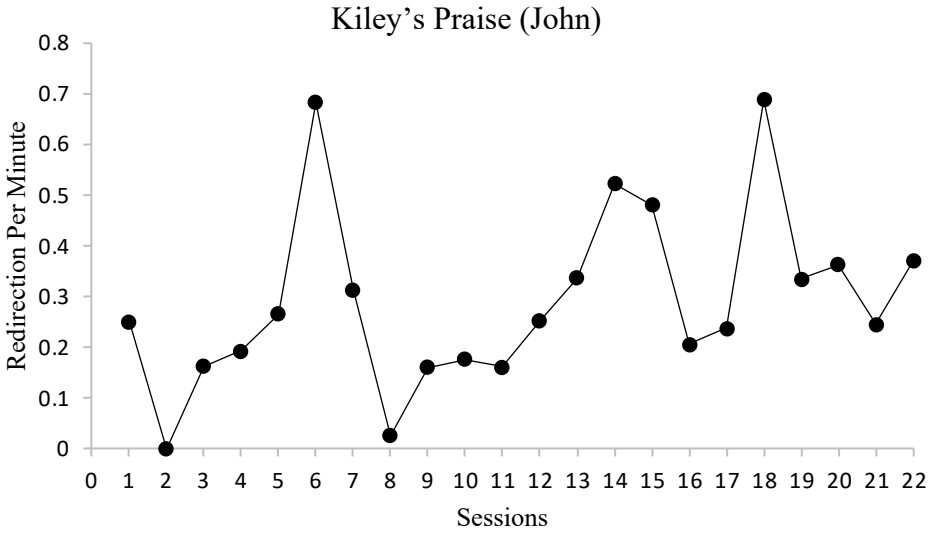
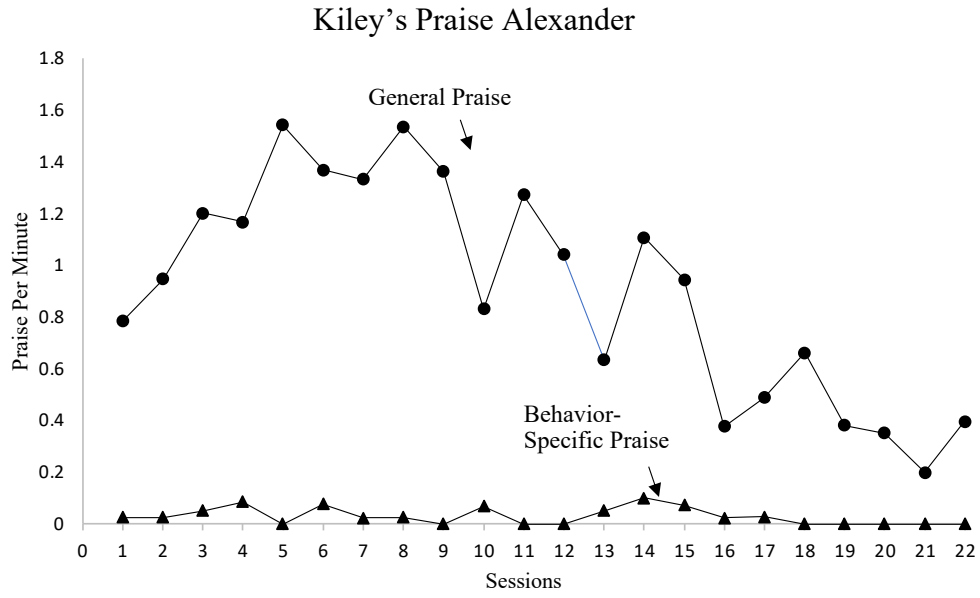



**Figure 14**

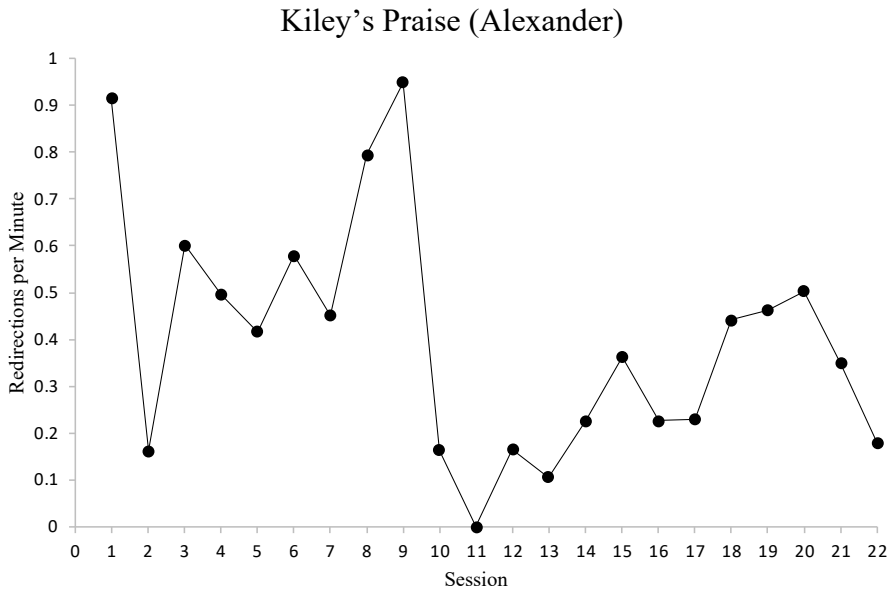
*Kiley's Praise While Tutoring Alexander*




Alexander has a large decrease at session 10, which was the first intervention session (see Figure 15). Rates of refocusing gradually increase during intervention (session 10-17). Then, rates of

**Figure 15**

*Kiley's Refocus While Tutoring Alexander*




refocusing had a moderate increase during the return to baseline (sessions 18-19) and then decreased again during the last two sessions of tutoring, when intervention was in place.

I did not analyze tutor behaviors to infer a functional relation between the TE intervention and tutoring behaviors. Instead, they depict what was being implemented during intervention, like procedural fidelity data. Overall, several trends were revealed across the three tutors. First, behavior-specific praise was not delivered in high-rates, regardless of phase of the study. Second, general praise tended to be variable and high, compared to behavior-specific praise. Last, refocusing tended to be variable but occurred on average less frequently than general praise across all tutors and participants.

## Social Validity Results

### *Objective Social Validity Results*

Because social validity was the focus of research question two, it was evaluated in several ways. First, the procedural fidelity results show that the TE intervention was somewhat feasible for tutors to implement. As summarized in the methods section and in Table 2, intervention fidelity ranged from 33.3-100% and averaged 89.45%, across all intervention sessions and tutors. Procedural fidelity was higher when Kiley implemented the intervention with John than with Alexander, which indicates that the TE intervention may be feasible to implement with some students but not others. Tutors were generally able to implement the TE, however they struggled with certain aspects of the TE intervention, like responding to misbehavior appropriately, and had variable consistency in applying the intervention across sessions.

Another way I evaluated social validity was through peer comparison. There were three participants that started the study and were discontinued because their engagement during baseline was too high to justify intervention. These students had engagement that averaged at 91% of intervals (SD =15). They also had disruptive behavior that averaged at 0.12 instances per minute (SD =0.2). Though I cannot conclude that this intervention was effective for any of the participants who received the TE, due to the lack of functional relations, their levels of engagement and disruptive behavior are comparable during intervention to students who did not have evidence of needing intervention. Of note, the students who were used as the comparison group were referred to participate in the study because either their caregiver or the VKC-RC director thought they might struggle with behavior during tutoring, so the comparison group is

not a population of students with no behavioral concerns, but rather a group that demonstrated high baseline engagement, despite concerns expressed by stakeholders.

### *Subjective Social Validity (Survey) Results*

I was able to collect survey data from two of the three tutors and all the students in the study. I sent a survey to Melinda's caregiver and John's and Alexander's caregiver, though neither caregiver returned the survey. Evan also did not return a survey, though I requested this via email from him twice. Table 3 reports the results from the tutor social validity survey.

**Table 3**

*Tutor Social Validity Survey*

| Question | Scale | Marnie | Kiley |
|---|---|---|---|
| How easy was the token economy to implement? | (1=not at all; 5=very easy) | 4 | 4 |
| If given the opportunity, how likely would you be to continue using the token economy with the same student or a similar student? | (1=not at all; 5=very likely) | 5 | 4 |
| How much did the token economy improve your student's behavior? | (1=not at all; 5= completely) | 5 | 4 |
| I would suggest using a token economy with similar procedures to other teachers who have students with challenging behavior. | (1= disagree, 5=agree) | 5 | 5 |
| The benefits of this intervention outweighed any negative side effects. | (1=disagree; 5=agree) | 5 | 4 |
| I liked implementing the procedures in this intervention. | (1=disagree; 5=agree) | 5 | 4 |

All results were very positive, with scores of either four or five (out of five points). Both tutors reported that they thought the TE intervention was "very effective" and both suggested adjusting the schedule of reinforcement (i.e., how frequently points were delivered). Marnie thought the reinforcement schedule interval should be lengthened to prevent it from interrupting the lesson as often. Kiley thought the interval should be shortened to prevent the student from losing focus.

Table 4 reports the results from the tutee social validity survey. All three participants gave the TE intervention positive scores, with most scores being five. The only question that had

**Table 4**

*Student Social Validity Survey*

| Question | Scale | Melinda | John | Alexander |
|---|---|---|---|---|
| I liked earning points during tutoring. | (1=disagree; 5=agree) | 5 | 5 | 5 |
| I liked the rewards I earned during tutoring. | (1=disagree; 5=agree) | 5 | 5 | 5 |
| Earning points helped me learn and behave better during tutoring. | (1=disagree; 5=agree) | 4 | 3 | 5 |
| I wish my other teachers at school could let me earn points or do something similar to help me. | (1=disagree; 5=agree) | 5 | 5 | 5 |
| I think other kids would like to earn points during tutoring like I did. | (1=disagree; 5=agree) | 5 | 5 | 5 |

variable feedback was whether students thought that earning points helped them learn and behave during tutoring. Students reported that they liked earning points and rewards. Melinda liked the TE intervention because she said "it helped her improve" and she liked "watching YouTube." John said he did not "like the buzzer" from the timer, but he "liked getting prizes," specifically earning the reward of getting to "show the tutor a toy or a pet." Alexander said he, "liked getting prizes," specifically he liked "looking at pictures of dinosaurs."

From the peer comparison data and survey social validity data, there is evidence that this TE intervention is socially valid and feasible. The tutor, who returned the social validity survey, liked delivering the TE intervention and students liked earning points and rewards. The TE intervention was implemented with variable fidelity, which raises questions about feasibility in a virtual format.

# CHAPTER 4

## Discussion

The purpose of the first research question was to determine whether a TE applied in a virtual reading tutoring setting increased engagement and decreased disruptive behavior for struggling readers. I examined data from three participants who completed all phases of the study and concluded there were no functional relations between the TE intervention and the dependent variables. This indicates that the TE intervention was not effective for increasing engagement or decreasing disruptive behavior for struggling readers. Notably, there were high levels of engagement and low levels of disruptive behavior across all six participants who participated in the study (with some participants having more variability across sessions than others). Though it was not the purpose of the study, these data prove that students who are struggling readers can have high engagement and good behavior during virtual reading tutoring, though some exhibit these behaviors less consistently than others. Prior to the COVID-19 pandemic, very few elementary-aged students participated in virtual learning, therefore there are very few studies that quantify and report how children behave when receiving educational instruction in this setting (Bernard et al., 2004). This study provides initial documentation about levels and trends of engagement and disruptive behavior of struggling readers when receiving virtual reading tutoring. Future research should explore interventions that are effective for increasing engagement and decreasing disruptive behavior, for students who need more support in the virtual setting.

The purpose of the second research question was to examine the feasibility of implementing a TE in a virtual tutoring session and to report information about the social validity of the TE intervention. The procedural fidelity data reveal that the TE intervention was

somewhat feasible for tutors to implement. Tutors were able to deliver reading instruction to their students via zoom while embedding a point system, however some tutors implemented the TE intervention procedures more consistently than others, depending on their students. Tutors were logistically able to navigate the demands of the computer, namely making points visible to the students and using PowerPoint slides to explain and review the behavior intervention. Tutors were also able to deliver rewards (i.e., back-up reinforcers) virtually to students that took up less than five minutes of the tutoring session and were free. Through surveys, students and tutors reported that they enjoyed using the TE intervention and liked earning prizes. These findings about the social validity of the TE intervention highlight the potential of using TEs in a virtual setting. Though this study does not demonstrate that this version of a TE is effective for increasing engagement and decreasing disruptive behavior, it establishes that TEs can be implemented virtually for certain students and that a tutor and participants find the goals and procedures to be socially valid. More research needs to be conducted to understand for what types of students and which behaviors create easier feasibility for implementing the TE intervention for tutors.

**Discussion of Oral Reading Fluency**

I collected data on oral reading fluency to monitor students' reading performance as they received virtual reading tutoring and to determine whether there was a functional relation between the TE and oral reading fluency performance. Though the rate of reading progress for Melinda was slow, when compared to typical word learning rates and benchmark data (Hasbrouck & Tindal, 2017), she made slight progress over the course of the study. I am unable to conclude whether the reading progress made during the study was due to the reading tutoring or from other experiences students encountered outside of the study; however, this provides

initial evidence of the potential of virtual reading tutoring in slightly improving students oral reading fluency performance. Similar to the findings of Beach and colleagues (2021), this study also provides an example in the scientific literature of reading curriculum being converted to a virtual format and being delivered virtually, though I did not collect procedural fidelity data on how well the tutors implemented the reading curriculum.

Though I hypothesized that the environment created by the TE could potentially increase oral reading fluency performance relative to baseline, the data did not reveal a functional relation between the TE and oral reading fluency performance. As mentioned earlier, oral reading fluency performance is distinct from reading achievement or learning (Soderstrom & Bjork, 2015) and is particularly susceptible to attentional or behavioral disruptions because the tasks are timed. Though there was no functional relation, the sessions in which Melinda had the highest oral reading fluency performance were also sessions in which Melinda had high engagement. Importantly, there were earlier sessions in which Melinda had high engagement but low oral reading fluency. There are a few reasons that may explain why oral reading fluency did not increase as a function of implementing the TE. First, the TE did not increase engagement or decrease disruptive behavior when it was in place. Because the TE was not effective, the environment in which the tutors administered the CBM-R probes was not improved (or different from baseline), meaning poor engagement and disruptive behavior may have prevented students from focusing on reading. Second, the tutor only conducted CBM-R probes during 10 out of 16 sessions. It is possible that more continuous data collection could have provided more evidence of experimental control. Third, it is possible that reading progress from the reading intervention competed with the effects on reading fluency from the TE intervention. I expected there to be an increasing trend in oral reading fluency performance across tutoring sessions, regardless of

phase, because the students were receiving an evidence-based reading intervention that should gradually improve oral reading fluency performance. I attempted to detect increases in oral reading fluency during intervention sessions in response to the TE that were above and beyond the progress that was supposed to be made from receiving the reading intervention. This change in oral reading fluency performance might be too small to detect in the context of an increasing trend. Last, there is substantial measurement error in CBM-R data (Christ, 2007; Poncy et al., 2005) and this error can pose threats when evaluating differences in performance in terms of level, trend, and phase overlap (Klingbiel et al., 2017). Future research should explore ways in which oral reading fluency performance can be influenced in virtual settings, perhaps with interventions that are more aligned with improving oral reading fluency. For example, researchers could evaluate Repeated Reading in a virtual setting, since it has substantial evidence of increasing oral reading fluency performance, with reported mean gain effect sizes of 0.85 for students with disabilities (Therrien, 2004).

**Discussion of Tutor Behaviors**

I observed and monitored several tutor behaviors throughout the course of the study. These data documented low rates of praise during instruction. In particular, behavior-specific praise occurred at very low rates across tutors and students. It is typical for teachers to use more general praise statements than behavior-specific praise statements. In fact, Reinke and colleagues (2013) reported an average of 25.8 general praise statements per hour and only 7.8 behavior-specific praise statements per hour from 33 general education teachers of early elementary students. Notably, rates of praise to students with behavior disorders are substantially lower than their typical peers (Jenkins et al., 2015). For example, Wehby et al. (1995) found rates of praise to be as low as 0.03 instances per hour for students with aggression. This study indicates that tutor

praising does not increase automatically when tutors are trained to implement a TE. Future research should examine whether tokens alone or tokens delivered with praise change the effectiveness and maintenance of TEs (e.g., Novak & Hammond, 1983).

In the current study, tutor refocus statements occurred on average at a lower rate than general praise across tutors and students. Generally, teachers tend to have higher ratios of emotionally negative-to-positive interactions with students (Beaman & Wheldall, 200), so this finding is important and shows that tutors can have a supportive tutoring environment in the virtual setting. Cook et al., (2017) trained teachers to use a 5:1 praise-to-reprimand ratio in elementary and middle school classrooms and found significantly fewer disruptive problems and higher academic engagement for students in the intervention group. Though that ratio may be difficult to achieve for students with poor reading or behavior performance, increasing tutor praise and decreasing tutor reprimands through intervention has the potential to create a supportive praise-to-reprimand ratio in a virtual setting.

**Potential Reasons for Null Behavioral Results**

There are several reasons why the TE may not have worked to increase engagement and decrease disruptive behavior for struggling readers. First, the virtual environment was difficult to control and could have caused the effects of the TE to be diminished. Each student had a different home environment where they attended virtual tutoring and sessions may have been interrupted by extraneous stimuli. For example, both John and Alexander attended tutoring in the same room on the same computer. Their caretaker tended to leave the room once tutoring began, but background noise from the rest of the house could be heard on some of the recordings. Therefore, changes in student or tutor behavior could have been impacted by background distractions, rather than the TE. Establishing operations, or events that occurred outside of

tutoring that changed the participants responding patterns or altered the TE's reinforcement effectiveness, may have also influenced the effects of the TE (Michael, 1993). For example, Melinda participated in a behavior program at school with various contingencies. At the start of each session, the tutor and the student often discussed how the school day was and whether Melinda earned her rewards at school. The outcome of the behavior program at school could have influenced Melinda's behavior during tutoring sessions and altered the power of the TE's reinforcement. Future researchers of interventions in virtual settings should work to build consistency in the virtual environment. Participants should attend virtual tutoring from the same location each session and on the same electronic device. Researchers should take data on the virtual environment and document any changes that could impede the internal validity of the study.

Second, instances of low procedural fidelity could have prevented the TE intervention from being effective and caused null effects in the study (Barton et al., 2018), especially if tutors missed key components of the TE intervention, like delivering the tokens consistently (Ivy et al., 2017). I was only able to conduct one 45-min tutor training session, which may explain the instances of low procedural fidelity. Behavior-skills training, which I was unable to conduct, has evidence of being delivered effectively in remote settings (e.g., Shriver, 2022); therefore, future research on TEs in virtual settings should expand training to include practice opportunities and feedback (Horner & Sturmey, 2012). It is possible that other factors explained low procedural fidelity aside from limited tutor training. Prior research has shown that teacher self-efficacy beliefs and motivation predict curriculum fidelity (Aytac, 2021). Kabas and Yildiz (2020) found self-efficacy beliefs, or confidence to improve student performance, was positively associated with fidelity in implementing a Turkish literacy curriculum. Perhaps there is a third variable

(e.g., tutor self-efficacy) that explains which tutors had low verses high procedural fidelity, and in turn, could moderate the effectiveness of TEs.

Third, participant baseline behavior may have not allowed for enough improvement in engagement and disruptive behavior. Participants with average baseline engagement above 80% were discontinued from the study, however the participants who received the TE intervention only had slightly lower engagement in baseline. It is unclear from the literature what level of engagement should be expected and is acceptable for students during tutoring sessions. Scott et al. (2011) found that general education high school students were only actively engaged in the curriculum 39% of the time and passively engaged in the curriculum 42% of the time, which means overall engagement averaged at 81%. In the same study, disruptions occurred once every 16.67 minutes. For elementary students, average level of engagement for typically developing students is likely lower (Berlinger, 1978), with average estimates of 71-73% engagement in 2nd grade general education classrooms during reading instruction (Rosenshine, 2015). In the current study, I attempted to detect very small behavior changes that would have very slight increases in level and minimal changes in trend and variability. Future researchers could replicate this study with participants who have more extreme behavior problems or adjust the codebook to detect more nuanced changes in behavior. Specifically, researchers could recruit participants who had documented EBD diagnoses to increase the likelihood that disruptive behavior would be high in the virtual setting. Or the codebook could be adjusted to measure active and passive engagement separately (e.g., Wills et al., 2018) or by measuring another variable like active *participation* (e.g., Didion et al., 2020), that could monitor students' responses to tutor requests to read text or write words.

Last, it is possible that the theory of change was wrong for this study. I theorized that the comorbidity between reading and behavior was dependent on the environment and interruptions to instruction. If the relation between poor reading performance and behavior problems was explained by correlated liabilities, then that could explain the null results of the study. For example, it is possible that attention was a shared deficit for the student in these studies, particularly for John who had an ADHD diagnosis. For students with ADHD, oftentimes interventions are most effective when they are combined with medication use (Prasad et al., 2013; So et al., 2008). Future research should examine ways to address deficits that may be genetic (Willcutt & Pennington, 2000), and develop interventions that target behavioral and reading goals specifically for students with ADHD.

As mentioned earlier, TEs are based on theories of operant conditioning and use reinforcement to alter behavior. The TE intervention does not address potential skill/ability deficits and therefore may not have been effective in increasing engagement and decreasing disruptive behavior. It is possible that language ability is a shared deficit for the students who participated in the study, especially for John and Alexander who had speech delays. Again, this is especially plausible because language deficits are well documented in the literature for students with externalizing behavior (Benner et al,, 2009; Chow & Wehby, 2018) and poor reading performance (Catts & Kamhi, 2005; Lombardino et al., 1997). Future research could explore adding components to the TE intervention that support language development (e.g., Curtis et al., 2017) or focus on other deficits that may share correlated liabilities, like executive function (Sesma et al., 2009). Bruhn and colleagues (2022) found significant effects for increasing engagement and decreasing disruptive behavior during academic instruction when students used a self-monitoring app. Interventions that use principles of applied behavior

analysis while addressing skill deficits are promising because they can simultaneously address the causal model and correlated liability model of comorbidity.

**Possible Behavioral Intervention Adaptations**

Because so many students struggle with both reading and behavior (Lane et al., 2008; Nelson et al., 2004) and because virtual instruction is likely to persist (Pitts et al., 2022), it is imperative to consider ways that the TE intervention could be adapted to meet the behavioral needs of struggling readers who are receiving virtual reading tutoring. The following recommendations could be used to potentially strengthen the effects of the TE to better interrupt the cycle of negative reinforcement, which aligns with the causal model of comorbidity.

There are several options for ways that TE can be adjusted and individualized (Ivy, et al., 2017). First, the schedule of reinforcement can be easily adjusted to meet the needs of learners and tutors. Tutors suggested a change to the token delivery interval on their social validity surveys. The interval could be shortened to increase the frequency of reinforcement for participants. On the other hand, the interval could be lengthened to increase feasibility of delivering tokens and perhaps increase procedural fidelity. It is also possible to change the schedule of reinforcement by adjusting the response requirement. The TE in the current study was implemented on a fixed-ratio scale of 3 min. Instead of using a time-based (ratio scale) schedule of reinforcement, tutors could use a response-based schedule of reinforcement (Ivy et al., 2017). For example, the Road to Reading curriculum has five distinct steps during each lesson. Tutors could award tokens after students successfully completed each step of the program. Notably, this adjustment is more likely to impact task completion rather than engagement and disruptive behavior (because of the shift in contingency for earning tokens), though task completion could have transfer effects to other student behaviors.

Second, the TE intervention could be adapted to include a formal preference assessment to determine appropriate back-up reinforcers. Using formal preference assessments is uncommon in the TE literature, but could be useful (Kim et al., 2022). Instead of simply asking the participants which reward they would like to receive, tutors could conduct a virtual preference assessment by observing how much time children engaged with certain back-up reinforcers when given free time (i.e., Free Operant Observation) (Chazin & Ledford, 2016). A forced hierarchy could be created by ranking back-up reinforcers as most preferable if the student spent the most time engaging with that reward. Another option for objectively measuring student preference would be to collect data on what students chose to purchase with their points after each session (i.e., Multiple Stimulus with Replacement Preference Assessment) (Chazin & Ledford, 2016). In this model, all rewards would initially have the same point value. After time, tutors could adjust the cost of rewards based on how often a child chose them in previous sessions. This method may be preferable because it does not require losing an instructional session to free play. Third, the back-up reinforcer could be a "mystery reward" instead of having students select from a menu of rewards (e.g., Tan et al., 2022). Mystery rewards are most often used in general education settings, but they have the potential to be effective in special education settings (Kim et al., 2022) and should be experimentally evaluated in virtual settings.

In contrast to the ideas described to adapt the dimensions of the TE, it may be advantageous to change the TE intervention to focus on the tutor rather than the participant. If the intervention could be modified to improve tutor behaviors, then researchers could look for a transfer effect to student behavior. As mentioned earlier, praise is most effective when it is behavior-specific (Brophy, 1981; Jenkins et al., 2015) and has substantial evidence of increasing student engagement and decreasing disruptive behavior for school-aged children (Royer et al., 2019).

Furthermore, Sutherland and colleagues (2000) demonstrated that teacher's rates of behavior-specific praise can be increased through observation and feedback. Increasing behavior-specific praise also has preliminary evidence of working in a virtual format, through email feedback (Gage et al., 2018). Instead of the student receiving tokens, the tutor could receive tokens for instances of behavior-specific praise and changes in both tutor and student behaviors could be evaluated. Or the TE procedures could be abandoned completely and instead, an intervention that uses performance feedback could be developed to alter tutor behaviors in a virtual setting (e.g., Sutherland et al., 2000).

**Limitations**

There are several limitations of the current study- some that threaten the study's internal validity and others that may limit the study's external validity and generalizability. First, IOA for certain sessions was lower than 80%. IOA for John was particularly low, which limits the internal validity of the study. Fortunately, all sessions were video recorded, so the data can be recoded and reviewed to resolve discrepancies in dependent variable observation.

As mentioned earlier, practical limitations in the context of the applied research setting threatened the experimental rigor of the study. Schedule constraints dictated the tutor-participant pairings. Melinda was assigned to two tutors that shared initial sessions and then rotated sessions. This created a lack of consistency among control variables in the study and threatens its internal validity. Similarly, John and Alexander are siblings and were assigned to the same tutor, for practical purposes. Consequently, phase changes were synchronized for John and Alexander to limit confusion for the tutor about whether the students were in baseline or intervention. Phase change decisions were informed by student responding and conformed to the guidelines described in the methods; however, decisions were also weighed against having a finite number

of tutoring sessions. The balance of weighing student data and also moving through phases to complete the entire withdrawal design threatens the experimental control of the study. Having the same tutor paired with two participants also reduced the amount of social validity survey data I was able to collect.

Third, there were instances of poor procedural fidelity during intervention sessions, especially for Alexander. This threatens the study's internal validity because it makes environmental differences between baseline and intervention sessions less distinct. Errors in procedural fidelity also caused interruption to phase changes for John and Alexander, which caused there to be fewer data points in later phases and prevented maintenance data from being collected.

Fourth, the study has missing data. Neither Melinda's or John/Alexander's caregiver returned social validity surveys, so I am unable to conclude what the children's caregivers thought about the goals, procedures, and effects of the intervention. Kiley did not return the oral reading fluency data she collected on John or Alexander. The oral reading fluency data for Melinda is also limited; probes were only conducted for 10 out of 16 sessions. Though this does not impact the internal validity of the study, it limits my ability to draw conclusions about reading performance during the study, which was an important variable to monitor.

Fifth, the reading curricula differed between students, with Melinda receiving Road to Reading and John and Alexander receiving Friends on the Block. Had I been able to conclude that there was a functional relation between the TE intervention and the dependent variables, then having two different curricula would have impacted the strength of the inter-person replication. However, had the TE been effective, regardless of curricula, then that could have expanded the generalizability of the findings.

Sixth, the normative comparison data used to assess the effects of the social validity of the TE intervention do not come from a "typical" sample of students without behavior problems. I used the baseline data of the three students who were initially identified by their caregivers or the VKC-RC director as potentially needing behavioral support during tutoring. Therefore, though the baseline engagement data of the comparison group were very high, it is possible that students who did not get referred at all to the study would have even higher baseline engagement during tutoring.

Last, the tutors had different levels of teacher training and experience. Notably, none of the tutors had applied behavior analysis training. It is possible that procedural fidelity and TE intervention effects could have been stronger if the tutors were practicing teachers, rather than pre-service teachers or if the tutors had applied behavior analytic training. Similar to the conclusions about the differences in curricula, had the TE intervention been effective for all participants, the diversity or lack of tutor experience could have expanded the generalizability of the findings.

**Conclusion**

In summary, I ran a proof-of-concept study to evaluate a TE in a virtual setting for students receiving one-on-one reading tutoring. The TE did not prove to be effective in increasing student engagement or decreasing disruptive student behavior; however, there is some evidence to support the TE intervention's social validity. Reading performance had a small increase for one participant that had complete data, but reading performance did not increase as a function of implementing the TE. Tutors displayed low rates of behavior-specific praise across students. Future research should explore ways in which the TE can be adapted to better meet the needs of students who struggle with both reading and behavior in a virtual setting.

## REFERENCES


Arnold, E. M., Goldston, D. B., Walsh, A. K., Reboussin, B. A., Daniel, S. S., Hickman, E., & Wood, F. B. (2005). Severity of emotional and behavioral problems among poor and typical readers. *Journal of abnormal child psychology*, *33*(2), 205-217.

Aaron, P. G., Joshi, R. M., Palmer, H., Smith, N., & Kirby, E. (2002). Separating genuine cases of reading disability from reading deficits caused by predominantly inattentive ADHD behavior. *Journal of Learning Disabilities*, *35*(5), 425–436. https://doi.org/10.1177/00222194020350050301

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special Education*, *23*(5), 300–316. https://doi.org/10.1177/07419325020230050501

Allor, J. H., Cheatham, J. P., & Al Otaiba, S. (2018). *Friends on the Block.*

Allor, J. H., Yovanoff, P., Al Otaiba, S., Ortiz, M. B., & Conner, C. (2020). Evidence for a literacy intervention for students with intellectual and developmental disabilities. *Education and Training in Autism and Developmental Disabilities*, *55*(3), 290-302.

Aytaç, A. (2021). A study of teachers' self-efficacy beliefs, motivation to teach, and curriculum fidelity: A path analysis model. *International Journal of Contemporary Educational Research*, *8*(4), 130-143.

Aziz, N. A. A., & Yasin, M. H. M. (2018). Token economy to improve concentration among students with learning disabilities in primary school. *Journal of ICSAR*, *2*(1), 32–36.

Baker, J. A., Grant, S., & Morlock, L. (2008). The teacher-student relationship as a developmental context for children with internalizing or externalizing behavior problems. *School Psychology Quarterly*, *23*(1), 3–15. https://doi.org/10.1037/1045-3830.23.1.3

Barton, E. E., Lloyd, B. P., Spriggs, A. D., & Gast, D. L. (2018). Visual analysis of graphic data. In *Single Case Research Methodology* (pp. 179–214). Routledge.

Barton, E. E., Meadan-Kaplansky, H., & Ledford, J. R. (2018). Independent variables, fidelity, and social validity. In *Single Case Research Methodology* (pp. 133-156). Routledge.

Beach, K. D.,Washburn, E. K., Gesel, S. A., & Williams, P. (2021). Pivoting an elementary summer reading intervention to a virtual context in response to COVID-19: An examination of program transformation and outcomes. *Journal of Education for Students Placed at Risk (JESPAR)*, *26*(2), 112-134.

Beaman, R., & Wheldall, K. (2000). Teachers' use of approval and disapproval in the classroom. *Educational Psychology*, *20*, 431–446.

Berliner, D. C. (1978). *Allocated Time, Engaged Time and Academic Learning Time in Elementary School Mathematics Instruction.* [Paper presentation]. National Council on Teaching Mathematics, San Diego, CA, United States.

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., Wallet, P. A., Fiset, M., & Huang, B. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research*, *74*(3), 379–439.

Benner, G. J., Mattison, R. E., Nelson, J. R., & Ralston, N. C. (2009). Types of language disorders in students classified as ED: Prevalence and association with learning disabilities and psychopathology. *Education and Treatment of Children*, 631-653.

Breyer, N. L., & Allen, G. J. (1975). Effects of implementing a token economy on teacher attending behavior. *Journal of Applied Behavior Analysis*, *8*(4), 373–380. https://doi.org/10.1901/jaba.1975.8-373

Bruhn, A., Wehby, J., Hoffman, L., Estrapala, S., Rila, A., Hancock, E., ... & Copeland, B. (2022). A randomized control trial on the effects of MoBeGo, a self-monitoring app for challenging behavior. *Behavioral Disorders*, 01987429221093987.

Burns, M. K., Silberglitt, B., Christ, T. J., Gibbons, K. A., & Coolong-Chaffin, M. (2016). Using oral reading fluency to evaluate response to intervention and to identify students not making sufficient progress. In *The fluency construct* (pp. 123-140). Springer, New York, NY.

Cavanaugh, C., Gillan, K. J., Kromrey, J., Hess, M., & Blomeyer, R. (2004). The effects of distance education on K-12 student outcomes: A meta-analysis. *Learning Point Associates/North Central Regional Educational Laboratory (NCREL)*.

Cavanaugh, C. S., Barbour, M. K., & Clark, T. (2009). Research and practice in K-12 online learning: A review of open access literature. *The International Review of Research in Open and Distributed Learning*, *10*(1).

Catts, H. W., & Kamhi, A. G. (Eds.). (2005). *The connections between language and reading disabilities*. Psychology Press.

Chazin, K.T. & Ledford, J.R. (2016). Free operant observation. In *Evidence-based instructional practices for young children with autism and other disabilities.* Retrieved from http://ebip.vkcsites.org/free-operant

Chen, N.-S., Ko, H.-C., Kinshuk*, & Lin, T. (2005). A model for synchronous learning using the Internet. *Innovations in Education and Teaching International*, *42*(2), 181–194.

Chow, J. C., & Wehby, J. H. (2018). Associations between language and problem behavior: A systematic review and correlational meta-analysis. *Educational Psychology Review*, *30*(1), 61-82.

Christ T. J., Silberglitt B. (2007). Estimates of the standard error of measurement for curriculum-based measures of oral reading fluency. *School Psychology Review*, 36, 130-146.

Christle, C. A., & Yell, M. L. (2008). Preventing youth incarceration through reading remediation: Issues and solutions. *Reading & Writing Quarterly*, *24*(2), 148–176. https://doi.org/10.1080/1057356070180843

Cook, C. R., Dart, E., Collins, T., Restori, A., Daikos, C., & Delport, J. (2012). Preliminary study of the confined, collateral, and combined effects of reading and behavioral interventions: Evidence for a transactional relationship. *Behavioral Disorders*, *38*(1), 38-56.

Cook, C., Grady, E., Long, A., Renshaw, T., Codding, R., Fiat, A., & Larson, M. (2017). Evaluating the impact of increasing general education teachers' ratio of positive-to-negative interactions on students' classroom behavior. *Journal of Positive Behavior Interventions*, *19*(2), 67–77.

Coy, K. A. (2013). *Online instruction with Universal Design for Learning in the synchronous K–8 classroom* (Publication No. 3598057) [Doctoral dissertation, Washington State University]. ProQuest Dissertations Publishing. http://search.proquest.com/docview/1458632526/abstract/C041ED0F89544E3BPQ/1

Curtis, P. R., Kaiser, A. P., Estabrook, R., & Roberts, M. Y. (2019). The longitudinal effects of early language intervention on children's problem behaviors. *Child development*, *90*(2), 576-592.

Daucourt, M. C., Erbeli, F., Little, C. W., Haughbrook, R., & Hart, S. A. (2020). A meta-analytical review of the genetic and environmental correlations between reading and attention-deficit/hyperactivity disorder symptoms and reading and math. *Scientific Studies of Reading*, *24*(1), 23-56.

Dews, P. B. (1978). Studies on responding under fixed-interval schedules of reinforcement: II. The scalloped pattern of the cumulative record. *Journal of the Experimental Analysis of Behavior*, *29*(1), 67-75.

Didion, L. A., Toste, J. R., & Wehby, J. H. (2020). Response cards to increase engagement and active participation of middle school students with EBD. *Remedial and Special Education*, *41*(2), 111-123.

Dietz, S., & Montague, M. (2006). Attention deficit hyperactivity disorder comorbid with emotional and behavioral disorders and learning disabilities in adolescents. *Exceptionality*, *14*(1), 19–33. https://doi.org/10.1207/s15327035ex1401_3

Donolato, E., Cardillo, R., Mammarella, I. C., & Melby-Lervåg, M. (2022). Research Review: Language and specific learning disorders in children and their co-occurrence with internalizing and externalizing problems: A systematic review and meta-analysis. *Journal of Child Psychology and Psychiatry*, *63*(5), 507-518.

Echols, L. D., West, R. F., Stanovich, K. E., & Zehr, K. S. (1996). Using children's literacy activities to predict growth in verbal cognitive skills: A longitudinal investigation. *Journal of educational psychology*, *88*(2), 296.

Engelmann, S. (1999). The benefits of Direct Instruction: Affirmative action for at-risk students. *Educational Leadership*, *57*(1), 77-79.

Foorman, B. R., Herrera, S., Petscher, Y., Mitchell, A., & Truckenmiller, A. (2015). The structure of oral language and reading and their relation to comprehension in Kindergarten through Grade 2. *Reading and writing*, *28*(5), 655-681.

Francis, D. A., Caruana, N., Hudson, J. L., & McArthur, G. M. (2019). The association between poor reading and internalising problems: A systematic review and meta-analysis. *Clinical Psychology Review*, *67*, 45-60.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. In *The Role of Fluency in Reading Competence, Assessment, and instruction* (pp. 239-256). Routledge.

Fuchs, D., Fuchs, L. S., & Vaughn, S. (2014). What is intensive instruction and why is it important? *Teaching Exceptional Children*, *46*(4), 13–18.

Gable, R. A., & Shores, R. E. (1980). Comparison of procedures for promoting reading proficiency of two children with behavioral and learning disorders. *Behavioral Disorders*, *5*(2), 102–107.

Gage, N. A., Grasley-Boy, N. M., & MacSuga-Gage, A. S. (2018). Professional development to increase teacher behavior-specific praise: A single-case design replication. *Psychology in the Schools*, *55*(3), 264-277.

Glassberg, L. A., Hooper, S. R., & Mattison, R. E. (1999). Prevalence of learning disabilities at enrollment in special education students with behavioral disorders. *Behavioral Disorders*, *25*(1), 9–21.

Good, R. H., & Kaminski, R. A. (2011). *DIBELS Next assessment manual*. Eugene, OR: Dynamic Measurement Group. Available from https://dibels.org/

Gunter, P. L., Denny, R. K., Jack, S. L., Shores, R. E., & Nelson, C. M. (1993). Aversive stimuli in academic interactions between students with serious emotional disturbance and their teachers. *Behavioral Disorders*, *18*(4), 265-274.

Gunter, P. L., Kenton Denny, R., Shores, R. E., Reed, T. M., Jack, S. L., & Nelson, M. (1994). Teacher escape, avoidance, and counter control behaviors: Potential responses to disruptive and aggressive behaviors of students with severe behavior disorders. *Journal of Child and Family Studies*, *3*(2), 211-223.

Hastie, M., Chen, N.-S., & Kuo, Y.-H. (2007). Instructional design for best practice in the synchronous cyber classroom. *Educational Technology & Society, 10*(4), 281-294.

Hasbrouck, J., & Tindal, G. (2017). An update to compiled ORF norms (Technical report no. 1702). Behavioral Research and Teaching, University of Oregon.

Herrnstein, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the experimental analysis of behavior*, *4*(3), 267.

Hinshaw, S. P. (1992). Academic underachievement, attention deficits, and aggression: comorbidity and implications for intervention. *Journal of consulting and clinical psychology*, *60*(6), 893.

Horbach, J., Mayer, A., Scharke, W., Heim, S., & Günther, T. (2020). Development of behavior problems in children with and without specific learning disorders in reading and spelling from kindergarten to fifth grade. *Scientific studies of reading*, *24*(1), 57-71.

Ivy, J. W., Meindl, J. N., Overley, E., & Robson, K. M. (2017). Token Economy: A systematic review of procedural descriptions. *Behavior Modification*, *41*(5), 708–737.

Jenkins, L. N., Floress, M. T., & Reinke, W. (2015). Rates and types of teacher praise: A review and future directions. *Psychology in the Schools*, *52*(5), 463-476.

Kabaş, İ., & Yıldız, S. (2020). Primary school teachers' perceptions of self-efficacy on first literacy teaching and their commitment to the Turkish lesson first literacy teaching curriculum. Researcher: Social Science Studies, 8(1), 105-134. https://doi.org/10.29228/rssstudies.41826

Kara, M., & Yildirim, Z. (2020). Identification of the optimal faculty behaviors for performance improvement in distance education. *Asia Pacific Education Review*, *21*(1), 83-97.

Kim, J. Y., Fienup, D. M., Oh, A. E., & Wang, Y. (2022). Systematic review and meta-analysis of token economy practices in K-5 educational settings, 2000 to 2019. *Behavior Modification*, *46*(6), 1460-1487.https://doi.org/10.1016/j.appdev.2008.12.025

Kim, Y. S., Petscher, Y., Schatschneider, C., & Foorman, B. (2010). Does growth rate in oral reading fluency matter in predicting reading comprehension achievement?. *Journal of Educational Psychology*, *102*(3), 652.

Klingbeil D. A., Van Norman E. R., Nelson P. M. (2017). Precision of curriculum-based measurement reading data: Considerations for multiple baseline designs. *Journal of Behavioral Education*, 26, 433-451.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., et al. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013) Single-case intervention research design standards. *Remedial and Special Education, 34*, 26-38.

Kuhfeld, M. & Tarasawa, B. (2020). *The COVID-19 slide: What summer learning loss can tell us about the potential impact of school closures on student academic achievement*. NWEA.

Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual review of clinical psychology*, *2*, 111.

Lamb, S., Maire, Q., Doecke, E., Macklin, S., Noble, K & Pilcher, S. (2020) *Impact of learning from home on educational outcomes for disadvantaged children.* Centre for International Research on Education Systems and the Mitchell Institute, Victoria University.

Lane, K. L., Barton-Arwood, S. M., Nelson, J. R., & Wehby, J. (2008). Academic performance of students with emotional and behavioral disorders served in a self-contained setting. *Journal of Behavioral Education*, *17*(1), 43-62.

Ledford, J. R., Ayres, K. A., Lane, J. D., & Lam, M. F. (2015). Accuracy of interval-based measurement systems in single case research. *Journal of Special Education*, *49*, 104–117.

Ledford, J. R., & Gast, D. L. (2018). *Single case research methodology: Applications in special education and behavioral sciences (3rd Ed)*. Routledge.

LeJeune, L. M., Samudre, M. D., Whitaker, S. C., & Bailey, C. (2022). The effect of CW-FIT on group engagement during virtual instruction. *Journal of Special Education Technology*, 01626434211065436.

Lemons, C. J., Kearns, D. M., & Davidson, K. A. (2014). Data-based individualization in reading: Intensifying interventions for students with significant reading disabilities. *Teaching Exceptional Children*, *46*(4), 20–29.

Leslie, L., & Caldwell, J. (1995). *Qualitative Reading Inventory-Ii (Qri-Ii).* Harper Collins College Publishers.

Lombardino, L. J., Riccio, C. A., Hynd, G. W., & Pinheiro, S. B. (1997). Linguistic deficits in children with reading disabilities. *American Journal of Speech-Language Pathology*, *6*(3), 71-78.

Maggin, D. M., Chafouleas, S. M., Goddard, K. M., & Johnson, A. H. (2011). A systematic evaluation of token economies as a classroom management tool for students with challenging behavior. *Journal of School Psychology*, *49*(5), 529–554.

McNamara, J. K., Scissons, M., & Gutknecth, N. (2011). A longitudinal study of kindergarten children at risk for reading disabilities: The poor really are getting poorer. *Journal of learning disabilities*, *44*(5), 421-430.

Michael, J. (1993). Establishing operations. *The behavior analyst*, *16*(2), 191-206.

Morgan, P. L., Farkas, G., Tufis, P. A., & Sperling, R. A. (2008). Are reading and behavior problems risk factors for each other?. *Journal of learning disabilities*, *41*(5), 417-436.

National Center for Education Statistics (NCES) (2019). The nation's report card: Reading 2019. National Assessment of Educational Progress at grades 4 and 8. Washington, DC: US Department of Education.

National Center for Education Statistics (NCES) (2022). NAEP Long-Term Trend Assessment Results: Reading and Mathematics. Washington, DC: US Department of Education.

National Reading Panel. (2000) *Report of the National Reading Panel--Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and Its Implications for Reading Instruction.* Washington, D.C.: National Institute of Child Health and Human Development.

Nelson, J. R., Benner, G. J., Lane, K., & Smith, B. W. (2004). Academic achievement of K-12 students with emotional and behavioral disorders. *Exceptional children*, *71*(1), 59-73.

NICHD Early Child Care Research Network. (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental psychology*, *41*(2), 428-442.

Novak, G., & Hammond, J. M. (1983). Self-reinforcement and descriptive praise in maintaining token economy reading performance. *The Journal of Educational Research*, *76*(3), 186–189.

Oakes, W. P., Mathur, S. R., & Lane, K. L. (2010). Reading interventions for students with challenging behavior: A focus on fluency. *Behavioral Disorders*, *35*(2), 120–139. https://doi.org/10.1177/019874291003500204

Pickren, S. E., Chow, C. J., & Cutting, L.E., (in prep). The relation between reading and externalizing behavior: A correlational meta-analysis.

Prasad, V., Brogan, E., Mulvaney, C., Grainge, M., Stanton, W., & Sayal, K. (2013). How effective are drug treatments for children with ADHD at improving on-task behaviour and academic achievement in the school classroom? A systematic review and meta-analysis. *European child & adolescent psychiatry*, *22*(4), 203-216.

Pitts, C., Pillow, T., Dusseault, B., & Lake, R. (2022). Virtual Learning, Now and Beyond. *Center on Reinventing Public Education*.

Poncy B. C., Skinner C. H., Axtell P. K. (2005). An investigation of the reliability and standard error of measurement of words read correctly per minute using curriculum-based measurement. *Journal of Psychoeducational Assessment*, 23, 326-338.

Quick, L. (2014). *Creating an individualized positive reinforcement menu.* Georgia Association for Positive Behavior Support Conference, 60. https://digitalcommons.georgiasouthern.edu/gapbs/2014/2014/60

Regmi, K., & Jones, L. (2020). A systematic review of the factors–enablers and barriers–affecting e-learning in health sciences education. *BMC Medical Education*, *20*, 1–18.

Reinke, W., Herman, C., & Stormont, M. (2013). Classroom-level positive behavior supports in schools implementing SW-PBIS: Identifying areas for enhancement. Journal of Positive Behavior Interventions, 5, 39–50.

Reid, R., Gonzalez, J. E., Nordness, P. D., Trout, A., & Epstein, M. H. (2004). A meta-analysis of the academic status of students with emotional/behavioral disturbance. *The Journal of Special Education*, *38*(3), 130–143. https://doi.org/10.1177/00224669040380030101

*Resources RTI$^2$-B | Tier 2 | TBSP Memphis*. (n.d.). TBSP | Supporting RTI2-B in West Tennessee Schools. Retrieved August 18, 2020, from http://www.tbspmemphis.com/resources/tier2

Rosenshine, B. V. (2015). How time is spent in elementary classrooms. *The Journal of Classroom Interaction*, *50*(1), 41-53.

Royer, D. J., Lane, K. L., Dunlap, K. D., & Ennis, R. P. (2019). A systematic review of teacher-delivered behavior-specific praise on K–12 student performance. *Remedial and Special Education*, *40*(2), 112-128.

Rutter, M., Tizard, J., Yule, W., Graham, P., & Whitmore, K. (1976). Isle of Wight Studies, 1964–19741. *Psychological Medicine*, *6*(2), 313-332.

Scott, T. M., Alter, P. J., & Hirn, R. G. (2011). An examination of typical classroom context and instruction for students with and without behavioral disorders. *Education and Treatment of Children*, *34*(4), 619-641.

Skinner, B. F. (1948). 'Superstition' in the pigeon. *Journal of experimental psychology*, *38*(2), 168.

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review*, *6*(1), 1–26. ProQuest Social Sciences Premium Collection. https://doi.org/10.1016/j.edurev.2010.07.002

So, C. Y., Leung, P. W., & Hung, S. F. (2008). Treatment effectiveness of combined medication/behavioural treatment with Chinese ADHD children in routine practice. *Behaviour Research and Therapy*, *46*(9), 983-992.

Soares, D. A., Harrison, J. R., Vannest, K. J., & McClelland, S. S. (2016). Effect size for token economy use in contemporary classroom settings: A meta-analysis of single-case research. *School Psychology Review*, *45*(4), 379–399. https://doi.org/10.17105/SPR45-4.379-399

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, *10*(2), 176-199.

Solis, M., El Zein, F., Vaughn, S., McCulley, L. V., & Falcomata, T. S. (2016). Reading comprehension interventions for students with autism spectrum disorders: An alternating treatments comparison. *Focus on Autism and Other Developmental Disabilities*, *31*(4), 284–299. https://doi.org/10.1177/1088357615583464

Speight, R., Whitby, P., & Kucharczyk, S. (2020). Impact of CW-FIT on student and teacher behavior in a middle school. *Journal of Positive Behavior Interventions*, *22*(4), 195-206.

Sesma, H. W., Mahone, E. M., Levine, T., Eason, S. H., & Cutting, L. E. (2009). The contribution of executive skills to reading comprehension. *Child neuropsychology*, *15*(3), 232-246.

Shriver, S. K. (2022). Increasing Implementation Fidelity Of Behavior Intervention Plans In Public Schools Following Delivery Of Remote Behavior Skills Training For Staff. (Publication No. 4376) [Doctoral dissertation, University of North Dakota]. Theses and Dissertations.

Stanovich K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*. 21:360–407.

Sutherland, K. S., & Oswald, D. P. (2005). The relationship between teacher and student behavior in classrooms for students with emotional and behavioral disorders: Transactional processes. *Journal of Child and Family Studies*, *14*(1), 1.

Tan, K. H., Kasiveloo, M., & Abdullah, I. H. (2022). Token Economy for Sustainable Education in the Future: A Scoping Review. *Sustainability*, *14*(2), 716.

Tangel, D., & Blachman, B. (2008). *Road to Reading: A Program for Preventing and Remediating Reading Difficulties* (1st edition). Brookes Publishing.

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education*, *25*(4), 252-261.

Thompson, L. A., Ferdig, R., & Black, E. (2012). Online schools and children with special health and educational needs: Comparison with performance in traditional schools. *Journal of Medical Internet Research*, *14*(3), e62. https://doi.org/10.2196/jmir.1947

Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (1999). *TOWRE: Test of word reading efficiency*. Pro-ed Austin, TX.

Vasquez, E., Forbush, D. E., Mason, L. L., Lockwood, A. R., & Gleed, L. (2011). Delivery and evaluation of synchronous online reading tutoring to students at-risk of reading failure. *Rural Special Education Quarterly*, *30*(3), 16–26. https://doi.org/10.1177/875687051103000303

Vasquez, E., & Serianni, B. A. (2012). Research and practice in distance education for K-12 students with disabilities. *Rural Special Education Quarterly; Morgantown*, *31*(4), 33–42.

Vasquez, E., & Slocum, T. A. (2012). Evaluation of synchronous online tutoring for students at risk of reading failure. *Exceptional Children*, *78*(2), 221–235. https://doi.org/10.1177/001440291207800205

Vasquez III, E., & Straub, C. (2012). Online instruction for K-12 special education: A review of the empirical literature. *Journal of Special Education Technology*, *27*(3), 31–40. https://doi.org/10.1177/016264341202700303

Vaughn, S., Levy, S., Coleman, M., & Bos, C. S. (2002). Reading instruction for students with LD and EBD: A synthesis of observation studies. *The Journal of Special Education*, *36*(1), 2–13. https://doi.org/10.1177/00224669020360010101

Vreeburg Izzo, M., Yurick, A., Nagaraja, H. N., & Novak, J. A. (2010). Effects of a 21st-century curriculum on students' information technology and transition skills. *Career Development for Exceptional Individuals*, *33*(2), 95–105.

Ward-Horner, J., & Sturmey, P. (2012). Component analysis of behavior skills training in functional analysis. *Behavioral Interventions*, *27*(2), 75-92.

Warmbold-Brann, K., Burns, M. K., Preast, J. L., Taylor, C. N., & Aguilar, L. N. (2017). Meta-analysis of the effects of academic interventions and modifications on student behavior outcomes. *School Psychology Quarterly*, *32*(3), 291.

Wehby, J. H., & Copeland, S. R. (2000). Effect of varying rates of behavior-specific praise on the on-task behavior of students with EBD. *Journal of Emotional and Behavioral Disorders*, *8*(1), 2-8.

Wehby, J. H., Symons, F. J., & Shores, R. E. (1995). A descriptive analysis of aggressive behavior in classrooms for children with emotional and behavioral disorders. *Behavioral Disorders*, 20, 87–105.

Willcutt, E. G. (2014). Behavioral genetic approaches to understand the etiology of comorbidity. In *Behavior genetics of psychopathology* (pp. 231-252). Springer, New York, NY.

Willcutt, E. G., Pennington, B. F., Olson, R. K., & DeFries, J. C. (2007). Understanding comorbidity: A twin study of reading disability and attention-deficit/hyperactivity disorder. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *144*(6), 709-714.

Willcutt, E. G., & Pennington, B. F. (2000). Psychiatric comorbidity in children and adolescents with reading disability. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(8), 1039-1048.

Willcutt, E. G., Petrill, S. A., Wu, S., Boada, R., DeFries, J. C., Olson, R. K., & Pennington, B. F. (2013). Comorbidity between reading disability and math disability: Concurrent psychopathology, functional impairment, and neuropsychological functioning. *Journal of learning disabilities*, *46*(6), 500-516.

Wills, H., Wehby, J., Caldarella, P., Kamps, D., & Swinburne Romine, R. (2018). Classroom management that works: A replication trial of the CW-FIT program. *Exceptional Children*, *84*(4), 437-456.

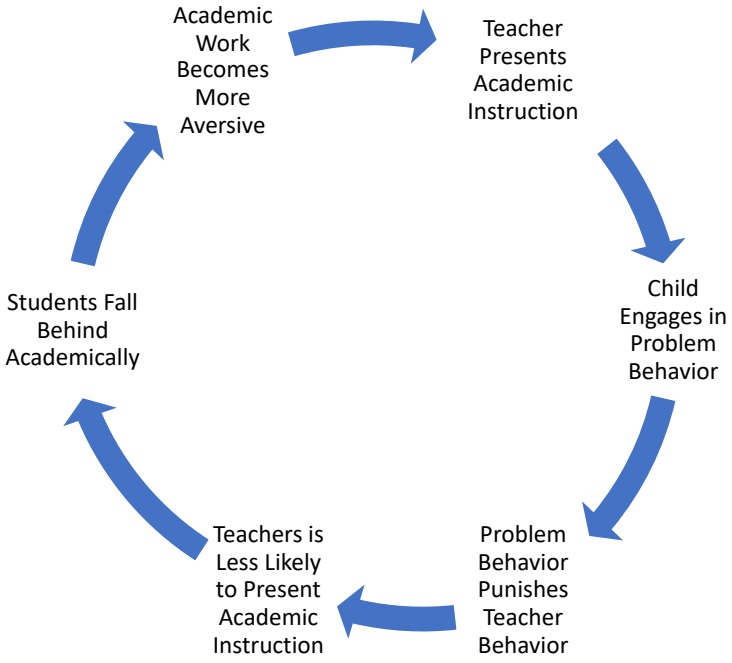
Wise, J. C., Sevcik, R. A., Morris, R. D., Lovett, M. W., Wolf, M., Kuhn, M., ... & Schwanenflugel, P. (2010). The relationship between different measures of oral reading fluency and reading comprehension in second-grade students who evidence different oral reading fluency difficulties. *Language, Speech, and Hearing Services in Schools, 41*, 340-348.

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart 1. *Journal of Applied Behavior Analysis*, *11*(2), 203–214.

Yong, T. L., & Ping, L., C. (2008). Engaging academically at risk primary school students in an ICT mediated after school program. *Australasian Journal of Educational Technology* 24, 521-539.

# Appendix A

## Cycle of Negative Reinforcement/ Non-instruction



Appendix B

Baseline Engagement for Discontinued Participants

| | Percentage of Intervals Engaged | | |
|---|---|---|---|
| | **Ryan** | **Cody** | **Ed** |
| **Session 1** | 91.70 | 88.44 | - |
| **Session 2** | 92.31 | 91.43 | - |
| **Session 3** | 99.07 | 43.75 | - |
| **Session 4** | 99.15 | 100.00 | - |
| **Session 5** | - | 98.97 | 99.50 |
| **Session 6** | - | - | - |
| **Session 7** | - | - | 98.49 |
| **Session 8** | - | - | 99.05 |
| **Mean** | 95.56 | 84.52 | 99.01 |

Appendix C

Tutoring Rules and Reward Menus

Melinda's Tutoring Rules

# Reading Clinic Rules

---

- Be safe
  - Stay in the room with your computer
  - If you need to get up to use the bathroom or get a tissue, please ask.
  - Sit up in your chair.
- Be respectful
  - Use kind words + a kind voice
  - Do not talk when I am talking; I will not talk when you are talking.
- Be responsible
  - Look at the screen when I'm teaching
  - Respond to questions
  - Ask a question if you don't understand something

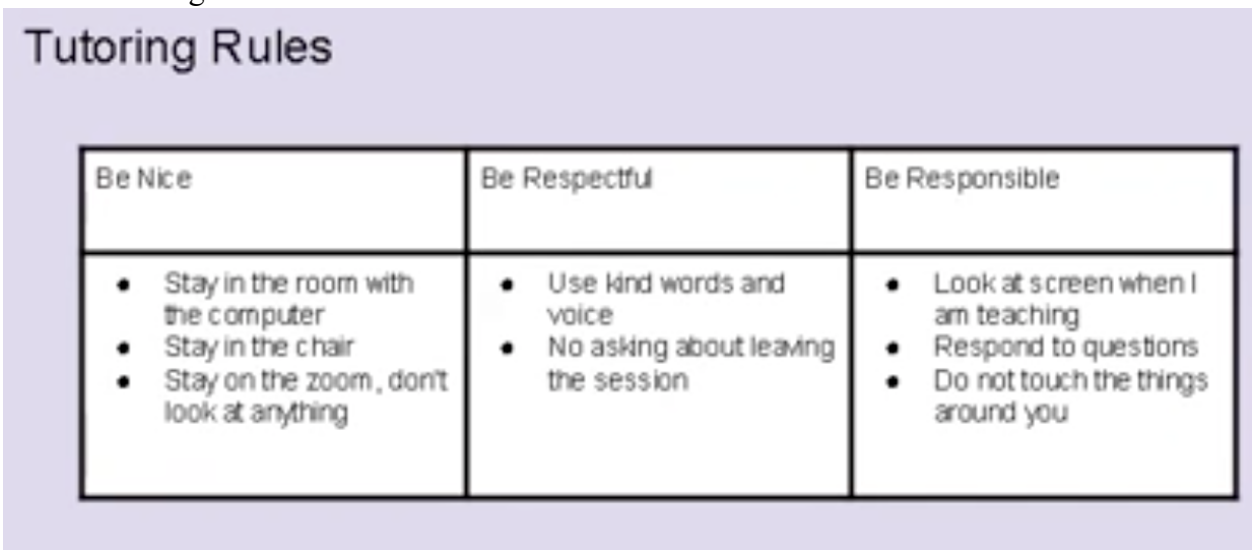
Melinda Reward Menu

# List of Prizes

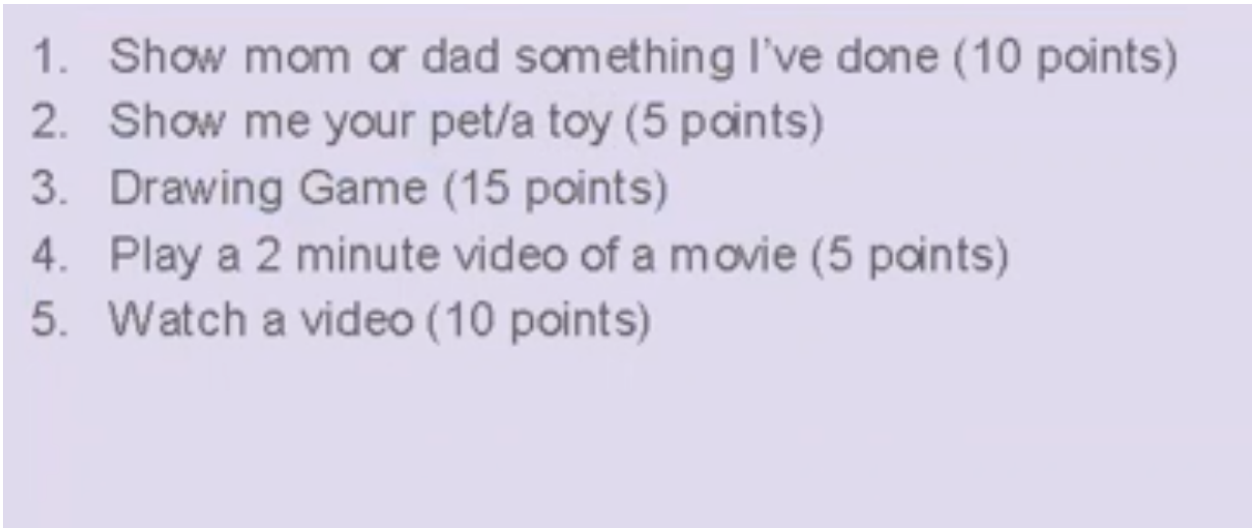
---

| | |
|---|---|
| 1-minute Talking Break | **3 Points** |
| See Archie (my dog) | **3 Points** |
| Show me a dog on PetFinder | **5 points** |
| 2-minute talking break | **5 points** |
| Pick a HILARIOUS zoom background | **8 points** |
| Pick a HILARIOUS theme for next session | **10 points** |
| Watch a YouTube video | **10 points** |

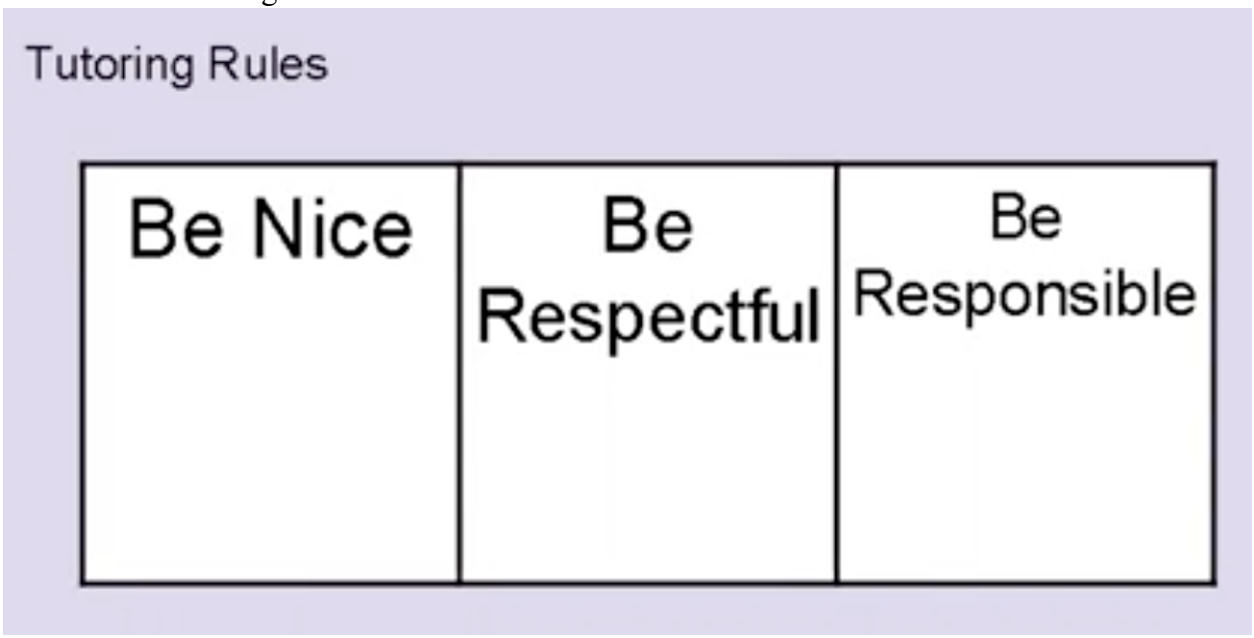
John's Tutoring Rules




John's Reward Menu



Alexander's Tutoring Rules




Alexander's Reward Menu

1. Getting to leave 5 minutes early (15 points)
2. Playing a 2 minute video from a movie (10 points)
3. Getting to draw a picture on the white board (5 points)
4. Getting to play a game (5 points)
5. Look at pictures of dinosaurs (5 points)

Appendix D

Coding Manual

Engagement and disengagement are duration codes. Students must be marked as either engaged or disengaged at the end of every 15 second interval. These codes are opposites of each other. We are using momentary time sampling, so make this judgement call only based on the student's behavior *at the moment* that the interval ends. If the tutor allows the student to take a break during the session (e.g., leaves to go to the bathroom) do not code anything for that amount of time.

| **Engagement (Duration)** | | | |
|---|---|---|---|
| Code | Definition | Examples | Non-Examples |
| eg | Engagement:<br>Student demonstrates engagement in the activity at hand. Examples of engagement can be active, such as responding to the tutor's directions and prompts, sounding out target words, reading aloud, thinking while looking down at paper/whiteboard, and asking academic questions. Examples can also be more passive forms of engagement, such as attentive body language such as orienting towards the computer camera, nodding, eye contact, and staying in the seat. | - Student responds to the instructor's cues by reading the word presented on the screen.<br>- Student raises hand to ask a question.<br>- Student looks at surroundings but returns to the task within [a reasonable amount of time, say, 5 seconds].<br>- Student responds to the instructor, but not verbally by holding up fingers to indicate "four."<br>- Student sits still in chair and listens as instructor explains behavior expectations for the lesson<br>- Student is focused on the screen while practicing pronunciation.<br>- Student is nods 'yes' when asked if directions are clear. | - Student gets up and walks around mid-lesson without instruction to do so.<br>- Student is thinking out loud about a topic that is unrelated to the lesson.<br>- Student is utilizing the materials such as the whiteboard in a manner that is not relevant to the lesson.<br>- Student is not responsive to the instructor's directions.<br>- Student is consistently looking around the room instead of the screen.<br>-Student is taking a tutor-sanctioned break |

| **Disengagement (Duration)** | | | |
|---|---|---|---|
| Code | Definition | Examples | Non-Examples |
| dg | Disengagement: | - Student takes longer than [5] seconds to | - Student responds to instructor's prompt |

| | | | |
|---|---|---|---|
| | Student is not actively engaged in the activity at hand. This includes both physical, verbal, and nonverbal acts that slow down completion of the tutoring activity. If student happens to be engaging in disruptive behavior at the end of the interval, this is categorized as disengagement. (Code both bd and dg.) However, there are some forms of minor disengagement (like zoning out) that would be categorized as disengagement, but not count as a disruptive behavior. | respond to instructor's academic or behavioral prompt.<br>- Student is unresponsive to instructor's attempts to refocus/redirect.<br>- Student is talking about a subject that is unrelated to the lesson (unless tutor is encouraging the conversation- then code engaged)<br>- Student stands up from seat and walks around the room, without being instructed to. | with an incorrect academic response.<br>- Student's body behavior is wiggly (shifting in chair, looking around room) but student is still responding to instructor's cues.<br>- Student responds to the instructor, but not verbally by holding up fingers to indicate "four."<br>-Student is taking a tutor-sanctioned break |

The remaining codes are frequency codes. Mark these instances of behavior every time they occur within the appropriate interval.

| **Disruptive Behavior: (Frequency)** | | | |
|---|---|---|---|
| Code | Definition | Examples | Non-Examples |
| bd | Disruptive Behavior:<br>Student is not actively engaged in the activity at hand, but rather, is engaged in a behavior that actively disrupts the flow of the lesson. Examples of disruptive behavior include getting up from seat without directions, touching toys, and playing with stimuli on computer in a non-productive manner. Count each time the behavior starts or changes form. If student is demonstrating disruptive behavior that is persisting, count it once for each interval or each time it changes form. | - Student uses the computer controls (dragging items) in an unproductive manner.<br>- Student eats/drinks in a manner that disrupts the lesson.<br>- Student disappears from camera view intentionally, in an attempt to disrupt the lesson.<br>-Student complains, groans, or says a bad word.<br>-Student pretends to minimize a distraction, but instead uses it as an opportunity to delay tutoring (e.g., leaves computer for a long time to help a pet, takes an | - Student stretches for a few seconds before returning to the lesson.<br>- Student disappears from camera view accidentally.<br>-Student asks how much time is left in the tutoring session.<br>-Student gets up from the computer briefly to minimize a distraction (e.g., shuts a door to a noisy room, puts on headphones). |

| | | | |
|---|---|---|---|
| | | excessively long time to put on headphones). | |

| **Praise (Frequency)** | | | |
|---|---|---|---|
| Code | Definition | Examples | Non-Examples |
| gp | Generic Praise:<br>Teacher praises and provides generic, positive feedback to the student that does not specifically mention the activity at hand. Praise includes *verbal* statements that indicate approval of behavior that is deemed satisfactory by the instructor or an acknowledgement of a correct response. Even if the praise is two sentences, if referring to one positive behavior, count only once. If a praise is one sentence but references two positive behaviors, count twice. | - "Very good job, Amelia!"<br>- "Perfect!"<br>- "[Your work] looks beautiful!"<br>- "Awesome!"<br>- "Yup!" (When in response to a correct answer.")<br>-"Wow!" | - "Awesome job staying on task!"<br>- "I noticed that you're doing a great job reading those words."<br>-Tutor smiles and shows a student a thumbs up |
| sp | Behavior-Specific Praise:<br>Teacher praises and provides positive feedback that goes beyond generic phrases and explicitly mentions the students' action. Praise includes *verbal* statements that indicate approval of behavior or academic work that is deemed satisfactory by the instructor. Even if the praise is two sentences, if referring to one positive behavior, count only once. If a praise is one sentence but references two positive behaviors, count twice. | - "Thank you for reading out loud that word so nicely."<br>- "You're doing a great job of staying focused on our activity."<br>- "You are becoming such a good reader. You said that word correctly and quickly." (1 count)<br>-"You did such a good job sounding out that word and you were in your seat the whole time." (2 count)<br>-"Wow! You got 4 out of 5 right!" (1 count) | - "Good!"<br>- "Great job on that exercise!"<br>- Instructor says, "Thank you," to student without specificity.<br>-Tutor smiles and shows the student a thumbs up |

| **Refocus (Frequency)** | | | |
|---|---|---|---|
| Code | Definition | Examples | Non-Examples |

| | | | |
|---|---|---|---|
| r | Refocus:<br>Teacher responds to disengaged behavior by refocusing the student's attention to the activity. Refocusing may or may not verbally reference the disruptive behavior. Examples include changing the conversation topic and restating/rephrasing the question. Refocusing can be gentle and sound like reminders or be aggressive and sound like reprimands.<br>Consecutive statements of refocus are coded as one instance unless there is at least a five second interval between the statements]. | - "You need to look at the screen, Amelia."<br>- "Amelia, [asking a second time] can you tell me how many letters are in this word?"<br>- In response to student's disruptive behavior: "Oh, okay. Well, let's try this activity instead."<br>-Stop spinning in your chair!<br>-"Pay attention to what I'm doing please."<br>-"If you do that again, I'm going to have to tell your mom." | - Tutor talks to the student about non-academic things<br>-Tutor makes a mean facial expression at the student<br>-Tutor transitions to a new activity and explains what the student will have to do. |

*If a behavior occurs right when an intervals end, mark it only once, in the interval that it ends on. (Ex. if praise occurs between 0:14-0:16, indicate this in the 0:15-0:30 mark).

*If student is on a tutor-approved break or is receiving tutor-approved award time, do not code any buttons during that time.

Appendix D

Baseline and Intervention Procedural Fidelity Forms

**VKC-RC Token Economy**
**Procedural Fidelity Form (Intervention)**

**Instructions**: Check whether each item was present (yes) or absent (no) during the session.

| **Participant ID:** | **Date:** | **Data collector:** | **Prim or Reli** (circle one) |
|---|---|---|---|

| **Session #** | | |
|---|---|---|
| | **Yes** | **No** |
| **Before instruction** | | |
| Tutor reminds participant of the target behaviors (visible) | | |
| Tutor reminds student that tokens will be earned by following the rules and that the tokens can be used to purchase rewards. (can be verbal) | | |
| **During Instruction** | | |
| Tutor sets vibrating timer to programmed interval. | | |
| Token board is visible to student. | | |
| Tutor delivers tokens correctly (i.e., token delivered when child is meeting target behavior and token *not* delivered when child is not meeting target behavior). | Correct Tally: | Incorrect Tally: |
| **After instruction** | | |
| Reward menu with pricing is made available to student. (visible) | | |
| Tutor allows student to cash in tokens for a reward at the end of the session. | | |
| **PF score = (yes)/(yes + no) x 100** | | |
| | | |

| **Notes:** |
|---|
| |

**VKC-RC Token Economy**
**Baseline Procedural Fidelity Form**

**Instructions**: Check whether each item was present (yes) or absent (no) during the session.

| **Participant ID:** | **Date:** | **Data collector:** | **Prim or Reli** (circle one) |
|---|---|---|---|

| **Session #** | | |
|---|---|---|
| | **Yes** | **No** |
| **Before instruction** | | |
| Tutor reminds participant of the target behaviors | | |
| Tutor reminds student that tokens will be earned by following the rules and that the tokens can be used to purchase rewards. | | |
| **During Instruction** | | |
| Tutor sets vibrating timer to programmed interval. | | |
| Token board is visible to student. | | |
| Tutor delivers tokens | Tally: | |
| **After instruction** | | |
| Reward menu with pricing is made available to student. | | |
| Tutor allows student to cash in tokens for a reward at the end of the session. | | |
| **PF score = (no)/(yes + no) x 100** | | |

| **Notes:** |
|---|
| |

# Appendix F

## Tutor and Student Social Validity Surveys

### Tutor Social Validity

VKC-RC APBS

How easy was the token economy to implement? (1=not at all; 5=very easy)

1 2 3 4 5

If given the opportunity, how likely would you be to continue using the token economy with the same student or a similar student?
(1=not at all; 5=very likely)

1 2 3 4 5

How much did the token economy improve your student's behavior?
(1=not at all; 5= completely)

1 2 3 4 5

I would suggest using a token economy with similar procedures to other teachers who have students with challenging behavior. (1= disagree, 5=agree)

1 2 3 4 5

The benefits of this intervention outweighed any negative side effects.
(1=disagree; 5=agree)

1 2 3 4 5

I liked implementing the procedures in this intervention.
(1=disagree; 5=agree)

1 2 3 4 5

Comments regarding the intervention's feasibility:

_____

_____

_____

_____

What changes would you suggest for implementing this intervention in the future?

_____

_____

_____

## Caregiver Social Validity

VKC-RC APBS

In the past, my child has had trouble focusing or behaving well during school or tutoring. (1=disagree; 5=agree)

1 2 3 4 5

In the past, my child has complained about attending reading clinic sessions or classes at school. (1=disagree; 5=agree

1 2 3 4 5

Earning points helped my child behave better during tutoring.
(1=disagree; 5= agree)

1 2 3 4 5

Earning points helped my child be more excited about tutoring sessions.
(1= disagree, 5=agree)

1 2 3 4 5

I think the rewards my child earned were appropriate and did not take up too much time.
(1=disagree; 5=agree)

1 2 3 4 5

Comments about participating in the research study:

_____

_____

_____

_____

What did you or your child like or dislike about participating in the research study?

_____

_____

_____

_____

## Student Social Validity

VKC-RC APBS

I liked earning points during tutoring. (1=disagree; 5=agree)

1 2 3 4 5

I liked the rewards I earned during tutoring.
(1=disagree; 5=agree

1 2 3 4 5

Earning points helped me learn and behave better during tutoring.
(1=disagree; 5= agree)

1 2 3 4 5

I wish my other teachers at school could let me earn points or do something similar to help me.
(1= disagree, 5=agree)

1 2 3 4 5

I think other kids would like to earn points during tutoring like I did.
(1=disagree; 5=agree)

1 2 3 4 5

What were some of your favorite rewards?

_____
_____
_____
_____

What did you like or not like about earning points?

_____
_____
_____
_____