Quasi-Experimental Study of Middle School Tokens, Behaviors, Goals, Preferences, and Academic Achievement

A Dissertation submitted
to the Graduate School
Valdosta State University

in partial fulfillment of requirements
for the degree of

DOCTOR OF EDUCATION

in Curriculum and Instruction

in the Department of Curriculum, Leadership, and Technology
of the Dewar College of Education and Human Services

May 2017

Kelly C. Dreger

M.Ed., Albany State University, 2010
B.S., Albany State University, 2008
A.S., Darton State College, 2005

This dissertation, "Quasi-Experimental Study of Middle School Tokens, Behavior, Goals, Preferences, and Academic Achievement," by Kelly C. Dreger, is approved by:

Dissertation Committee Chair/Researcher

Steven E. Downey, Ph.D.
Associate Professor, Department of Curriculum, Leadership, and Technology

Dissertation Co-Chair

Jennifer M. Branscome, Ph.D.
Associate Professor, Department of Psychology, Counseling, and Family Therapy

Committee Members

Heather M. Kelley, Ph.D.
Associate Professor, Department of Psychology, Counseling, and Family Therapy

William F. Truby, Ph.D.
Assistant Professor, Department of Curriculum, Leadership, and Technology

Interim Dean of the Graduate School

James T. LaPlant, Ph.D.
Professor of Political Science

FAIR USE

DUPLICATION

I authorize the Head of Interlibrary Loan or the Head of Archives at the Odum Library at Valdosta State University to arrange for duplication of this dissertation for educational or scholarly purposes when so requested by a library user. The duplication shall be at the user's expense.

Signature 

I refuse permission for this dissertation to be duplicated in whole or in part.

Signature _____

ABSTRACT

There is a high demand for standardized instructional strategies that align with decision-making frameworks found in public schools. Teachers who use tokens in the classroom can create more avenues for standardization, provided that their methods address the contradictory nature of past research. A lack of knowledge about reinforcement options and their possible effects has been found within a variety of research studies that attempted to bridge gaps between research and practice. This study aimed to create an incentive system that demonstrated how teachers could use tokens as academic and behavioral supports for students in southwest Georgia.

The purpose of this convergent parallel study was to assess the degree to which token use, type, and timing affected performance on nine weeks achievement tests for students in grades 6-8. Additionally, there was an investigation about the extent to which achievement scores were influenced by environmental factors such as the amount of behavioral referrals received, the goals students met to pass tests, and the preferences that students had concerning motivation. Students' and teachers' perspectives on tokens were explored throughout this process in order to accurately gather information on their experiences.

Overall results for this study showed that those without tokens outperformed students who received tokens during the study. Within the token groups, the highest performers were those who received points in the first half and coins in the second half. Students who received no referrals had better performance and behavior than students who did receive referrals. Students accurately assessed how well they would do on tests, and those who met goals for math performed significantly better than those who did not meet math goals. Motivational preferences alone did not guarantee good performance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my committee for their support. Dr. Downey, there are no words to express the gratitude I have for your time and dedication. I worked with you in RSCH 9800, in RSCH 9870, and on this project. You are truly one of a kind, and you deserve a promotion for your patience. Dr. Branscome, thank you so much for sharing your expertise about psychology and dissertation writing. You have supplied me with the fundamental details that I needed for making this study credible. Dr. Kelley, I would not have had this topic if I skipped your PSYC 8020 class. You have provided me with advice, resources, and the optimism I needed to complete the project. Dr. Truby, your tireless efforts to get schools on board with this are awe-inspiring. You did not know me, but you had faith in what I could do. Thank you for helping me make this design a reality.

Also, I would like to take this opportunity to show my sincere thanks to the administrators, faculty, and staff at the school in which I conducted my study. I could not have done this without all of you. I do understand the pressures and time constraints that are placed on principals, instructional coordinators, and math educators. This is especially true during the last quarters of the school year. Thank you for volunteering and for giving me the essentials to see this through to the end. The students and parents deserve a thank you as well. They obviously play a big role within teaching and learning.

I must take the time to thank every teacher who had the chance of ever instructing me. You know who you are. You provided me with the foundation I needed for my life pursuits. Teachers are often unappreciated for the work they do, and this is my time to say that you really do make a difference in this world.

A special thank you goes to one of my fellow colleagues, Megan Ames. I enjoyed working and communicating with her during coursework. Out of all the students I asked to help me with my validity check, she was the one who came through for me in my time of need. Megan, thank you so much for your time and perseverance.

Furthermore, I must thank my mom for everything she has ever done for me. She has always encouraged me to follow my dreams, use common sense, and expand my mind. She is not very fond of computer technology, and she cringes whenever I decide to have personal discussions with her on operant conditioning. Somehow she has been able to endure them for my sake. I love you mom! I am very fortunate to have you in my life.

Last, but certainly not least, are my friends who helped me keep sane during this endeavor. Roki, those readings and proofreading exercises kept my mind fresh. Look forward to more of your wisdom in the future. Plus, I may need some more technical support, so you can look forward to that. Neola, your free spirit encourages me to keep an open mind about what I am doing. Thanks for your guidance. Your experiences are just as wacky as mine! I am glad you are not afraid to show it. Eleanor, your support has made me a better person. I have been able to discover who I am and what I want out of life because you have been there for me. I unreservedly admire you as a person because of your ability to be comforting to others. F.A., a big hug is in order for you! Because of you, I have learned that being a fighter is not so bad after all. My emotions and frustrations were the fuel I needed to get through this. You have taught me that asking questions and stirring the pot are necessary for personal development. Thanks!

DEDICATION

There are two people who brought me into this world, and I will forever be in gratitude. It is because of them that I have been able to have these experiences and opportunities.

Chapter I

INTRODUCTION

Background

Within modern-day education, there is a need to determine strategies that will improve how well middle school students perform on assessments (American Psychological Association, 2015; Wolfe, Dattilo, & Gast, 2003). The use of incentives is a possibility that has been investigated within education and research (Anderman & Maehr, 1994; Cameron & Pierce, 1996; Deci, Koestner, & Ryan, 2001; Nelson, 2010). One type of incentive is the token. Tokens are tangible items presented to students after the display of desirable behaviors (Hackenberg, 2009; Marinak & Gambrell, 2008). Tokens are later exchanged for prizes that are meaningful to students. Money, points, stars, badges, and checks are examples of tokens (Kazdin & Bootzin, 1972). The main idea of using tokens is to positively reinforce desirable behaviors so the likelihood of those behaviors would increase in the future.

Tokens were initially used as counting devices for developing civilizations (Paschalis, 1987; Schmandt-Besserat, 1992). Clay counting tokens have been found during archeological excavations in the Middle East, and token use dates as far back as 8000 B. C. (Schmandt-Besserat, 1992). They later became an efficient tool of exchange within bartering systems. Token exchange systems were utilized by Greeks during the fifth century B. C. in order to solve problems associated with government and commerce (Iverson, 2010; Paschalis, 1987). Tokens have been useful within academia as well.

They have been used within school settings to encourage academic success and modify behaviors (Skinner, Williams, & Neddenriep, 2004). The educational use of tokens within the United States started around the 1960s, and it continues to this day (Gaughan, 1985; Hackenberg, 2009; Taylor, 2000). Unfortunately, the results of past studies on tokens are contradictory in nature. There have been studies where increases in student participation and academic achievement have been observed with the use of such tokens as money, points, badges, and tickets (Abramovich, Schunn, & Higashi, 2013; Coyle, 2013; Miller, 1981; Truchlicka, McLaughlin, & Swain, 1998). There are also researchers who have recommended the use of tokens for motivation, behavior management, and goal development due to significant increases in classroom performance (Simon, Ayllon, & Milan, 1982; Strahan & Layell, 2006). Others have found that tokens diminish performance gains if they are used for long periods of time (Hayenga & Corpus, 2010; Wulfert, Block, Santa Ana, Rodriguez, & Colsman, 2002). They have asserted that tokens negatively affect the academic achievement of high performing students and middle-income students (Miller & Eller, 1985; Unrau & Schlackman, 2006). The relevance of tokens can depend upon a variety of environmental factors, such as school culture, classroom rules, and resource access (Anderman & Maehr, 1994; Cross, 1981; Self-Brown & Mathews, 2003). In fact, the middle school environment accounts for at least 11% of the variability in student motivation (Maehr, 1990). This means that to a certain extent, issues dealing with student motivation can be attributed to the routines and practices that are present within school systems.

Token systems can help students to learn about individual responsibilities and group expectations (Kazdin, 1982; McLaughlin, 1975; Skinner et al., 2004). In order for

token use to be practical for the public school system, there are certain traits that must be integrated within the token system itself. McLaughlin (1975), for instance, outlines five essential characteristics of practical token systems: (a) effectiveness, (b) ease of implementation and management, (c) low expense, (d) compatibility with school and community attitudes, and (e) high approval rating with students. Effectiveness requires that tokens are able to provide a benefit to the school environment. They could help decrease inappropriate classroom behaviors, increase students' engagement during instruction, or improve students' academic performance over time. Additionally, it is necessary for token systems to be implemented in such a way that they do not hinder the efficiency of everyday activities within the school environment. This would make it easy for teachers to implement and manage the system on an as-needed basis. A cost-effective system would not make unnecessary demands on the school's budget, and the low expense would encourage a variety of individuals to participate. Individuals within the school system would be more receptive to a token system if it aligns with the beliefs and practices of those who participate within the school environment. It would gain additional approval from community members if it shows compatibility with the educational expectations of community organizations. Finally, token systems can be designed in ways that address the interests, needs, and preferences of students. Students must find a reason to invest time into learning the rules associated with the systems that they encounter.

The feasibility of token systems within schools is further discussed by Parish and Parish (1991). The authors state five recommendations for effective token systems. Firstly, they recommend that the amount of tokens given should exceed the amount of

punishment given in the classroom. Positive feedback must outweigh any corrective feedback that is received. Secondly, behavioral conditioning is effective when authority is clearly established by teachers and when students feel that teachers care about them. Students must have a sense that the incentives are relevant to their needs and wants. Thirdly, behavioral conditioning is less effective when students experience psychological reactance due to the perception that they have a lack of choice when interacting with teachers. When students perceive that they do not have freedom of choice, they may do the opposite of what is expected of them as a result of their perceptions. Fourthly, students who understand the rules are less likely to violate them. This understanding leads to more cooperation in order to avoid psychological conflicts within themselves. Lastly, positive and negative consequences for behavior need to be fully understood by teachers before token systems are implemented. Teachers must look at the degree to which the strategies they use could impact what happens during instruction.

Statement of the Problem

Existing literature about token economies emphasizes that tokens can be used to meet desired achievement outcomes; however, many research designs that examine the use of tokens lack the rigor and consistency necessary to develop a standardized, research-based system that would benefit all students (Gaughan, 1985; Maggin, Chafouleas, Goddard, & Johnson, 2011). It is vital that teachers utilize token systems in ways that align research with practice. Educational practices regarding tokens often lack a salient connection to objective research-based concepts, and information about their specific use may be vague within research (Hackenberg, 2009; Maag, 2001). More research is necessary in order to thoroughly examine issues pertaining to scheduling

options, social validity, behavior disruptions, and motivation in the use of token economies (McDonald, Reeve, & Sparacio, 2014; Wolfe et al., 2003). There are major gaps found within past literature that center around these issues. One gap that exists within the literature concerns reinforcement and scheduling within token systems. Token systems require more than one scheduling procedure, but not all studies outline those procedures in detail. Therefore, it is important that more studies are conducted about the use of scheduling options within educational research. Another apparent gap is that studies frequently lack discussion on how the token systems are socially acceptable for the participants involved. Students and teachers would have different perspectives on what is appropriate within the school setting, and more qualitative exploration would be critical in establishing social validity for interventions that incorporate tokens economies within them.

Moreover, it is important to determine the extent to which token systems can address behavior problems for groups who are underrepresented or marginalized within society. The U.S. Department of Education Office for Civil Rights (2014), within its Civil Rights Data Collection (CRDC) database, found that African Americans, American Indians, and Native Alaskan students are disproportionately suspended within public schools. The data within the CRDC further indicated that students with disabilities are more likely to be suspended than students without disabilities. According to the National Center for Educational Statistics (2016), the most recent figures from the Department of Education estimate that there are only 6.5 million public school students who receive special education services nationwide. This makes up approximately 13% of all students who attend public schools in the United States. More research needs to be conducted on

whether or not behavioral interventions, which include token systems, can decrease the amount of suspensions and referrals for these particular groups. Percentages for enrollment, out-of-school suspension, and law enforcement referrals are presented in Table 1.

Table 1

Percentages of Enrollment, Suspension, and Referral for P-12 Schools by Subgroup (CRDC Data)

Subgroup	Enrollment	OSS	Law Enforcement Referrals
Ethnicity			
African American	16%	33%	27%
American Indian	0.5%	2%	3%
Asian	5%	2%	2%
Caucasian	51%	36%	41%
Hispanic	24%	23%	24%
Pacific Islander	0.5%	0.4%	0.3%
Two or More Races	2%	3%	3%
Designation			
SWD	12%	13%	75%
SWOD	88%	6%	25%

Note. OSS = Out of School Suspension; SWD = Students With Disabilities; SWOD = Students Without Disabilities.

Finally, there is a gap that exists in terms of how tokens can be used to measure performance and mastery goals within a specific learning environment (Marinak & Gambrell, 2008). Tokens are often linked with performance goals, but additional

research can be conducted in instances when there are multiple types of domain-specific goal orientations and motivations that exist within a particular subject area. Further research in this area would help to clarify whether or not token systems can be used to meet achievement goals within education. For instance, not enough research is available about how token systems could be influenced by the implementation of the Common Core Georgia Performance Standards (CCGPS), which were derived from achievement goals outlined within the Common Core State Standards Initiative (CCSSI). Georgia educators currently use the CCGPS to guide instructional decisions. Since 2010, standards from the CCSSI have been adopted by 43 states (National Governors Association, n.d.). Many past studies about tokens do not take into consideration these standards and how they might influence instructional practices.

Conceptual Framework

Operant Conditioning

There are two theories that can explain concepts about token use, motivation, behavior management, classroom performance, and goal development. These theories are operant conditioning and goal theory. Operant conditioning is a learning theory that is a part of the behaviorist movement. Behaviorism is a family of theories that emphasizes the need for researchers to look at how overt human behavior can be consistently measured through scientific techniques (Goodman, 2010). According to Skinner (1938), operant conditioning includes the following ideas about behavior:

1. Individual behavior is determined by the consequences that occur after a stimulus-response (S-R) pair is established.
2. A consequence that strengthens a behavior is known as a reinforcer.
3. A consequence that weakens a behavior is known as a punishment.

4. A consequence that neither strengthens nor weakens a behavior is known as a neutral operant.

In operant conditioning, token systems can be used to reinforce behavior (Skinner, 1953). There are different schedules of reinforcement that can be created according to the behaviors that teachers would like students to exhibit within the classroom. Ferster and Skinner (1957) outline these schedules of reinforcement. One schedule of reinforcement is the continuous schedule. In a continuous schedule, individuals receive the same reinforcement each time they display a particular behavior. There are also fixed ratio, variable ratio, fixed interval, and variable interval schedules (Ferster & Skinner, 1957). These four specific schedules are known as partial reinforcement schedules. Fixed ratio schedules reinforce behaviors after a predetermined number of correct behaviors. The rate of reinforcement does not change. It is possible for students to receive reinforcements faster if they display the target behavior at a faster pace. Unlike fixed ratio schedules, variable ratio schedules do not have a constant rate of reinforcement. There can be, however, an average number of responses that an individual has in mind. For instance, a VR 10 schedule would require an overall average of ten responses, but the number of responses needed for each reinforcement delivery would still vary. Fixed interval schedules are implemented in order to award reinforcers to students after a behavior has been displayed for a predetermined amount of time, but variable interval schedules deliver reinforcers after different periods of time. Also, it is possible to reinforce behavior through shaping. Shaping occurs when there is a gradual reinforcement of behaviors that eventually leads to meeting the requirements of a particular target behavior. Shaping is used in situations where the target behavior is nonexistent or rare. Skinner (1958), for instance, described how he used shaping in order

to help a pigeon learn how to use a wooden ball to knock over pins in a miniature bowling alley. The pigeon did not display the target behavior at the onset of the experiment, and it did not peck at the ball on its own. Because of this, he decided to first reinforce when the pigeon looked at the ball. Once it looked at the ball, he then reinforced any behavior that resembled a swipe until the pigeon was able to use its beak to swipe the ball at the pins.

Goal Theory

Like operant conditioning theory, goal theory attempts to explain factors that can impact behavior. It is also known as achievement goal theory, and it discusses how cognitive goals help to direct aspects of motivation and behavior (Dowson & McInerney, 2003; Haselhuhn, Al-Mabuk, Gabriele, Groen, & Galloway, 2007). Before frameworks were derived for goal theory, the concept of goal direction was discussed in the past by various theorists. Within behaviorism, for instance, Tolman (1932) created what is known as purposive behaviorism. He felt that individuals produced behavior for a purpose. Their behavior was more than a response to stimuli. There were cognitive components to behavior as well. It was possible for individual actions to be based on beliefs, attitudes, and the environment (Tolman, 1932, 1948). Behavior would be an attempt to strive toward their goals. Figure 1 provides an illustration of purposive behaviorism.

Because of the work of Tolman (1932, 1948) and others, there was a need within research and education for frameworks that addressed goal orientations and achievement goals. Actual frameworks for goal theory have been developed by a variety of researchers. The traditional version of the theory can be traced back to Diener and

Dweck (1978). It contains two types of goal orientations: helpless and mastery. Helpless goal orientation is commonly referred to as performance goal orientation (Ames & Archer, 1988). Performance goals are those that are competitive short-term goals in the school year.

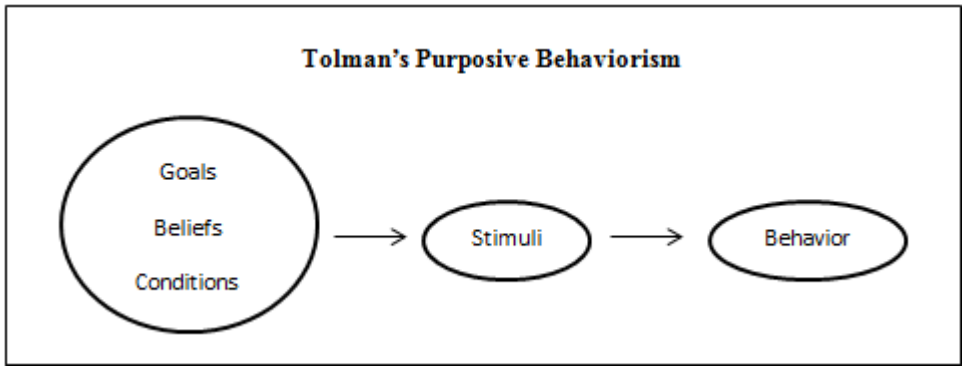


Figure 1. Process Diagram of Goal-direction. Note: This figure attempts to explain goal-directed behavior have been made by different theorists. It provides an explanation of Tolman's perspective on goal direction.

Mastery goals, also known as learning goals, are long-term goals that require individuals to utilize a particular set of skills. Recent research in goal theory has focused on revised frameworks that include the traditional goal orientations. The need for revised frameworks has stemmed from the fact that individuals can have a variety of goals and motivations. Pintrich (2000), for instance, mentions that the traditional goals can have avoid and approach subsets. Individuals who are worried that their mastery and performance will not improve have an avoidance orientation. Those who are confident that they will learn new skills and demonstrate performance goals have an approach orientation. Learning environments that encourage mastery and performance approach goals tend to show more achievement gains than those that only promote one type of goal orientation (Pintrich, 2000; Robustelli, 2006). Furthermore, Robustelli (2006) emphasizes that the type of motivation that students prefer depends on the type of goal

orientations they have. Performance goals typically require extrinsic motivators, whereas mastery goals usually require intrinsic motivators. Extrinsic motivators come from outside of an individual, and intrinsic motivators come from within an individual. Tokens would be important for performance goals since they are often used as extrinsic motivators; however, they have also been utilized in studies to determine their effects on intrinsic motivation (Cameron & Pierce, 1996; Marinak & Gambrell, 2008). Teachers can categorize the goals they want to accomplish in order to know what motivators are needed and how long those motivators would be implemented within the classroom. This study will focus on performance approach and mastery approach goals since the CCGPS require that teachers set high expectations for achievement.

Purpose of the Study

The purpose of this study was to determine if the use of token systems produced any effects on achievement scores from Math Nine Weeks Tests. There are a variety of reinforcement systems used in education today. It is important to look at what works within schools and what helps to keep students performing well on tests. There is an increased emphasis on standardized test scores within the field of education, particularly within math. This study addressed middle school students who were taught within a southwest Georgia school district. This district contained different races and income levels that were often found within schools in southwest Georgia. This diverse mix allowed for the collection and analysis of data from a variety of subgroups. Additional discussion for this school district is provided in Chapter 3. The scope of the study included token systems that have been used with students in grades 6-8. There was a discussion about the major theories and practices concerning tokens. The discussion was

centered around the tenets of operant conditioning, goal theory, and phenomenology. Studies about tokens within the context of operant conditioning focused on how tokens are associated with academic achievement and behavior. Studies that pertain to goal theory discussed how tokens and other tangible stimuli would impact goal development and motivational preferences. Phenomenological studies on tokens required an investigation into how individuals construct meaning in relation to token system experiences.

Research Questions

This study addressed three questions:

1. To what degree, if any, are differences found in Math Nine Weeks Test scores among students who participate in a token program and students who do not?
    a) To what degree are test scores influenced by the timing of token use in the classroom, i.e., students who receive tokens in the first half of the quarter versus students who receive tokens in the second half versus students who do not receive them at all?
    b) To what degree are test scores influenced by the type of token used in the classroom, i.e., points-based system versus coin-based system versus no tokens used at all?
2. To what extent are the test scores influenced by (a) classroom behavior referrals, (b) classroom achievement goals, and (c) students' motivational preferences?
3. What are the perceived experiences of (a) students and (b) teachers involved with the use of token-based systems in their classrooms?

Research Question 1 helped to determine if there are significant differences in academic achievement between middle school students in grades 6-8. The research design for this study required two treatment groups and one control group. Differences among the groups were compared. Academic achievement was measured through a Nine Weeks pretest and posttest. These were tests that were given at the end of each quarter.

Subquestions 1a and 1b helped to describe additional characteristics that could influence the findings. The timing of the pretest and posttest could have impacted results, and the type of tokens used could have affected the findings.

Research Question 2 assessed if there were factors in addition to tokens that could have affected academic achievement scores obtained within the quantitative strand. These variables could have interacted with token systems during the implementation phase. The first factor mentioned is classroom behavior referrals. Data were collected from middle school teachers about the amount of referrals that students received during the study. The second factor mentioned is classroom achievement goals, where it was determined whether or not students met the goals outlined within CCGPS. This included the criterion that students score a 70 or higher on the Nine Weeks pretest and posttest. The third factor listed is students' motivational preferences. Students, at the end of the quasi-experiment, described whether they preferred coins, points, both, or neither during instruction. Their preferences described what worked for them in terms of reinforcement.

Research Question 3 required an investigation about how students and teachers described their experiences with tokens, particularly coins and points. Qualitative data were collected through student focus group interviews and individual teacher interviews that incorporated phenomenology within their designs. These data collection methods helped to document any past experiences with tokens, present experiences with tokens, and current perceptions on the use of token systems during instruction. The interviews took place around the same time as the quasi-experiment. Criteria were used to select a diverse range of participants. These criteria are listed within Chapter 3.

Significance of the Study

This study was significant in that it provided a standardized incentive system that was based on the current policies and practices of Georgia public schools. It helped to develop a practical instructional strategy that addressed diverse issues pertaining to academic achievement. Due to the changing nature of the school environment, these issues were not adequately covered within existing studies. The system proposed within this study was utilized for a wide range of performance levels. It was used to manage classroom behaviors and to motivate students to do their best. It took into account past research concerning token systems, and it had a solid foundation in what makes sense within the school system.

It is important that token systems are implemented in accordance with the CCGPS in order to determine the extent to which teachers and students in southwest Georgia could benefit from them over time. This study utilized the CCGPS within its design, and it utilized school-wide decision-making models that informed procedures. Decision-making models used in Georgia schools today include Positive Behavioral Interventions and Supports (PBIS) and Response to Intervention (RTI) (Georgia Department of Education, 2011, 2013). Professionals within education can share the findings within this study and improve upon them in future research.

Role of the Researcher

Researcher roles can range from completely unobtrusive to completely interactive, and roles may shift according to the research methodology (Punch, 2014; Thomas, 2003). The researcher's role differed for each part of this study. The quantitative strand required an objective, detached outlook on what occurred during the

quasi-experiment. It was necessary that the influence of biases and perceptions was minimized in order to maintain procedural integrity. The researcher controlled for extraneous variables that could compromise the findings in later stages. The procedures of the design were outlined so that others could replicate or adapt it within future research. It was important to point out, however, that there were interactions between the researcher and the participants that were unique in their own right. Consent and data were obtained by interacting with individuals in the public school system. Data collection about classroom events required visits from the researcher on a regular basis. Participant feedback about the study was integral to its success.

The qualitative strand, in contrast, required a subjective approach to inquiry that involved an exploration of biases, values, and perceptions (Punch, 2014). These aspects of the researcher's identity, when thoroughly examined, had less of an influence on the implementation and coding of all interviews. Within qualitative aspects of a study, the researcher is utilized as a human instrument that interacts with all data (Patton, 2002; Seidman, 2006). It was possible for new personal perspectives to emerge during the study due to novel insights. The interview process took into account any unique developments that occur during the study. Therefore, the qualitative aspect of this research required continual reflection of the researcher actions, thoughts, and feelings.

Limitations, Assumptions, and Design Controls

Limitations

There were limitations that existed which affected the design and implementation of this study. They included the following considerations:

1. The findings may not have been generalizable to token systems used in other school environments.

2. For the Nine Weeks pretest and posttest, the ability to establish validity was dependent upon the level of permission given by those who developed and used the testing instruments.

3. Block scheduling and instructional time constraints within the middle school setting affected the study's design and implementation.

4. The amount of tokens awarded and the amount of documentation provided for the study was dependent upon teachers' preferences and biases.

5. Behavior referrals only accounted for those behaviors that the teacher judged to be inappropriate. There were biases that exist as to who receives a written referral for behavior.

6. The cultures, curriculum plans, and practices currently found within the school districts limited the acceptability of any plans associated with token economy systems.

7. The use of verbal feedback and other contingency systems were difficult to control during token reinforcement of behavior.

8. Non-normal distributions existed for the dependent variables.

9. The leaking and dissemination of knowledge about the token treatments occurred among the treatment and control groups.

10. The experiences that were documented within the qualitative strand were subject to different interpretations.

Assumptions

There were also assumptions that were inherent within the study's design:

1. The results did not violate the principle of homogeneity of variances, which states that variances found within the population are equal.

2. Test scores helped to collect data on students' skills, but they could not account for all that students have learned and experienced.

3. Any issues or differences found within the results were byproducts of the protocols, treatments, and processes that were contained within the study's design.

Delimitations

Finally, there were delimitations that provided boundaries for the study:

1. The quantitative and qualitative strands took place within a public school district in southwest Georgia.
2. Teachers and students within grades 6-8 participated in the study.
3. Points-based and coins-based token systems were used for the treatment program found within the quantitative strand.
4. The achievement goals and intervention plans within the study's design aligned with the Common Core Georgia Performance Standards (CCGPS) for math.
5. Students and teachers who were interviewed for the qualitative strand participated in the quantitative strand.

Definition of Key Terms

The following operational definitions are necessary for the successful comprehension of the information presented about this study:

Academic achievement – For the purposes of this study, academic achievement is defined as a passing score on standardized tests. For all Math Nine Weeks Tests, a passing score of 70 is required.

Appropriate classroom behaviors – Student behaviors that adhere to classroom rules. Students who exhibit appropriate behaviors at all times do not receive behavior referrals.

Classroom behavior referrals – A quantitative variable that is defined as the number of written office referrals that students receive during instruction.

Mastery goals – "With a mastery goal, importance is attached to developing new skills. The process of learning itself is valued, and the attainment of mastery is seen as dependent on effort" (Ames & Archer, 1988, p. 260).

Motivation – "Motivation may be defined as the energization (i.e., instigation) and direction of behavior" (Elliot & Covington, 2001, p. 73). Behavior can be energized or directed by different events, and these events are interpreted according to the preferences of the individuals who experience them.

Negative reinforcer – "We define a negative reinforcer (an aversive stimulus) as any stimulus the withdrawal of which strengthens behavior" (Skinner, 1953, p. 185).

Performance goals – "A performance goal reflects a valuing of ability and normatively high outcomes" (Ames & Archer, 1988, p. 260).

Positive reinforcer – "We first define a positive reinforcer as any stimulus the presentation of which strengthens the behavior upon which it is made contingent" (Skinner, 1953, p. 185).

Punisher – "An operation in which an aversive or conditioned aversive stimulus is made contingent upon a response" (Ferster & Skinner, 1957, p. 731).

Students with disabilities – Form of classification designated for special needs students who receive speech services. Academic services are not provided for students with this classification.

Tokens – "A generalized reinforcer distinguished by its physical specifications is the token" (Skinner, 1953, p. 79). Tokens are known as secondary reinforcers, which derive their value from the stimuli that are associated with them. Skinner (1953) uses the example of money to illustrate the exchange process concerning tokens. For the purposes of this study, points and coins are used as tokens.

Summary

This chapter discussed crucial aspects of the study, including information on background, overall purpose, research questions, and key concepts. Due to the diverse nature of the school environment as a whole, it is increasingly important that there are educational strategies that can be differentiated according to the needs of all students. These strategies must also be aligned with national, state, and local standards. It is possible for students to benefit from the use of token incentives. Tokens are instructional interventions that require more research in order to determine how they can be applied to student learning in the 21st century. Existing literature does provide a starting point for how tokens systems can be used within education, but more consistent methods of administration and scheduling are needing in order to collect systemic data about their use. More information must be gathered about how token use might affect academic achievement and how token systems may interact with the variables of classroom behavior, achievement goals, and motivational preferences.

The next chapter discusses existing literature about essential concepts and theories for this study. The literature focuses on how tokens have been used within studies about academic achievement, operant conditioning, goal theory, and phenomenology. There will also be explanations for additional concepts that are not discussed in this chapter. Chapter 3 provides further information about the study's methodology. It includes explanations for sample selection, data collection, instrumentation, and data analysis.

Chapter II

REVIEW OF RELATED LITERATURE

Introduction

This chapter presents a review of findings from existing research studies that explore problems concerning token use. Concepts and theories that are discussed in this chapter are relevant to explore for this study. The purpose of this study was to determine if any effects existed between token use and academic achievement within standardized tests. This study included an examination of other factors that can interact with token designs, such as classroom behaviors, achievement goals, and motivational preferences. Additionally, gaps within the literature are identified that require further exploration within research. There is an overall need for more studies that address token scheduling as well as and how tokens can be utilized across disciplines. The chapter is divided into four sections: (a) Tokens and Academic Achievement, (b) Application of Tokens to Operant Conditioning, (c) Application of Tokens to Goal Theory, and (d) Application of Tokens to Phenomenology.

Tokens and Academic Achievement

Past studies on token economies and achievement showed positive and negative results (Hayenga & Corpus, 2010; Simon et al., 1982). In their study on token reinforcement and academic achievement, Simon et al. (1982) indicated that a positive correlation existed between the use of point tokens and academic achievement. They also found that as academic achievement increased, the amount of disruptive behaviors during class decreased. A case study conducted by Strahan and Layell (2006) found that token

economy use helped a group of seventh-grade students to have higher gains in math and reading test scores. The token system also helped students understand classroom rules and routines. Hayenga and Corpus (2010), however, found in a survey of 388 middle school students that there were negative effects on achievement when external stimuli were used at a high rate. High achievers were not as dependent on external stimuli as low achievers were. Wulfert et al. (2002) found in their study of monetary incentives that high school and middle school students who exhibited low academic performance preferred to receive immediate gratification, even when the amount of money increased with delayed scheduling.

Furthermore, culture and socioeconomic status can play a role in how tokens affect performance (Miller & Eller, 1985; Truchlicka et al., 1998). Unrau and Schlackman (2006) conducted a study on extrinsic stimuli and academic achievement. The participants of the study were predominantly Asian and Hispanic. Their study indicated that significant differences were found with the Asian students in terms of performance. They also found that Asians had high academic achievement scores when low amounts of extrinsic stimuli were used. No significant differences were found for Hispanic students. Miller (1981) conducted an experimental study on 135 middle school students to determine the effects of praise and money on IQ test performance. There were two experimental groups and one control group for the design. The two experimental groups received money and praise at different intervals, but the control group did not receive either intervention. Results showed that African Americans with low socioeconomic status had significant increases in IQ test scores when they were presented with money. Caucasians from middle-income and low-income families had

significant increases in IQ test scores when presented with verbal praise. Miller and Eller (1985) found similar results in their survey about feedback preferences and achievement. Results indicated that African Americans, specifically those who had a low socioeconomic status, preferred money. Caucasian students, however, were more likely to want praise. Devers and Bradley-Johnson (1994) utilized a token economy system to determine the effects of token reinforcement on the Detroit Test of Learning Aptitude-2 (DTLA-2) and the Wechsler Intelligence Scale for Children (WISC-R). There were 31 Native American students in grades 5 through 9 who participated in the study. Poker chips were given to students who answered questions in class correctly, and the chips were exchanged for prizes that were based on students' preferences. The prizes included cash, a curling iron, a tape player, eye shadow, a frisbee, sunglasses, and edible items. Findings from a one-way ANOVA procedure showed that students within the experimental group scored significantly higher on the WISC-R. There were no significant differences found with the DTLA-2.

Application of Tokens to Operant Conditioning

Behaviorist Principles

Classical Conditioning and the Behaviorist Movement

Before the behaviorist movement began in the Western hemisphere, classical conditioning was being studied in the 1890s by a Russian researcher named Ivan Pavlov (Windholz, 1997). His experiments focused on salivary reflex conditioning in animals, particularly dogs (Batuev & Sokolova, 2004; Windholz, 1997). From those experiments, it was found that a new behavior could be learned after links were created between two stimuli. According to Pavlov (1927), this conditioning process occurred in stages:

- Stage 1: Before the conditioning process could actually occur, there is a natural stimulus within the environment that is called the unconditioned stimulus (UCS). This is followed by an unconditioned response (UCR) from the participant. There is also a neutral stimulus (NS) that produces no response from the participant.

- Stage 2: During conditioning, the NS is associated with the UCS. The NS becomes a conditioned stimulus (CS), and the participant learns about the association.

- Stage 3: The participant has learned the association to the point where there is now a conditioned response (CR) to the conditioned stimulus (CS).

Another researcher, John B. Watson, experimented with classical conditioning in the United States. Watson (1913) officially kickstarted the behaviorist movement in the Western hemisphere. The movement was started in response to the general lack of interest within the field of psychology to study behavior from an objective point of view. He argued that data about human behavior could provide information on humans that was not dependent on emotions or mental states of consciousness. Traditional behaviorism centered on the following tenets: (a) organisms use aspects of heredity and behavior to adjust to their environment, (b) individual responses depend on specific stimuli, and (c) stimuli and responses can be predicted if enough information is available about habit patterns (Watson, 1913). One example of a classical conditioning experiment of his involved the conditioning of fear in humans (Watson & Raynor, 1920). An 11-month old infant named Albert learned to fear a white rat due to associations between the rat and loud noises. He later generalized this fear to furry objects. More explanation on classical conditioning and the aforementioned experiments are provided in Figure 2:

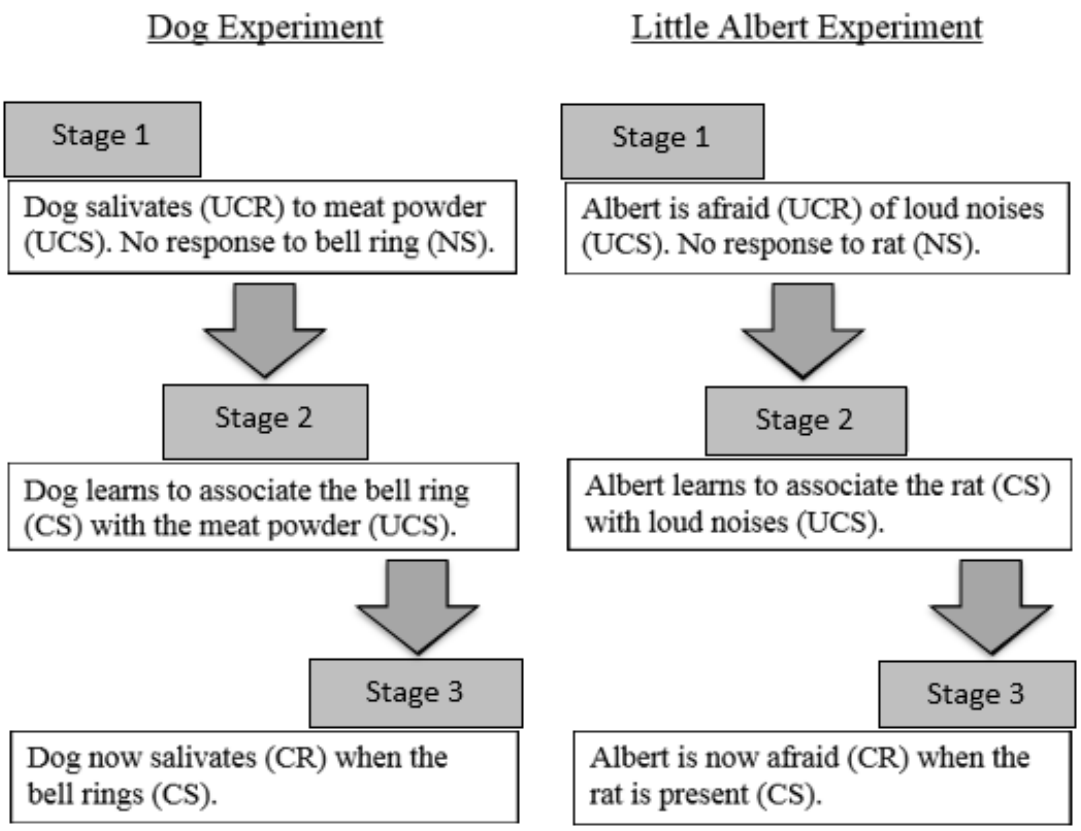


Figure 2. Sequence Diagrams of Classical Conditioning. Note: This figure contains sequence diagrams for classical conditioning that compare one of Pavlov's dog experiments with the Little Albert Experiment.

Thorndike (1913), another behaviorist, was able to build on these ideas by creating the theory of connectionism. Thorndike's (1913) theory suggested that learning occurred when neural connections were made between stimulus and response. These connections formed stimulus-response (S-R) pairs. Furthermore, Thorndike (1913) describes three laws of learning within this theory that explain how the learning process occurs. The first law is the law of readiness, which states that learning is satisfying only when individuals are ready to form the connections necessary for it to occur. When they are not ready, the learning process is annoying to them. The second law is the law of exercise. It states that neural connections strengthen with use, but they weaken with

disuse. The final law, known as the law of effect, states that an established connection can be strengthened if followed by a satisfying condition; however, the connection will be weakened if it is followed by an annoying condition (Schunk, 2012; Thorndike, 1913).

The final behaviorist that must be discussed is B. F. Skinner, whose approach to science and research became known as behavior analysis. The guiding philosophy behind behavior analysis is radical behaviorism (Baum, 2011; Moore, 2011). This branch of behaviorism was initially coined by Calkins (1921) to describe Watson's approach to behaviorism. This term later became associated with Skinner (1945) as he described his own ideas about operationism and psychology. In radical behaviorism, all behavior can be explained by natural occurrences within the environment (Baum, 2011). Subjective terms that explain behavior are therefore inadequate for describing any event, and they are often categorized as primitive verbal constructs (Schneider & Morris, 1987). The ideas found within radical behaviorism and Thorndike's law of effect were later utilized by Skinner (1938) to create the theory of operant conditioning. Like Thorndike, he included the idea of the S-R behavior framework. He, however, provided additional information about the types of consequences that could occur in connection with different behaviors. The idea of operant conditioning is often represented as S-R-S in order to include the use of consequences in association with S-R behavior pairs (Goodman, 2010; Tuckman & Monetti, 2011).

Operant Conditioning and Consequences for Tokens

The theory of operant conditioning has four types of consequences that are relevant to this study: positive reinforcement, negative reinforcement, positive punishment, and negative punishment. Token economies are known as a type of

reinforcement system for appropriate behavior; however, it is important to mention that tangible stimuli can be reinforcers or punishers depending on how individuals respond to them (Goodman, 2010; Maag, 2001; Skinner, 1953). Positive reinforcement occurs when a behavior is increased after an external item is presented. Tokens can be positive reinforcers if students increase appropriate behaviors after receiving them. Negative reinforcement occurs when a behavior is increased after the removal of an external item. The removal of tokens would only occur during a response cost system, where teachers would remove tokens that had been previously received (Tuckman & Monetti, 2011; Young-Welch, 2008). Positive punishment happens when there is a decrease in behavior after an external stimulus is given to the student. This is possible within a token economy because an appropriate behavior could decrease if students do not like the token that they receive. Negative punishment would not be a consequence for a token economy system, but it would be a consequence for the response cost system. Negative punishment involves a decrease in behavior after an external item is removed. Teachers can use response cost and token economies during instruction in order to observe a variety of student responses. Truchlicka et al. (1998), for instance, used a token system and a response cost system within their study. Points were awarded when students scored an 85% or higher on daily spelling exams. Points and playtime were taken away when students scored less than an 85% on the tests. They found that greater spelling accuracy was demonstrated by the students when the systems were in place.

The effects of tokens are dependent on whether or not they are directly tied to performance and behavior (Cerasoli, Nicklin, & Ford, 2014; Cross, 1981). When tokens are based on how students perform or behave in the classroom, they are contingency

incentives. Tokens that are presented regardless of performance or behavior are noncontingency incentives. Swain and McLaughlin (1998) conducted a study in a classroom where a points-based token system and response cost system were already established. They added a bonus point contingency system to the design. Bonus points were awarded according to problem completion, neatness, and accuracy on math problems. Results showed that math accuracy increased when the bonus points system was in place. Hansen and Lignugaris/Kraft (2005) found that a group contingency plan could be used to increase the amount of positive verbal interactions within the classroom. In their meta-analysis of motivators and performance, Cerasoli et al. (2014) found that contingent incentives were better predictors of performance than noncontingent incentives.

Any reinforcer used within operant conditioning is a contingency incentive (Schunk, 2012; Skinner, 1953). In terms of contingency incentives, there are 4 types of contingency programs for reinforcement: individual programs, independent group-oriented contingencies, dependent group-oriented contingencies, and interdependent group-oriented contingencies (Lynch, Theodore, Bray, & Kehle, 2009; Skinner et al., 2004). Individual programs are those where 1 person receives reinforcements that are contingent on their behavior. In independent group-oriented contingencies, students receive reinforcements according to their individual needs. Independent programs would require at least 2 individuals who could receive reinforcements (Skinner et al., 2004). Dependent group-oriented contingencies are reinforcement programs where reinforcement for all students is based on the performance of one student or a small group of students. Interdependent group-oriented contingencies supply reinforcements that are

based on the performance of all members of a group (Popkin & Skinner, 2003). Skinner et al. (2004) recommended the use of interdependent group-oriented contingencies, but they did mention that teachers are more likely to use individual programs and independent group-oriented contingencies in the school setting. They additionally found that token economies are frequently used within individual programs or independent group-oriented contingencies, but they can be utilized within interdependent group-oriented contingencies to reinforce performance based on the behavior of all students. Stage and Quiroz (1997), however, had a different perspective concerning token economies. Within their study, they categorized token economies and response cost separately from group-oriented contingencies. They stated that the difference between the 2 was that tokens and response cost usually did not have parameters concerning group behavior. Their findings showed that incentive systems that used group-oriented contingencies were more likely to decrease disruptive behaviors in the classroom.

Considerations for Token Design

Token Systems and Behavior Types

The type of token design may be dependent on the behavior characteristics of the students. Wulfert et al. (2002) found that high school and middle school students who exhibited problem behaviors preferred to receive instant gratification when concerning tokens. In the study, problem behaviors included inappropriate behavior in class, substance abuse, a poor self-concept, low performance, and low self-regulation. Students who were not known for exhibiting problem behaviors preferred to receive money using a delayed schedule. During their experiments on multiple ratio and single ratio schedules, Lovitt and Esveldt (1970) found that a student with behavioral disorders

preferred a reinforcement schedule that contained multiple rate contingencies within a points system. Math performance and classroom participation increased when the reinforcement schedule included multiple ratios. Another important study was conducted by Yager (2008) on the relationship between the use of token programs, the use of praise programs, and the number of behavior referrals received by sixth-grade students within the Mississippi public school system. A survey was given to administrators about token use as well as the type of feedback schedules that were utilized during the year. The use of tokens and praise tended to decrease the amount of referrals for students if they were given on a weekly, monthly, or yearly basis. The behaviors of special education students, at-risk students, and students from low-income families were more likely to improve when tokens and praise were used in the classrooms (Yager, 2008).

Studies associated with tokens frequently employ different designs for different problem behaviors. For instance, Hansen and Lignugaris/Kraft (2005) utilized a token system where nine students with emotional disturbances were given tally marks when they gave each other verbal praise. When the group received four tallies for positive statements, each student received a prize from a bag. The results of the study indicated that the amount of positive verbal statements that students gave to each other increased over time, but the amount of negative statements decreased. Anderson, McLaughlin, Derby, and Williams (2012) created a token system designed specifically for students with learning disabilities. The system was used to determine if the amount of improper verbalizations during class could be decreased over time. Improper verbalizations included yelling, interrupting without permission, and swearing. Students would receive free time to play basketball at the end of each day when improper verbalizations

decreased and when they completed schoolwork. Along with this, students received praise during correct verbalizations, and they received a verbal correction when exhibiting improper verbalizations. Results showed that the students did decrease their improper verbalizations when the system was in place. Sansosti (2012) employed the case study method to examine whether a personalized token system could help reduce the threatening and aggressive behaviors of a seventh-grade student with Asperger's syndrome. The student used an index card that contained directions about how to behave appropriately. Points were given to the student when he used the card in class. These points were later exchanged for prizes. The findings indicated that the number of threatening and aggressive behaviors decreased during the intervention.

Token Use and Behavior Modification

Behavior modification is an important concept that pertains to operant conditioning. Token economy systems have been utilized within behavior modification studies (Novak & Hammond, 1983). Tokens are frequently utilized for special education students (Maggin et al., 2011). In a study conducted by Alberto, Heflin, and Andrews (2002), two students with moderate intellectual disabilities wore wristbands, known as timeout ribbons, when they were not disruptive during instruction. The wristbands were removed when inappropriate behaviors were exhibited, and students had to sit in timeout until the wristbands were returned. Students would receive tokens every 5 minutes that the wristband was worn, and they could later exchange the tokens for various prizes. Results showed that there were significant decreases in self-touching, yelling, inappropriate vocalizations, and uninvited approaching during the use of the timeout ribbon intervention. Habaibeh-Sayegh (2014) also used students with intellectual

disabilities in her study to determine if there were significant differences in behavior and achievement between students who received tangible stimuli and students who received intangible stimuli during token exchanges. Students were given cards that listed goals for each day of class. Points were awarded daily according to whether or not students met behavioral goals. The points were later exchanged for either tangible or intangible stimuli. Results indicated that there were significant differences in behavior between the groups, but no significant differences existed in terms of achievement. Students who received the tangible stimuli exhibited a higher rate of appropriate behaviors when compared to students who received intangible stimuli. Furthermore, Lewis-Lancaster and Reisener (2013) used a token economy along with different interventions to determine the effects of various reading strategies on the reading fluency of a self-contained special education student. They used a points system to encourage the student to improve reading scores and work hard during a reading task. Gummy bears were awarded to students at the end of an intervention session for staying on task. Results showed that fluency improved when the student received unlimited time for repeating readings, immediate feedback, tokens, and phonemic awareness training. Finally, Strahan and Layell (2006) collected interviews, observations, work samples, and test data on a year-long program that encouraged the use of a 2-week token system at the beginning of the school year. The program incorporated four major perspectives of successful teaching that addressed the needs of struggling students: a learner-centered perspective, an assessment-centered perspective, a knowledge-centered perspective, and a community perspective. The program helped students improve in math and reading test scores over time.

Models for Academic and Behavioral Support

Teachers who make decisions about tokens and similar instructional interventions must consider the decision-making models that their schools follow. One decision-making model used in Georgia schools is Response to Intervention (RTI), which involves a multi-tiered decision-making process to determine the academic support that students need (Georgia Department of Education, 2011). According to the Georgia Department of Education (2011), Georgia's RTI framework has four tiers: (a) Tier 1 - Standards-Based Classroom Learning, (b) Tier 2 - Needs-Based Learning, (c) Tier 3 - SST-Driven Learning, and (d) Tier 4 - Specially-Designed Learning. The first tier delivers a universal, standardized support system to all students. Struggling students who cannot get their needs met within Tier 1 are given more focused instruction that may include more small group collaboration or more individual help. They receive Tier 2, and they are monitored over time to see if their needs are met. Those who are still having problems after Tier 2 receive Tier 3, which requires the assistance of the school's Student Support Team (SST). The SST is employed to determine the individual strategies needed for each student. They could recommend that a student receives more interventions within the general education classroom, or they could recommend Tier 4 learning. Tier 4 requires specialized programs for students. Examples of specialized programs include self-contained classrooms, interventions for the Deaf and Hard of Hearing, and gifted education programs.

There are Georgia districts, however, that follow the Positive Behavioral Intervention and Supports (PBIS) model (Georgia Department of Education, 2013). This model also uses the Tiers explained in RTI (Georgia Department of Education, 2011).

The difference is that there is more focus on behavioral interventions and reinforcements associated with behavior. Because token systems can be used to address academic performance issues and behavioral problems, they could be utilized within either model. School-wide token systems would be considered a Tier 1 intervention. Token systems designed for small groups and individuals would be useful within Tiers 2, 3, and 4. An outline of the tiers used for RTI and PBIS is provided in Figure 3.

Basic Principles for Schedules of Reinforcement

Existing studies on tokens systems contain a wide range of different reinforcement schedules (Fisher, Piazza, & Roane, 2011; Maggin et al., 2011; Zeiler, 1977). Research on reinforcement schedules shows that behaviors can change according to the schedule that is employed within a study (Ferster & Skinner, 1957). Continuous schedules of reinforcement are not as effective as partial reinforcement schedules (e.g., fixed ratio, variable ratio, fixed interval, and variable interval) in establishing and maintaining behaviors over time (Ferster & Skinner, 1957; Lovitt & Esveldt, 1970; Zeiler, 1977). Partial reinforcement schedules are likely to have more response variability than continuous schedules, and they are more resistant to extinction (Ferster & Skinner, 1957; Lee, Sturmey, & Fields, 2007). Within the partial reinforcement schedules, Ferster and Skinner (1957) found in their experimental studies on pigeons that ratio schedules typically produce higher response rates than interval schedules. When the number of responses approached or exceeded 300, however, interval schedules maintained a much faster rate of response. In general, their findings indicated that fixed ratio schedules were most effective for teaching new behaviors. Variable ratio schedules

were the most effective option for generating high response rates and for maintaining behaviors over time.

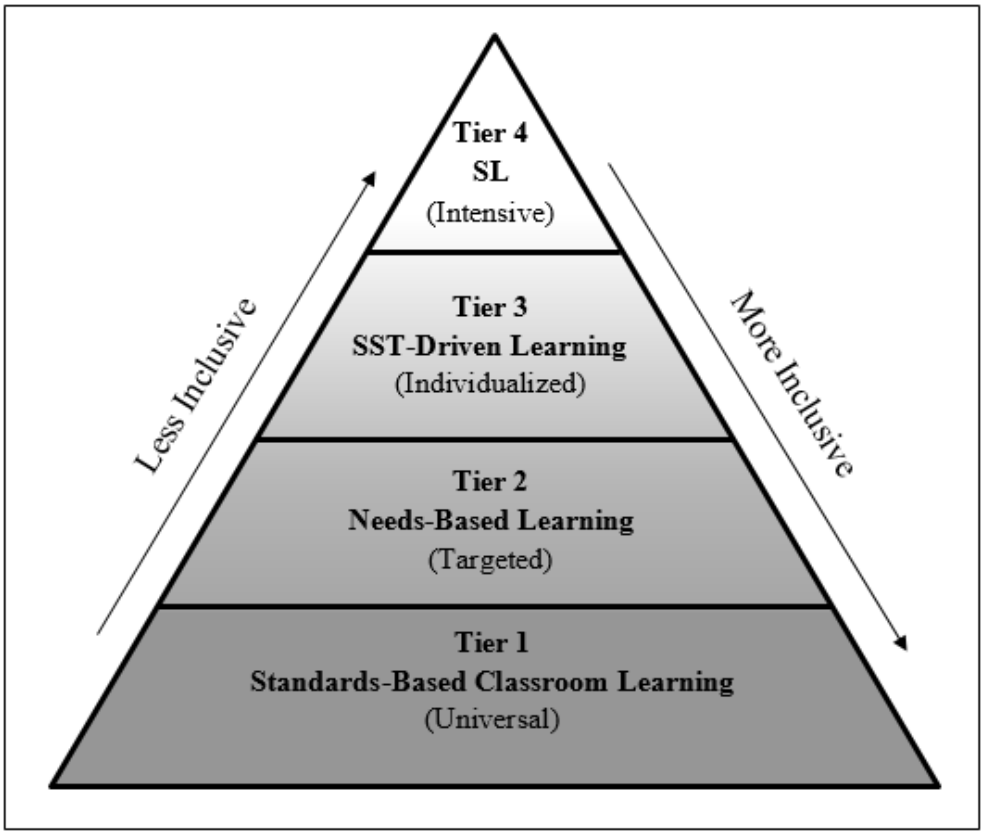


Figure 3. Diagram of RTI and PBIS Tiers. Note: There are four tiers used in Georgia's RTI and PBIS models. As the tier level increases, the number of students who would receive instruction in that tier decreases.

Ferster and Skinner (1957) also saw a break-run pattern when they were observing the behavior of pigeons within fixed ratio and fixed interval schedules. A pause in response was present at the beginning of each reinforcement cycle, but high rates of response would reestablish themselves later in the schedule. They saw little to no pausing in variable ratio and variable interval schedules. Waddell, Leander, Webbe, and Malagodi (1972) had similar results when they looked at fixed interval and fixed ratio schedules with rats. They experimented with the use of tokens on rats where the rats had

to press levers to receive marbles. The marbles were later exchanged for food. The study showed that there was a significant break-run pattern that was apparent within the behavior of the rats. Webbe and Malagodi (1978) conducted a follow-up study with rats that included fixed ratio and variable ratio schedules. They noticed that the rats did exhibit a clear break-run response pattern with fixed ratios, whereas variable ratio schedules produced little to no pausing. Furthermore, Foster, Hackenberg, and Vaidya (2001) compared fixed ratio schedules with variable ratio schedules to determine response patterns in pigeons. Both fixed and variable ratio schedules were able to establish and maintain behaviors, but the variable ratio schedules elicited higher response rates than fixed ratio schedules. With the variable ratio schedule, there was a much shorter time between high response rates.

Not all studies have reached the same conclusions about variable ratio schedules though. Schlinger, Derenne, and Baron (2008) reviewed 50 years of research pertaining to pausing within fixed ratio schedules and variable ratio schedules. Some studies within the literature review did show that variable ratio schedules had shorter pauses than fixed ratio schedules, but there were other studies that demonstrated that variable ratio schedules could produce longer pauses between responses. In the studies that contradicted the findings of Ferster and Skinner (1957), the amount of pausing that occurred within variable ratio schedules depended on the magnitude of the reinforcer and the size of the ratio. Variable ratio schedules that contained a high concentration of reinforcement and a high ratio size produced behaviors that included significantly long pauses between responses. These pauses exceeded the amount of pausing demonstrated with fixed ratio schedules.

Scheduling of Tokens in Education

Tokens require the use of at least two reinforcement schedules, which are called first-order and second-order schedules (Bullock, 2006; Webbe & Malagodi, 1978). The first-order schedule is used so that participants can receive the tokens as reinforcers for behavior. The second-order schedule allows participants to exchange tokens for more desirable reinforcers. Ferster and Skinner (1957) list several possibilities for combination scheduling, including such options as differential reinforcement schedules, concurrent schedules, interlocking schedules, and multiple schedules. Differential reinforcement is a type of schedule that involves reinforcement, extinction, or a combination of the two on specific behaviors (Ferster & Skinner, 1957; Mace, Pratt, Zangrillo, & Steege, 2011). There are different kinds of differential reinforcement schedules. Four types of differential reinforcement are prominently discussed within past studies about reinforcement schedules: differential reinforcement of alternative behavior (DRA), differential reinforcement for low rates of responding (DRL), differential reinforcement for high rates of responding (DRH), and differential reinforcement for other behavior (DRO) (Stage & Quiroz, 1997; Zeiler, 1977). A DRA reinforcement schedule involves reinforcing an alternative behavior and ignoring an inappropriate behavior. The purpose of the DRA schedule is to have higher response rates for the alternative behavior and lower response rates for the inappropriate behavior (Mace et al., 2011). Within a DRL schedule, reinforcement for the target behavior only occurs after the response rate has been lowered to a predetermined rate. DRH schedules also have a specified rate, but reinforcement occurs when the response rate has met or exceeded the goal in mind. DRL and DRH schedules can be used in instances where teachers would want students to

follow a specified criterion for behavior (Hersen & Gross, 2008). DRO schedules are utilized in order to ignore an inappropriate behavior and to reinforce any appropriate behaviors that are exhibited by students. This schedule would help to significantly lower the rate of inappropriate behavior while increasing the rate of desired behaviors (Mace et al., 2011).

There are studies that specifically address the use of differential reinforcement schedules within the field of education. Stage and Quiroz (1997) conducted a meta-analysis of 99 studies in order to determine interventions that would be able to reduce the amount of disruptive behaviors in public schools. There were 16 categories of interventions for the meta-analysis: (a) token reinforcement, (b) differential reinforcement, (c) response cost, (d) group contingencies, (e) teacher behavior, (f) peer management, (g) exercise programs, (h) multimodal treatments, (i) home-based contingencies, (j) functional assessment, (k) self-management, (l) stimulus cueing, (m) punishment, (n) cognitive-behavioral interventions, (o) parent training, and (p) individual therapy. The differential reinforcement category included studies that used DRA, DRL, and DRO reinforcement schedules with students. Within the studies that were analyzed, it was found that the students who received one or more of the 16 treatment conditions were less likely to exhibit disruptive behaviors than those who did not receive any of the interventions. Out of all of the categories, however, three interventions were most likely to be successful in reducing disruptive behaviors: differential reinforcement, self-management, and group contingencies. Anderson et al. (2012) used a DRL schedule within their study to see if it would help to lower the amount of improper classroom verbalizations. They gave positive reinforcement only if students lowered the amount of

verbalizations to a designated number. This would require that a decrease in the rate of incorrect verbalizations would be ignored unless it was less than or equal to that specified number. Negative reinforcement was given when students made improper verbalizations. They found that a differential reinforcement schedule did help to decrease the amount of inappropriate verbalizations in the classroom. Hagopian, Kuhn, and Strother (2009) also used a DRL schedule to address improper social behaviors of a student with Pervasive Developmental Disorders (PDD). The student was reinforced when inappropriate comments were reduced by 85%. The decrease in behavior was ignored until the student's responses reached the appropriate rate. Like Anderson et al. (2012), corrective feedback was a part of the study, and it was given each time the student made inappropriate comments in class. Over time, this scheduling did help to reduce the amount of inappropriate comments that the student displayed on average. Finally, Truchlicka et al. (1998) had a DRH schedule within their points-based token system. Students would receive points after answering at least 85% of the questions correctly from their spelling exams.

Concurrent schedules are a second option where two or more schedules are available at the same time (Morgan, 2010; Taylor, 2000). Morgan (2010) emphasized that concurrent schedules are useful for looking at choice within behavior analysis. He mentioned that studies with concurrent scheduling often have two variable interval schedules that occur simultaneously. Borrero et al. (2010), however, had an experiment with four conditions that included three concurrent schedules and one DRO schedule. The schedules addressed the inappropriate and appropriate behaviors of three children. The four conditions were as follows: (a) two concurrent schedules where problem

behaviors were reinforced at a higher rate than appropriate behaviors, (b) two concurrent schedules where inappropriate and appropriate behaviors were reinforced at the same rate, (c) two concurrent schedules where appropriate behaviors were reinforced at a higher rate than problem behaviors, and (d) one differential reinforcement schedule where reinforcement for problem behaviors was extinguished, and appropriate behaviors were continuously reinforced. It was found that the first child preferred the first scheduling condition when trying to escape from instructional demands, but he preferred the third condition when given tangible reinforcers for appropriate behaviors. The third condition was where he engaged in more appropriate behaviors. The second child engaged in a higher rate of appropriate behaviors only during the second condition. The third student engaged in more appropriate behaviors during the third and fourth conditions, but the fourth condition had a lower extinction rate for those behaviors.

Interlocking schedules are another option within research that uses two or more simple schedules. When participants are reinforced in one schedule, it helps them to progress in other schedules. Lewis-Lancaster and Reisener (2013), for instance, used interlocking schedules for their study. They utilized a continuous reinforcement schedule where a middle school student was awarded a point each time he improved his reading accuracy score. Then, they used a fixed ratio schedule so that the student could receive a token every time he received twenty points. Hansen and Lignugaris/Kraft (2005) followed similar scheduling protocols. They used a continuous reinforcement schedule where points were awarded for each time students gave appropriate verbal praise to others. Later on, they used a fixed ratio schedule where students exchanged their tokens for prizes every time they had accumulated four points.

Multiple schedules can be found within the literature as well. Multiple schedules contain two or more schedules, and the schedules will alternate due to a different context or stimulus. Wulfert et al. (2002) used two fixed interval schedules in their study on middle school students and monetary reinforcement. The schedules were implemented a week apart from each other. Students were able to choose the reinforcement schedule that they wanted to receive. More complex uses of multiple schedules can be seen with Lovitt and Esveldt (1970) and Hoeltzel (1973). Lovitt and Esveldt (1970) conducted three experiments on point-based token systems that involved fixed ratio schedules and variable ratio schedules. The first contained a single fixed ratio schedule that was followed by multiple variable ratio schedules. The second experiment used only a fixed ratio schedule, but the reinforcement frequency was modified over time. The third experiment utilized multiple variable ratios. Hoeltzel (1973) conducted a study that contained fixed ratio and variable ratio schedules in order to determine how they affected reading accuracy over time. A token system was in place where students received galvanized washers for reading accuracy. There were three schedules altogether, and each schedule had a separate phase. The first phase of the experiment had a fixed ratio schedule where a galvanized washer was awarded each time students read 50 words correctly. The second phase had a variable ratio schedule that contained multiple bands. Tokens were received according to the reading rate and accuracy of students. Students would receive a higher amount of tokens as the amount of correct words increased. The reinforcement rate for the tokens, however, decreased as the amount of correct words increased. The third phase had a single variable ratio schedule. The average amount of

tokens for each day of the experiment was calculated, and daily ratios were established that were based on those calculations.

Feedback Combinations

Within scheduling, it is possible to have different feedback combinations. This includes combinations for verbal feedback, symbolic feedback, or tangible feedback. These combinations can be expressed in ratios (Madsen et al., 1970; Parish & Parish, 1991). There are three general feedback combinations that can be seen within research studies: right-blank, wrong-blank, and right-wrong (Barringer & Gholson, 1979). Right-blank involves the awarding of feedback for appropriate actions, but there is no corrective feedback given for inappropriate actions. Wrong-blank involves corrective feedback for inappropriate behaviors, but it does not involve positive reinforcement for appropriate behaviors. Right-wrong feedback occurs when individuals are given positive feedback for appropriate actions and corrective feedback for incorrect actions. According to a literature review by Barringer and Gholson (1979), right-blank is the least effective combination for maintaining behavior over time, whereas wrong-blank is most effective. In contrast, Parish and Parish (1991) have found that right-wrong is more practical for the school setting. They suggest the use of the 5 to 1 rule. The 5 to 1 rule states that for every 5 instances of reinforcing feedback, there should be 1 instance of corrective feedback. This, in turn, creates a safe environment where corrective feedback is not seen as a hindrance to the learning process. Knoster (2014) has emphasized the need for right-wrong combinations for classroom management. He states that teachers need to utilize a 4:1 ratio for current school-wide intervention programs, especially when implementing

PBIS intervention frameworks. This is similar to the 5 to 1 rule, except that 4 instances of positive feedback are given for every one instance of corrective feedback.

Application of Tokens to Goal Theory

Goal Orientation, Behavior, and Achievement

Students can have a combination of different goal orientations (American Psychological Association, 2015; Dowson & McInerney, 2003). These goal orientations are often linked to specific thoughts and behaviors. Ames (1984), for instance, studied 88 students in the fifth and sixth grades to determine if there were differences in how students who had performance and mastery goal orientations thought about their performance on puzzles. Performance goal orientation was called the competitive goal orientation, and mastery goal orientation was known as individualistic goal orientation. Results indicated that students within different goal orientations attributed their academic success to different factors. Students who were in the competitive goal orientation attributed their success to their individual abilities, whereas students within the individualistic orientation attributed their success to the amount of effort they put into the task. It was also noted that those with individualistic orientation displayed more self-sufficiency and self-regulation. Like Ames (1984), a literature review created by Covington (2000) cited multiple studies that indicated that students with a mastery goal orientation were more likely to display self-regulatory behaviors. Anxiety about failure was less likely to be seen. Covington (2000) goes on to state that students with performance goal orientation tended to use more superficial thought processing strategies, and they were more likely to be disorganized in terms of how they attempted to solve problems.

Additionally, motivations can vary according to goal orientation. Baker and Wigfield (1999) found in their study on reading motivation and reading achievement that students were motivated to read because of grades. Grades were classified as performance goals. Students were also likely to be motivated by how important reading was to them. Importance to reading was labeled as a mastery goal. Mongillo (2006) showed in her study on instructional games, cognition, and understanding that it was possible for amusement to be a motivator for learning concepts within middle school. Students in seventh grade showed gains in their vocabulary comprehension and conceptual development when they actively participated in instructional games for science class. It is also possible for goals to be affected by different contexts within the school environment. Anderman and Maehr (1994) pointed out in their literature review on motivation and schooling that policies within the school environment do play a role in what goals students find most important. Eccles et al. (1993) support these results with what they found within their 2-year longitudinal study on the influence of the school environment on mathematics motivation. Results indicated that students who transitioned from elementary to middle school had a harder time being motivated due to more teacher control, less teacher efficacy, more ability grouping, and weaker student/teacher relationships. Maehr (1990) adds to this by saying that school cultural factors associated with accomplishment, power, recognition, and affiliation have a greater influence on motivation as students get older.

Token Use and Goal Creation

Token systems can be used to support goal development. Tokens can be utilized for short-term performance goals (Anderman & Maehr, 1994). If tokens are used for mastery goals, they would need to be administered in a low amount (Haselhuhn et al., 2007; Nelson, 2010). Urdan and Midgley (2003) found in their longitudinal study on goal structure and motivational strategies that the type of motivator that students preferred depended on the goal they set. When they set performance goals, students preferred tangible stimuli. Haselhuhn et al. (2007) added to the research further by stating that strategies used within the classroom depended on what teachers believed about goals. The more teachers believed that their students could set performance and mastery goals, the more likely they were to use a variety of motivational strategies during instruction. Self-Brown and Mathews (2003) found in their quasi-experimental study on the effects of classroom structure and goal development that students who experienced a token system were more likely to set performance goals. Abramovich et al. (2013) indicated that the performance goals for low-performing students and high-performing students changed according to the type of stimuli they received.

Token Use and Motivational Preferences

Research on goal orientation includes extrinsic and intrinsic goals. These orientations can be viewed as exclusive categories that describe how students are motivated within the classroom. McClintic-Gilbert, Corpus, Wormington, and Haimovitz (2013) conducted a survey where they examined the relationship between middle school academic achievement and motivation type. Extrinsic and intrinsic orientations did not overlap within the study, but they were linked to different learning strategies. Extrinsic

motivators were associated with surface learning strategies, whereas intrinsic motivators were associated with deep learning strategies. The study found that academic achievement was negatively correlated with extrinsic motivation and surface learning strategies. There was no relationship between academic achievement, intrinsic motivation, and deep learning strategies. A study by Joosten, Bundy, and Einfeld (2009) was conducted to determine intrinsic and extrinsic motivators for two groups: (a) students who only had intellectual disabilities and (b) students who had a combination of autism and intellectual disabilities. The participants completed a modified version of a questionnaire known as the Motivation Assessment Scale. The results showed that those with just intellectual disabilities were more likely to identify sensory seeking as an intrinsic motivator, whereas those in the combination group were more likely to view anxiety as an intrinsic motivator. In terms of extrinsic motivators, escape and attention were the most common motivators among those in the first group. The ones in the second group, however, were more likely to be motivated by escape and tangible objects.

On the other hand, there are studies about extrinsic and intrinsic orientations that do not treat these orientations as mutually exclusive. Goal orientation and reinforcement preferences can change according to the choices students are allowed to make within token systems (Mucherah & Yoder, 2008; Sheldon & Kasser, 2008). Cerasoli et al. (2014) found in their review that noncontingent incentives were more likely to increase intrinsic motivation than contingent incentives. In their study about monetary incentives and problem behavior, Wulfert et al. (2002) offered students choices related to how much money they could receive. Students could choose to receive an immediate amount of $7 or a delayed amount of $10 that they received at the end of 1 week. Students who had

performed well in school were more likely to choose the $10 option, even though that meant more wait time for monetary reinforcement. Abramovich et al. (2013) conducted survey research about the relationship between badges and motivation. Students in seventh grade and eighth grade received badges during tutorial sessions within the Carnegie Melon Online Computer Science Student Network (CS2N) learning system. There were participation badges and skill badges within the system. Results showed that low-performing students who earned participant badges had lower performance motivation. High-performing students who earned skill badges were more likely to have high performance expectations.

Tokens are often utilized as extrinsic motivators within research (Anderman & Maehr, 1994). Although there are many studies that cite tokens as being suitable for extrinsic motivation, there is debate as to their effects on intrinsic motivation. Cameron and Pierce (1996) conducted a meta-analysis of 100 studies to determine the effects of tangible stimuli on intrinsic motivation. Overall, they found that tokens had positive effects on intrinsic motivation, and any negative effects that were experienced were due to a great amount of restrictions within the design. Akin-Little, Eckert, and Lovett (2004) agreed that the negative effects of tangible stimuli on intrinsic motivation were due to design flaws that did not take into account current research-based reinforcement strategies that were known to be effective. Deci et al. (2001), however, found in their meta-analysis of 128 experiments that tangible stimuli undermined intrinsic motivation. The negative effects of these types of stimuli were more apparent when young children were used in the studies. Benabou and Tirole (2003) were also opponents of token use. They stated that many studies within psychology and sociology have found that the use of

tangible stimuli for performance had negative effects on intrinsic motivation, task perception, and task performance. They further said that the use of punishments was found to have similar effects.

Application of Tokens to Phenomenology

Foundations of Phenomenology

Existential phenomenology explained the qualitative aspect of this study. It is a combination of existentialism and phenomenology. Existentialism is a philosophy about the ability of individuals to construct their own reality. It investigates how a person experiences a phenomenon of interest and how an individual perceives what is experienced (Thompson et al., 1989). A well-known existentialist is Soren Kierkegaard. In his 1846 book, *Concluding Unscientific Postscript to the Philosophical Fragments,* he discussed the nature of truth. He argued that existence could be looked at objectively or subjectively; however, truth can only be found through personal development and subjective thinking. There are three stages of existence within his branch of existentialist philosophy. The first stage is the aesthetic stage, where individuals are interested in enjoyment on an instinctual level. The second stage is the ethical stage, and individuals develop personal commitment to their endeavors. They grow an increasing sense of awareness to the concepts of good, evil, and responsibility. The last stage is the religious stage. Personal commitment is the highest at this stage. There is an awareness that objectivity is uncertain and suffering is required for happiness. The paradoxical nature of this stage is an essential feature of the existentialist view of the truth (Kierkegaard, 1992 version).

Phenomenology is focused on recounting the events of individuals in order to understand lived experiences (Seidman, 2006). Edmund Husserl started phenomenology as a philosophy, and he established what is known as transcendental phenomenology. Within this theory, phenomenology is thought of as a scientific study of consciousness (Husserl, 1931). According to Husserl (1931), the purpose of transcendental phenomenology is to discover the pure essence of a phenomenon. Experience is thought to be separate from facts. In order to find the essence of a phenomenon, all subjectivity must be suspended through a process known as phenomenological reduction. Husserl (1931) refers to this as *epoche*. The essence of a phenomenon is separated from scientific facts and assumptions. This separation allows for the study of what is beyond the scope of experiential reality. Existential phenomenology was later developed as a revision of Husserl's transcendental phenomenology. Founders of this branch of phenomenology include Martin Heidegger and John-Paul Sartre (Machado, 2008; Patton, 2002). The theory itself pertains to how a phenomenon presents itself within real-world contexts. Heidegger's (1962) version of existential phenomenology was able to focus phenomenology within an ontological perspective. He posited that phenomenology should be utilized to understand being and different kinds of being. He thought that it was important to determine *beingness* as it is presented within the real world (Heidegger, 1962). Environmental factors such as time, space, and context would be analyzed in order to study different modes of being. Heidegger (1962) created the concept of *Dasein* to explain how human consciousness can understand itself. *Dasein* is a subjective being that changes its perspective as it experiences the real world. There are existential modes of being that are authentic to *Dasein*, and there are categorical modes of being that

conceal its true nature. Both can be used within the real world in order to experience different environments. Within this view of existential phenomenology, experiences themselves are shaped in a variety of ways where they are constructed according to who is participating within the environment (Heidegger, 1962; Thompson, Locander, & Pollio, 1989). Sartre (1943) also created a perspective of existential phenomenology that included concepts about consciousness, but he was more concerned about how consciousness exists and less concerned about how that consciousness interacts within the real world. He was of the view that *beingness* itself was difficult to define within the realm of experience because its existence extends beyond conscious experience. Being is an objective state, but existence depends on a subjective presence within the real world. It is possible to exist in two states of being: being-in-itself and being-for-itself. Being-in-itself is a non-conscious state of matter, whereas being-for-itself is a conscious state that understands the characteristics of non-consciousness. Being-in-itself makes it possible for nothingness to exist. This is contradictory to the idea of being-for-itself, which has the ability to create morality and values based on a constructed existence. Nothingness is an extension of *beingness*, so it is not completely separate from it. Information of participants' past experiences would help to document the being-in-itself. Present experiences and future expectations would be seen as aspects of being-for-itself.

Because reality can be constructed according to individual choices and experiences, it is important to be able to describe study participants in great detail. This includes noting typical behaviors, patterns, and habits that happen as the participants experience a particular phenomenon (Patton, 2002). Within this study, the phenomenon of importance would be token economy systems as they are experienced by students and

teachers within the middle school setting. The reality of the token system experience is dependent upon the values, characteristics, and perspectives of those who have participated in it first-hand. The way in which consciousness expresses itself within the context of token systems would be seen within qualitative data that are collected about experiences within the real world.

Phenomenology and Tokens

There is research on token economies that offer insights about lived experiences with phenomena. Coyle (2013) conducted a phenomenological study with 22 middle school students to determine the lived experiences of individuals who experienced PBIS. Students received tickets for appropriate behaviors during instruction. Data from qualitative interviews showed that students enjoyed receiving tickets from teachers, and they were encouraged to behave appropriately when they saw others receiving them. Wolfe et al. (2003) conducted a study with three special education students to determine how students experienced token economies within cooperative activities. They found that students increased appropriate behaviors over time as they received tokens for groupwork. The tokens made instruction more engaging for students as they cooperated with others. Kazdin (1982) also found within his review of token economy studies that peer involvement and group activities within token systems tended to increase the response rates of research participants. Alberto et al. (2002) utilized a token exchange system for students who participated in Community-Based Instruction (CBI), which required that students receive instruction in real-world settings. The performance tasks would occur in settings where the specific tasks were commonplace, and the token system that was used addressed inappropriate behaviors that were exhibited within the

CBI design. Simon et al. (1982) found that students who had moderate to severe hearing impairments were able to perform well with a points system. Factors within the environment, such as socioeconomic status, gender, classroom rules, and classroom relationships do play a role in how students respond to a token economy system (Kazdin & Bootzin, 1972; Miller & Eller, 1985). Students respond well to tokens that are relevant to what they are trying to learn. In their experiment with reading students, Marinak and Gambrell (2008) found that the participants preferred to receive a book for their efforts rather than the toys that were also available as prizes. The book helped them to practice their reading skills moreso than the other prizes.

Summary

There is existing research available about token use, but the results have not always been consistent. Tokens can help with academic achievement if they are not overused (Goodman, 2010; Hayenga & Corpus, 2010). There are environmental factors and student behaviors that can influence the design of token systems (Cameron & Pierce, 1996). Operant conditioning, goal theory, and phenomenology help to explain how token designs can be used within education today. Operant conditioning includes the use of tokens within reinforcement and punishment systems (Maag, 2001). Goal theory ties token use to classroom performance goals and students' motivational preferences. Phenomenology describes how teachers and students are able to perceive token experiences within the real world. This literature review has discussed concepts and findings that are relevant to middle school curriculum and instruction. Professionals within education can use many of the ideas described within the review to create research-based token systems for general education and special education students.

Chapter III

METHODOLOGY

Introduction

This chapter discusses the methods that were used for this study. It utilized a convergent parallel mixed methods design. Score data from two Nine Weeks Tests were collected for the quantitative strand of the study. Interviews from teachers and students about token system experiences were collected for the qualitative strand. The research design is further discussed within the first section of the chapter. This is followed by sections that describe the participants, instruments, data collection procedures, data analysis procedures, and limitations for the study. The major aspects of the study are summarized within the last section. The purpose of this study was to determine if any significant differences in math achievement existed between students who receive tokens and students who do not receive them. Additional factors associated with behavior, goals, and motivational preferences were investigated to identify whether or not they have effects on academic achievement. The research questions pertaining to this study were as follows:

1. To what degree, if any, are differences found in Math Nine Weeks Test scores among students who participate in a token program and students who do not?
    a) To what degree are test scores influenced by the timing of token use in the classroom, i.e., students who receive tokens in the first half of the quarter versus students who receive tokens in the second half versus students who do not receive them at all?

b) To what degree are test scores influenced by the type of token used in the classroom, i.e., points-based system versus coin-based system versus no tokens used at all?

2. To what extent are the test scores influenced by (a) classroom behavior referrals, (b) classroom achievement goals, and (c) students' motivational preferences?

3. What are the perceived experiences of (a) students and (b) teachers involved with the use of token-based systems in their classrooms?

Research Design

Design Structure and Worldview

Overview of Mixed Methods Design

A convergent parallel design was used for this study. In a convergent parallel design, the quantitative and qualitative approaches are separate strands that are implemented within one phase (Creswell & Plano Clark, 2011). This design was appropriate for the study because it allowed for ongoing data collection that was based on the practical needs and circumstances surrounding public schools. Real-time data were gathered in ways that maintained the integrity of each approach. It was more practical than sequential or multiphase designs because accurate, qualitative documentation of token experiences was required as the quasi-experiment was implemented. This provided an understanding of participants' experiences without too many academic and developmental changes taking place between the strands. It provided more stability than embedded designs because the researcher was able to collect different types of data without compromising the essential features of qualitative and quantitative designs. Transformative designs were also inappropriate for this study because the use of objective research roles was not advocated within those designs. The convergent parallel

design required different research roles, which were dependent on the circumstances that occurred during the research process.

Worldview Description

A worldview is a set of beliefs that guide action within research (Creswell, 2009; Guba, 1990). Given the focus of this study, it was imperative that the researcher looked at how the data were useful for those in education. There was not a one-size-fits-all method for this. The research questions and the procedures were based on what was viewed as the most practical solution, given the resources and skills that were available at the time. The knowledge generated from this study was used to solve a research problem, and different types of data were essential for understanding the full implications of that problem. Therefore, the worldview that was employed during the study was pragmatism. Pragmatism is utilized within mixed methods research so that researchers can apply what they know in order to address their research needs (Creswell, 2009). This required that multiple avenues were available for exploration. The methods and techniques that were chosen during the research process were appropriate for meeting the instructional needs of teachers and students in Georgia schools.

Structure Description

In the quantitative strand, a counterbalanced format was used to gather Nine Weeks pretest and posttest scores for students who participated in the study. According to Miller (1981), counterbalancing within a research design involves alternating treatment plans among groups in order to control for any possible changes in participants' responses over time. The participants in this study were divided into three groups. Two were treatment groups, and one was a control group. The treatment groups were a

reflection of the common token preferences of those who performed well on achievement tests and those who had low scores on achievement tests. The control group was necessary to establish a baseline of the typical activities that occurred during instruction.

The qualitative strand involved the use of teacher interviews and student focus groups. They provided ongoing feedback that helped to determine token experiences that occurred during the study. They helped to explore current practices with tokens, student behaviors, achievement goals, and motivational preferences. Documents were used as a source of information in order to retrieve data on behavior referrals, classroom achievement goals, and students' motivational preferences.

Population and Sample

Participants

The population for this study was all middle school teachers and students within southwest Georgia. The reasons for choosing this area were as follows:

1. There was a high incidence of Title I schools, which contained students who were more likely to receive behavior referrals according to CRDC data.
2. The school districts in the area aligned curriculum and instruction with the CCGPS.
3. It was the most cost-effective option for the implementation phase.
4. The districts in the area provided a diverse composition of students that allowed for comparisons among different ethnicities.
5. There was a mixture of different geographic territories, which allowed for comparisons among urban and rural settings.

The district for this study, which was referred to as District A, had 85 teachers and 1,569 students (National Center for Education Statistics, 2012). It was classified by the NCES as a town (fringe) territory that contained at least 25,000 people. There were

two middle schools in the district. The majority of District A had Caucasian students. There were also students who identified as African American, Asian, Hispanic, and Multiracial. Both girls and boys attended the school. The school system worked on a block schedule, where each grade level had different instructional blocks for the general education system. Each block was at least 50 minutes in length. There were different subjects taught during the instructional blocks. Special education students were included within the blocks.

Sampling Techniques

Initial Contact

In order to retrieve participants for all parts of the study, purposeful sampling was used. Participant selection was based on two criteria: (a) the instructional program of the school was reflective of what was seen within most Georgia middle schools and (b) the academic performance and behavior referrals received by students was typical of Title I schools. Teachers and administrators were contacted by phone and email so that participants were obtained for all aspects of the implementation phase. For the parties who were interested in assisting with the study, additional written information was given to them through email. This included the dissemination of written summaries of the events that took place.

Quantitative Strand

Sample selection for the quasi-experiment involved purposeful sampling. This was due to the fact that students were already preassigned into existing groups. After permissions to conduct the study were obtained by school officials, teachers, and parents, a list of students within the approved classrooms was acquired. Students were randomly

assigned into the treatment and control groups. An equal amount of students were placed into each group.

The software program G*Power was used to calculate minimum sample size. G*Power is a power analysis tool that contains a variety of procedures for statistical tests. Calculations in G*Power indicated that a minimum sample size of 187 participants was required for statistical credibility. This was the largest value obtained from the power analyses that were completed, and the factorial ANOVA was the test that produced this value. The number was based on a medium effect size (f = .25), a .7 power level, an alpha level of .05, a numerator df of 6, and a minimum of 48 category combinations for the independent variables.

Because the minimum sample size was known, estimates were made about the number of students required for the treatment and control groups. Table 2 provides information about the number of students that would be available, given specific scenarios within the educational setting. Approximately nine classrooms were needed for the study, which would require a sample size of 189 students. This number exceeds the minimum amount of students required for the study. Each classroom would need to have at least 21 students. The actual classrooms that were used for the study contained at least 28 students per classroom, and there were ten classrooms altogether.

Qualitative Strand

Maximum variation sampling was the specific type of purposeful sampling that was utilized for the qualitative aspect of the study. Seidman (2006) recommended maximum variation sampling for phenomenological interviewing so that a wide range of experiences can be documented. This arrangement helped to collect data in such a way

that allowed for a variety of comparisons based on diverse student characteristics. All teachers and students who participated in the study discussed their experiences through interviews.

Table 2

Scenarios for Sample Size

Grade Levels	Classrooms	Students (Treatment)	Students (Control)	Total
1	1 (1 each grade)	14	7	21
2	2 (1 each grade)	28	14	42
3	3 (1 each grade)	42	21	63
3	6 (2 each grade)	84	42	126
3	9 (3 each grade)	126	63	189

Note. For these scenarios, each classroom has 21 students.

There were 3 focus group interviews available for students and 3 individual interviews available for teachers. Each focus group represented a particular grade level. There were 10 teachers who consented to participate in the study, but there were 3 math teachers who actually implemented the token systems. The other teachers were there to support the math teachers during the study. The math teachers were the ones who participated in the individual interviews.

Within the student focus groups, there were students who represented all groups that were a part of this study. A minimum of 6 students were selected for each student focus group based on their achievement and behavior. There were students who reflected above average academic achievement, average academic achievement, below average academic achievement, a high frequency of referrals received, a low frequency of

referrals received, and no amount of referrals received. The ideal number of participants for noncommercial focus groups is from 5 to 8 members; however, the actual number depends on the purpose of the study and the main characteristics required for each focus group (Krueger & Casey, 2015).

Data Collection

Quantitative Strand

Teacher Training

Teacher consent, parental consent, and child assent forms were required for the study (see Appendices E, F, and G for details about these documents). For Institutional Review Board (IRB) purposes, readability statistics were calculated for the child assent forms, where Flesch Reading Ease = 82.8 and Flesch-Kincaid Grade Level = 5.2. Before the start of the quasi-experiment, teacher training was required for the intervention. A 1-hour training session was required for the educators who utilized the token systems for their classrooms. Information about implementation procedures was placed within a PowerPoint presentation, and teachers received handouts of the information during the session. Teachers who had questions about what was covered during the presentation were able to ask them at the end of the session. They were able to communicate with the researcher through phone and email in order to receive ongoing support about the information that was disseminated. The training session answered the following questions:

- Why are tokens important to consider as academic and behavioral supports?
- What are the benefits for using this in teaching?
- How long does this study last?

- How much time will this take from instruction?
- How hard is this to implement?
- What types of tokens are used for the study?
- What schedules are required for the study?
- How flexible is the design for instruction?
- Where are the behavioral supports placed within the PBIS and RTI frameworks?
- Where can educators get more information about the use of tokens in the classroom?

Because teachers may have difficulties understanding the differences in scheduling, real-world examples of the relevant scheduling types were discussed as well. The reinforcement schedules that were utilized within this study were two fixed ratio schedules and one differential reinforcement for high rates of responding (DRH) schedule. An example of a fixed ratio schedule would be awarding students a coupon each time they read three books. This would be a 3:1 ratio, and it would be labeled as an FR 3 schedule. DRH schedules are often used with grading criteria. For instance, students could receive five dollars for getting a score of 80 or above on a test. For the schedule to be in effect, the amount of correct responses would need to be high.

Intervention Design and Procedures

It took approximately one marking period (i.e., 9 weeks) to implement the program for the quasi-experiment (see Appendix A for more details on data collection). The treatment plan's design was based on the procedures implemented by Miller (1981), Truchlicka et al. (1998), Wulfert et al. (2002), and Knoster (2014) that fitted within the tenets of operant conditioning and goal theory. Miller (1981) implemented a

counterbalanced design framework that involved two treatment groups and one control group. This framework was used during the study. Like Miller (1981), a monetary system was included within the intervention. The praise system within that framework was replaced with a points system. Both token systems were able to address a wider variety of behaviors concerning operant conditioning, and the updated design was a more cost-effective design compared to the one that used praise. Points were easy to convert to coins when making transitions from one system to another. For example, four points converted to four coins, and vice versa. The control group did not receive these interventions, so conversion was not necessary for the participants in that group. Figure 4 summarizes the design for the study.

At the beginning of the quasi-experimental strand, a goal orientation questionnaire was administered to students (see Appendix B for details about the questionnaire items). Scores were obtained from a set of Math Nine Weeks Tests, and these scores served as pretest scores. This scoring method was adapted from Wulfert et al. (2002). Since the Nine Weeks pretests were administered at the end of the second 9 weeks, the earliest time the program was implemented was at the start of the third quarter. After the pretest, the experimental groups received different treatments. One experimental group received the points system treatment, and the other received the coins system. The token systems were based on the scheduling procedures administered by Truchlicka et al. (1998). The ratios created within this study, however, were based on the 4:1 feedback ratio described by Knoster (2014). The reason for choosing Knoster's (2014) ratio is because it is used for PBIS.

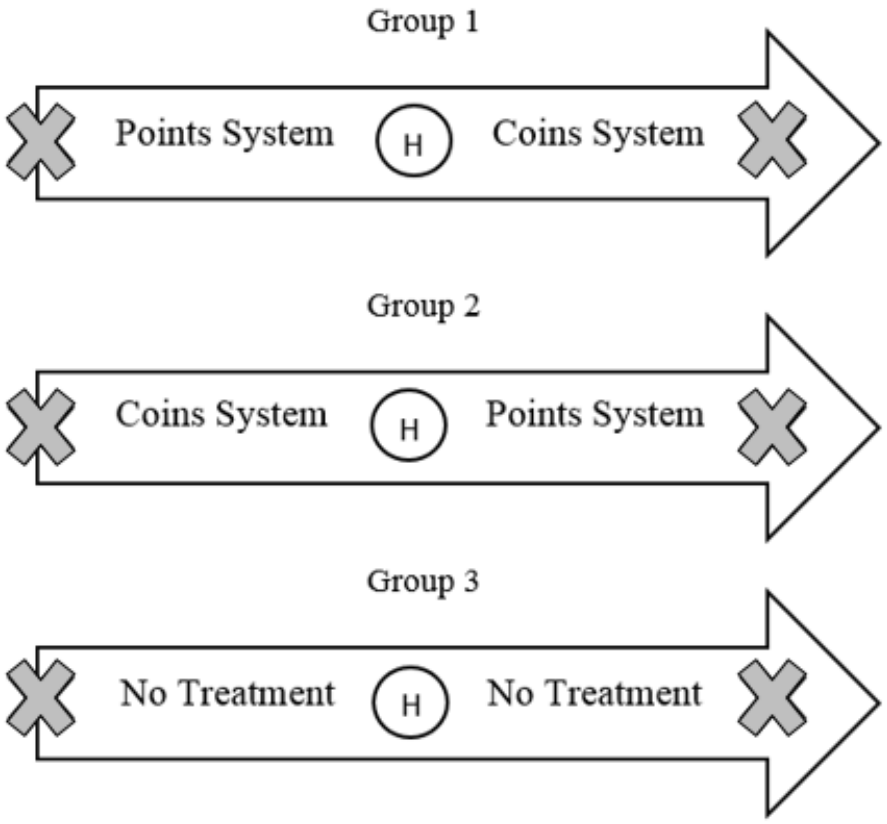


Figure 4. Schematic of Intervention Plan Procedures. Note: The study contained 3 groups, 2 treatment groups and 1 control group. The gray X areas represent when the Nine Weeks Tests were administered, and the circled areas represent the halfway point for the quasi-experiment.

Operant conditioning requires that token systems contain 2 schedules at minimum. There were 3 schedules of reinforcement that were employed for both token systems. The schedules are described in Figure 5. All schedules were based on the tenets of operant conditioning and goal theory. Once students were awarded tokens, they were not taken away as punishment. The research questions focused on token reinforcement systems, which were utilized to reinforce positive behaviors and ignore inappropriate behaviors. Punishment would require response cost or other contingency systems that are not the focus of this study.

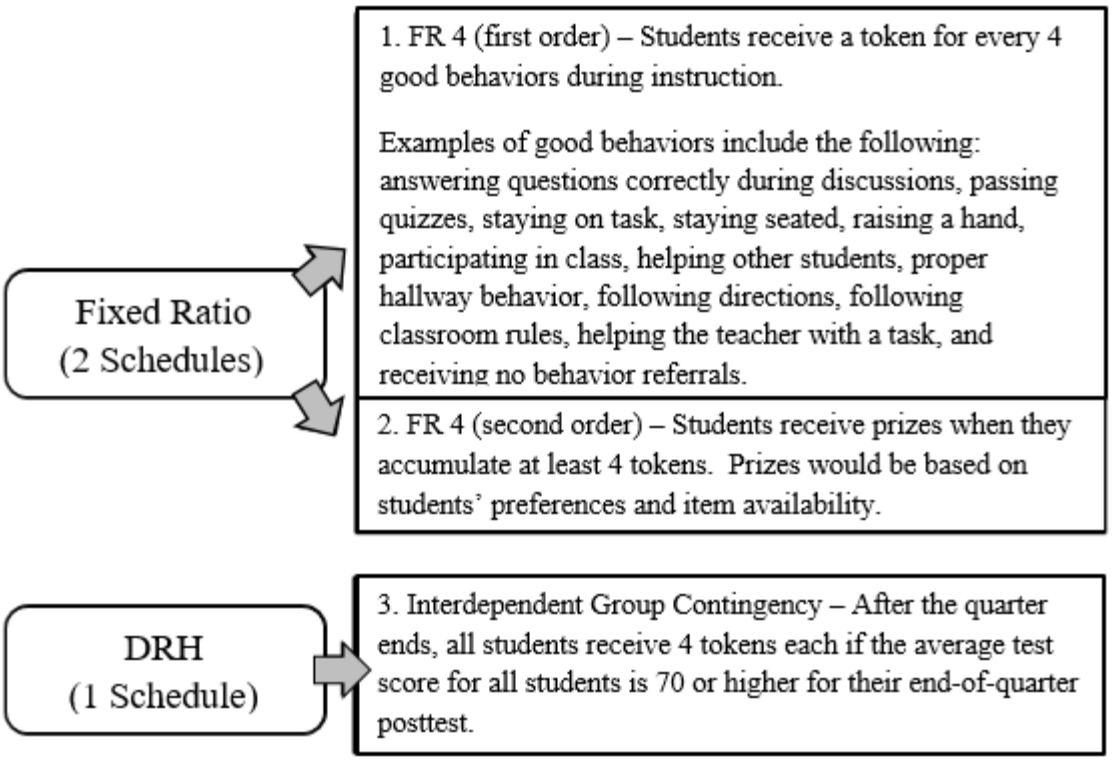


Figure 5. Schedules of Reinforcement for Study. Note: There are 3 schedules of reinforcement that were used for the points system and the coins system.

Schedules 1 and 2 were used to establish and maintain appropriate behaviors in class. There was 1 first-order schedule for token reinforcement, and there was 1 second-order schedule so that students could exchange tokens for prizes. The criteria for these reinforcements were performance-based goals for individual behavior. Schedule 2 prizes were based on what students liked to have for prizes as well as the affordability of the items requested. Examples of affordable prizes included school supplies, toys, and candy. Ideas for prizes were further discussed by Truchlicka et al. (1998). They used early dismissal, game time, library visits, and free time as prizes. Before the token systems were implemented, students were asked to choose the top 5 prizes they would like to see within the token systems. Their choices came from a list of 10 items that the teacher created. The 5 with the most votes became a part of the token exchange systems

(see Appendix I for the prize choices). Prize ratios were based on Knoster's (2014) 4:1 ratio. The values were as follows: 4 tokens – Prize 5; 8 tokens – Prize 4; 12 tokens – Prize 3; 16 tokens – Prize 2; 20 tokens – Prize 1.

Schedule 3 was a DRH schedule that has group-based contingencies. The passing rate for all tests was usually 70% in order to meet CCGPS standards within math. This was why a score of at least 70 is required within the criteria. Schedule 3 was in effect after the Nine Weeks posttests were scored at the end of the third quarter. All students received tokens according to how well the class did on average. If the average score for the posttest was at least 70 for the whole class, then all students in the treatment groups were awarded four tokens. The criterion for this schedule involved a mastery goal since the Nine Weeks posttest served as a summative assessment for the third quarter. There were written records created of all tokens awarded during the study. The document required for these records is provided in Appendix C. The teachers and the researcher kept track of all points and coins obtained by the students. Reliability checks were performed to demonstrate consistency within the systems.

After approximately 4 weeks, the treatment group that had the points system received the coins system. The second treatment group received the points system. A Nine Weeks posttest was administered at the end of the third quarter. At the end of the quasi-experiment, students in all groups filled out a check sheet to indicate whether they preferred a points system, a coins system, both, or neither for motivation (see Appendix D for sample check sheet). They completed the goal orientation questionnaire for a second time. All information about goals and motivational preferences was turned in to

the researcher. This information was entered into Statistical Package for the Social Sciences (SPSS) software and *R* statistical software.

Qualitative Strand

While the quasi-experimental strand was taking place, interviews were conducted for 18 students and 3 teachers. Criteria for selection were discussed in the Sampling Techniques section of this chapter. Each participant in this strand completed a three-part interview series, based on Seidman's (2006) phenomenological interviewing procedures. Interview guides were created for each part in the series. The first part collected information about past experiences with tokens before the intervention. The second part detailed classroom experiences with the intervention. The third part determined what meaning participants place on the concept of tokens. The interviews lasted approximately 60 minutes. Seidman (2006) typically recommends 90 minutes for each interview, but he also states that the length of time can be modified depending upon the age and characteristics of the participants. Sixty minutes was a length of time that was most practical for the time constraints of the schools and the ages of the participants. The date of each interview depended on the availability of the participants. They were all conducted within the school setting in a classroom environment where students and teachers were most comfortable. The interviews were at least 3 weeks apart to ensure that ongoing qualitative feedback about tokens occurred around the same time that the quasi-experiment took place. An audio recording was made of each interview, and written notes were created during the interview that documented what was observed by the researcher.

Variables and Instrumentation

Variables and Factors

For the quasi-experimental strand, there were six variables of concern. There were four independent variables that were measured on the nominal scale. Token media use was the first variable. It contained three values: 0 = Group A, 1 = Group B, and 2 = Group C. Students in Group A used the points system during the first half of the program and used the coins system treatment during the second half of the intervention. Students in Group B used the coins system during the first half of intervention program, and they used the points system during the second half. Group C received neither treatment. The second variable was classroom behavior. It contained three values: 0 = No Referrals, 1 = Low Frequency of Referrals, and 2 = High Frequency of Referrals. The first value indicated that students received zero referrals. Students who had a low frequency of referrals received no more than two referrals in a month. Students who had a high frequency of referrals received at least three referrals in a month. The third variable was achievement goals. Two values were used: 0 = Did Not Meet Goals for Math and 1 = Did Meet Goals for Math. The first value indicated that, on average, students received less than a 70 on the Nine Weeks pretest and posttest. The second value indicated that students received a score of at least 70 on those tests. The fourth variable was students' motivational preferences, which were based on a preferences check sheet that students filled out at the end of the study. There were four values for this variable: 0 = Points, 1 = Coins, 2 = Both, and 3 = Neither. These values accounted for all motivational preferences that were associated with the treatment plan.

There are two dependent variables, which were academic achievement and classroom referrals. Academic achievement was defined as a score of 70 or higher for the Math Nine Weeks pretest and posttest. It was measured on an interval scale. An average score of 0 did not indicate a lack of achievement, but it did indicate that an average of 0 items were correctly answered on those tests. The second dependent variable was classroom behavior referrals, which was defined as the number of referrals received by students during instruction. It was measured on a ratio scale. In this case, the number 0 indicated that the student had not received a referral.

The qualitative interviews required an inductive process to determine additional factors that could influence the implementation of a token economy. The interviews gave teachers and students a chance to share their thoughts, feelings, and perspectives about token economies. Feedback received during the interview process helped to determine patterns that emerged within the experiences that were documented. There were instances where the patterns discovered about teachers' experiences differed from those found within students' experiences. These differences between groups were documented within written notes about the interviews.

Quantitative Instrument

The Nine Weeks pretest and the Nine Weeks posttest were designed to measure if students met the CCGPS standards found within the curricular framework of Georgia's schools. They focused on math standards within the curriculum, and they aligned with content standards for the CCGPS. Tests like these were often required at the end of each marking period. They were different for each grade level. They usually contained multiple-choice items, but some of the tests did have two to four short answer items.

This study focused on the end-of-quarter tests during the second and third 9 weeks. A passing score for the Nine Weeks pretest and posttest was a 70. This means that students who passed these tests answered 70% of the items correctly. Students who scored less than a 70% failed the tests. The tests were made and developed by teachers at the school level. There were different tests made for every marking period. Validity and reliability were established for the tests when they were created. Features of the tests were documented by the researcher to establish face validity and item consistency.

Qualitative Instrument

Seidman's (2006) three-series interview design requires semi-structured interview guides. There was 1 guide for each interview. The guides followed Patton's (2002) suggestions for structuring interview guides. They contained open-ended questions for teachers and students, and there were word prompts under the questions in order to help the interviewer focus on topics that were relevant to the study. The actual interview guides for this study can be seen in Appendix H.

Seidman (2006) mentions that phenomenological interviewing requires interviewees to describe past experiences, present experiences, and constructive views about the phenomenon of study. The first interview guide included broad questions related to past experiences with token economies. The second guide discussed recent experiences that were related to the intervention plan. The third guide discussed how teachers and students constructed meaning when discussing token economies. In order to establish validity for the guides, memoing was required. Memos about interview perspectives and experiences helped to address issues related to bias. Teachers and committee members who had knowledge about how to structure interview guides looked

over the content of the guides in order to determine if there were any inaccuracies in language and content. Narrative memos were required for interview procedures in order to establish reliability.

Data Analysis and Interpretation

Quantitative Data Analysis

Screening Process

The data analysis stage followed suggestions made by Mertler and Vannatta (2013). Multiple rounds of analysis were envisioned for this stage. For all rounds of data analysis, data went through a screening stage. Screening was required in order to identify essential features of grouped data. Mertler and Vannatta (2013) described specific procedures for screening:

- Examine missing data for each variable.
- Examine outliers for the quantitative variable(s) within each group.
- Examine normality for the quantitative variable(s) within each group.
- Examine homogeneity of variances between and among groups.

Histograms and stem-and-leaf plots were generated to determine any outliers within the data. The mean, median, mode, range, variance and standard deviation were calculated for the variables. The screening process tested for normality and homogeneity. Normality was tested using the Shapiro-Wilk test. The significance level for this test needed to be greater than .05 ($p > .05$) for the normality assumption to hold. Homogeneity was tested using Levene's test, and the significance level needed to be greater than .05 ($p > .05$) in order to assume equal variances.

First Round

At the start of the quasi-experiment, names, demographic information, achievement scores, and referral information were gathered for all participants. Data were collected about previous end-of-quarter test scores within math. In other words, data for students at this time were accumulated at the start of the quasi-experiment. After the screening process, a between-subjects factorial ANOVA procedure was conducted to determine if any significant differences in achievement scores existed at this point in time. Demographic factors for subgroups (i.e., grade level, gender, and ethnicity) were included along with the independent variables named classroom behavior and achievement goals. Tukey's HSD was used for post-hoc testing. The significance level had to be less than .05 ($p < .05$) to say there was a statistically significant difference between the groups. The variables for token media use and motivational preferences were not used for this procedure. Token media use had relevance within Round 2 analysis since the treatments had not started yet. Because motivational preferences were assessed at the end of the quasi-experiment, the results for that variable were not available until the third round. A second factorial ANOVA procedure was conducted to determine significant differences in the amount of classroom referrals that were received among the groups, and the same post-hoc testing procedures were applied. Line plots for the independent variables and dependent variables were created in order to illustrate when interaction takes place between factors.

Second Round

The data available within this round covered the first half of the quasi-experiment (i.e., 4 weeks). The values included within classroom behavior and achievement goals

did change because of the information that was collected at that time. Because of this, Research Question 2 was partially answered. More conclusive results were available in the next round. All procedures that took place within Round 1 were recalculated to include the new data. Afterwards, a one-way ANOVA was performed within Round 2 to determine significant differences in achievement scores between Group A (Points Treatment), Group B (Coins Treatment), and Group C (Neither). Tukey's HSD was used again for post-hoc analysis.

Third Round

The rest of the quantitative data that were required for the study were collected at this stage. This included data from the motivational preferences check sheet and data for the second half of the quasi-experiment. Research Questions 1a, 1b, and 2 were answered. For Research Questions 1a and 1b, the categories for token media use were revised so that the achievement scores found in Round 3 can be compared with the achievement scores found in Round 2. Therefore, the categories would change to Group A1 (Points for first half), Group A2 (Coins for second half), Group B1 (Coins for first half), Group B2 (Points for second half), Group C1 (Control for first half), and Group C2 (Control for second half). The corresponding achievement scores were entered for all groups. A one-way ANOVA would be used to determine the degree to which achievement scores were affected by the various categories. Assuming that homogeneity was not violated, the *p*-value for this procedure needed to be less than .05 in order for the differences among the groups to be statistically significant. Histograms, boxplots, and normal q-q plots were used to determine the distribution of the scores. Means plots were created in order to compare means in test scores among the groups. Research Question 2

required a factorial ANOVA for three independent variables and one dependent variable. The independent variables were classroom behavior, achievement goals, and motivational preferences. The dependent variable was academic achievement. Significance levels for main effects and interactions were calculated. A validity check was conducted for a few of the goal orientation questionnaire items in order to verify the responses and categories obtained from the information. All quantitative data findings were connected to the theories of operant conditioning and goal theory.

Qualitative Data Analysis

Audio files and notes from all interviews were transcribed into plain text files in Microsoft Word. All text files for the interviews were imported into *R* statistical software for coding. An electronic journal of memos was created to keep track of all interactions that took place with the qualitative data. The journal was created to keep an audit trail during data analysis. A computer password was required to access the journals. The procedures for coding followed Patton's (2002) and Seidman's (2006) recommendations. Patton (2002) and Seidman (2006) recommended that codes are usually created in text or in color. Text codes are short words or phrases that describe key ideas and observations found within the qualitative data. Color codes would be used to highlight where these ideas and observations are located within the interviews. Both were used for the transcripts. Patton (2002) also stated that a list of codes and their definitions should be created for future reference. Research memos were created that stored all codes, definitions, and coding procedures. Furthermore, Seidman (2006) suggested that codes should be handwritten on a hardcopy of the transcripts, and they should be transferred to software files at a later time. A draft of the text and color codes was created on a hard

copy of each file. Text codes were created within *R* software after the codes were drafted, and a color was associated with each text code. All codes were inserted within the relevant passages. The coded data were placed into meaningful units, and recurring themes were summarized from the data. Patton (2002) described the steps that should be taken in order to determine themes within phenomenological interviews:

- Through the epoche process, set aside preconceived ideas about what is said in the text.
- Determine codes and categories through bracketing.
- Identify patterns seen within the text.
- Organize the data into meaningful clusters so that themes are easily located.
- Create a structural description of the overall experiences of the participants.
- Synthesize what is revealed from the textual and structural analyses.

Interview transcripts were read and coded multiple times to demonstrate consistency within the findings. Copies of the transcripts were given to the interviewees for member checking. Any additional feedback was noted within memos about the interview data. The qualitative data were interpreted according to the phenomenological framework that was outlined for the study. The results answered Research Question 3.

Data Mixing

Data mixing occurred within the interpretation stage. In this stage, data were mixed in order to explain the importance of all findings. There is a discussion provided in Chapter 5 about the similarities and differences between each strand. If certain findings for each approach contradict one another, then there is information provided about the contradictions. Interpretations for the findings were connected to the

conceptual framework, which included operant conditioning theory, goal theory, and existential phenomenology.

Research Permission and Ethical Considerations

Before conducting this study, approval was obtained from the IRB at Valdosta State University (see Appendix J). Permissions were also obtained from those who worked within the school system in order for data to be collected for each strand of the study. Accurate details about the procedures of the study were outlined to determine the risks and benefits. Teacher consent forms included the option to receive copies of teacher interview transcripts as well as copies of summarized results. Since most of the participants were adolescents, it was essential to consider their health and development as the study was implemented over time. Parents were informed of the possible risks and side effects concerning the study before the implementation stage, and their consent was required for student participation. They had options to receive copies of focus group transcripts and summarized results. Participation was on a voluntary basis. Any problems or concerns about the study were discussed with the researcher.

One major issue concerning research is the fairness of the treatment plan (American Psychological Association, 2010). Students who do not receive treatment may feel left out when other students do. In order to address this in the study, teachers were able to use positive intervention strategies that were already established within the curriculum design. This included the use of extra credit assignments, cooperative learning activities, computer-based assignments, and engaging games to review important concepts. Another concern is how data will be protected. A password-protected computer was used for all important data files. Students' names were not used when

reporting data results. The intellectual property of others was respected, and proper permissions were obtained to utilize any data that came from other sources.

Limitations

There were 10 limitations that affected how the results were documented and interpreted. The first limitation was that there was a lack of generalizability when it came to the findings. Although purposeful sampling was used for the quantitative strand, the amount of students available for sample selection was dependent upon the amount of access granted by the administrators, teachers, and parents. Maximum variation sampling was used in the qualitative portion of the study, so the results were only applicable to the individuals who decided to participate. In order to minimize threats to generalizability, detailed records were created about the characteristics of the population of interest. This included the use of memos, written field notes, school profiles, archived forms, and spreadsheets in order to document information. This data helped determine the overall generalizability of the study's findings to other populations.

Secondly, the ability to establish content validity and construct validity was limited. The number of teachers, specialists, and committee members who were able to look over the pretest and posttest depended on the amount of clearance given by the teachers who created the tests. Privacy and confidentiality were important concerns during the construction of tests. The school had strict policies as to who was able to help with test development. It was possible, however, to describe important features, constructs, and feedback opportunities within memos. A written summary of the process for developing the pretest and posttest was created for the dissertation in order to document the procedures involved with test development.

Thirdly, there were time constraints that affected when students were able to participate in the study. The middle school classrooms in the study had block scheduling, and the token intervention was not adapted by all teachers within grades 6-8. There were different instructional periods for the subjects that were taught. In order to fully address this threat to validity, the teachers who used the intervention had the same protocols to follow in terms of how they awarded tokens to students. The students only experienced the token systems within math classrooms. Interviews took place within the school environment, but they took place outside of the time periods designated for math instruction. All procedures for the study were documented by the researcher in order to determine consistency among the classrooms.

The fourth limitation concerning this study was that the awarding of tokens during the quantitative strand required teachers to make judgments about who received them. Personal preferences and biases came into play, despite the fact that guidelines were available for all systems that were used. The effects of this limitation were minimized by the use of multiple resources for storing data about the tokens that were given to students. This included having teachers, students, and the researcher keep a log of the tokens that were awarded. The teachers recorded the number of tokens more than once and checked students' token amounts at the beginning and end of each class session.

The fifth limitation pertained to the data found within behavior referrals, which were used to determine students who exhibited problematic behaviors. It was possible that teachers wrote referrals based on their prior experiences with students. There were environmental and social factors that were taken into account when teachers wrote referrals, and there were times when they did not catch instances of problematic behavior.

To minimize this threat, written notes were required so that the reasons for the referrals were sufficiently documented. There was at least 1 interview question during the qualitative strand that addressed how teachers made referral decisions during the intervention plan.

The sixth limitation was that there were some compatibility issues with the token intervention plan and the norms of the school environment. The token systems, which were whole-class systems, were designed with all students in mind. Not all teachers found the use of tokens to be acceptable for their students. This is why the token systems complied with the CCGPS and other policies with regards to teaching and learning. Site selection for the token systems was based on the recommendations of administrators and teachers. Students and teachers participated on a voluntary basis.

The seventh limitation pertained to the fact that the use of praise, grades, and other feedback systems was difficult to control. While the study was mainly concerned about the use of tokens, teachers were required to give verbal and written feedback about students' work. This included the use of praise, grades, and corrective feedback. Praise is often used as a reinforcer, but not all students would like praise. Grades can be used to reinforce students for good work or punish students for unsatisfactory work. Corrective feedback is often used as a punisher, yet students may enjoy the attention that it gives them. Teachers also used cooperative learning strategies and response cost to address behavior in the classroom. These strategies, when utilized, played a role in student learning within the token systems, and the feedback preferences were dependent on teacher preferences. Documenting the reinforcement and punishment procedures before,

during, and after treatment helped to establish important details about the school environment and its protocols.

The eighth limitation concerned the distribution of the interval and ratio variables. A non-normal distribution was found for the pretest, posttest, and referral information. The pretest and posttest scores were negatively skewed. There was a small frequency of students who received referrals each month, meaning that the data was positively skewed for the referral information. Creating memos about school data, quantitative analysis procedures, and normality helped to explain how non-normal distributions could occur within the school setting.

The ninth limitation pertained to the leaking of treatment information among groups. During study implementation, the participants interacted with each other throughout the day. The control groups would interact with the treatment groups during P.E., lunch, class transition periods, special school events, and extracurricular activities that were not part of math instruction. Those who did not receive the token treatments had reactions to the fact that they were not provided with those supports. To address this threat to internal validity, students who did not have the token treatment were provided with alternative supports and rewards that did not conflict with the procedures inherent within the study's design. These alternatives were noted within interview transcripts, IRB documentation, and written memos about the study's procedures.

Finally, perspectives regarding token experiences were dependent upon how students and teachers interpreted the events. There were differences in data interpretation as well. How the researcher interpreted qualitative data was different than how others interpreted the data. The effects of this limitation were mitigated through member checks

and ongoing participant feedback. The interview questions were discussed with participants before, during, and after their interviews. They were able to ask questions if they needed clarification for a particular question, prompt, or concept. Participants provided feedback about findings from their interviews, and they received copies of their individual transcripts. This helped to determine if there were any misunderstandings in terms of how their perspectives were interpreted.

Summary

The Methodology section outlined all of the major components of this mixed methods study. A convergent parallel mixed methods study was conducted to determine if any effects existed between token design and academic achievement. Participants for the quasi-experimental strand included middle school students in southwest Georgia who were enrolled in a Title I public school. Curricular frameworks and instructional strategies within the school environment were aligned with CCGPS. Participants for the qualitative strand included middle school teachers and students who were involved in the quantitative strand. Instruments were created for both approaches. For the quasi-experimental strand, a Nine Weeks pretest and a Nine Weeks posttest were created by the teachers that tested skills in math. For the qualitative strand, the interviews required the creation of interview guides by the researcher. The data analysis process and the interpretation process contained procedures that helped to answer all research questions. The study's design interests those who want to know more about token systems, achievement tests, middle grades education, and mixed methods research.

Chapter IV

RESULTS

Introduction

This study required multiple rounds of analyses for quantitative and qualitative data. Quantitative analyses involved procedures that assessed unit test information, token media information, demographic information, referral data, and goal orientation data. These forms of data were analyzed using recommendations from Mertler and Vannatta (2013). The level of significance for quantitative calculations was .05, and statistically significant items were noted from the procedures. Items that were not statistically significant ($p > .05$) were noted as well. These items were mentioned in order to identify avenues for future research. Qualitative analyses were based on data accumulated from teacher interviews and student focus groups. The interviews were transcribed, coded, and categorized according to recommendations made by Patton (2002) and Seidman (2006).

Organization of Chapter

Demographic data for the sample are presented first in this chapter. This information was obtained from school records. In addition, students described their race and gender on the goal orientation questionnaire. After the demographic data are reported, screening procedures for quantitative analysis are discussed. This is followed by results for the research questions. Major steps in data analysis procedures are outlined in the order in which they are performed. Because quantitative results answer Research Questions 1 and 2, they are presented before the qualitative results. The data for the

research questions were obtained from school records, unit tests, teacher printouts, the prize list, the goal orientation questionnaire, and the motivational preferences check sheet. Qualitative analyses were performed in order to answer Research Question 3 and to provide supporting evidence to the quantitative data used in answering Research Questions 1 and 2.

Demographics

The study was conducted within the sixth, seventh, and eighth grades. There were 208 students who participated at the beginning of the study. This number was reduced to 205 because there were 3 participants who withdrew during its implementation. Out of the 205 participants, there were 57 (27.8% of the sample) who were sixth graders, 79 (38.5% of the sample) who were seventh graders, and 69 (33.7% of the sample) who were eighth graders. The sixth and eighth grades had 3 groups each. The seventh grade had 4 groups of students. More information about the grade levels is provided in Table 3.

Table 3

Demographic Data for Grade Levels

Grade	Students	Groups	Males	Females	SWD	FRPL
Sixth	57	3	50.9%	49.1%	1.8%	36.8%
Seventh	79	4	45.6%	54.4%	1.3%	29.1%
Eighth	69	3	50.7%	49.3%	0.0%	40.6%

Note. SWD = Students With Disabilities; FRPL = Free and Reduced-Price Lunch. Percentages are approximate values.

There were 105 girls and 100 boys who participated in the study. This means that approximately 51.2% were girls, and approximately 48.8% were boys. In Grade 6, there were 28 girls and 29 boys. The seventh grade contained 43 girls and 36 boys, whereas the eighth grade contained 34 girls and 35 boys. The percentages for gender were similar for sixth and eighth grade. Both grades had more males than females. Seventh grade, in contrast, contained more females than males. There was a percentage difference of 8.8% between the gender categories in seventh grade. This was higher than the differences within the other grades, but it was not large enough to have statistical significance within analysis procedures. Discussions about free lunch status and disability status are provided in the Disability Status and ECD Status sections of this chapter.

Ethnicity

There were 5 categories for ethnicity: Asian, Black, Hispanic, Multiracial, and White. These categories were based on data labels obtained from official school records. Overall, the majority of students (n = 147) were identified as White, also known as Caucasian. This was approximately 71.7% of the total sample. The second largest category was Black, also known as African American. Approximately 18.5% of students (n = 38) identified themselves as African American. There were 10 students (4.9%) who were categorized as Asian, 5 (2.4%) who were categorized as Hispanic, and 5 (2.4%) who were categorized as Multiracial. Subgroup percentages were calculated for these categories, and they are presented in Table 4.

Caucasians, African Americans, and Asians make up the majority of students within each of the gender categories. The differences within the gender categories were very small. The widest differences were seen with African American and Multiracial

students. The male category had 3.0% more Multiracial students than the female category, and there were 5.0% more African Americans within the female category. Since the percentage differences were small within these categories, there is a low likelihood that they would have statistical significance during analysis procedures.

Table 4

Subgroup Percentages for Ethnicity by Gender and Grade Level

Subgroup	Asian	Black	Hispanic	Multiracial	White
Gender					
Male	5.0%	16.0%	2.0%	4.0%	73.0%
Female	4.8%	21.0%	2.9%	1.0%	70.5%
Grade Level					
Sixth	8.8%	26.3%	0.0%	5.3%	59.6%
Seventh	5.1%	11.4%	2.5%	0.0%	81.0%
Eighth	1.4%	20.3%	4.3%	2.9%	71.0%

Note. Data includes general education and special education students.

In terms of grade level, Caucasians and African Americans were the predominant groups for each of the 3 categories; Asians, however, were not the third highest group within all categories. Eighth grade had 2.9% more Hispanic students than Asian students. When comparing the percentages within each ethnicity, small gaps were found within the Asian, Hispanic, and Multiracial subgroups. The highest percentage gap found within these groups was 7.4%, which was found between the percentage of Asian students for sixth grade (8.8% of the sample) and the percentage of Asian students for eighth grade (1.4% of the sample). It is likely that these small differences would not be statistically

significant. In addition, there were no Hispanic students found within sixth grade. No students were identified as Multiracial in seventh grade.

Further analysis within ethnicity showed that there were noteworthy gaps found for the Black and White categories. Within the Black category, there was a 14.9% percentage gap between sixth and seventh grade. There was a 21.4% gap between sixth and seventh grade in the White category. One plausible reason for these wide gaps had to do with the fact that sixth grade had the lowest amount of students participate in the study. Seventh grade had the highest amount of students when compared to the other grade levels. These differences found within grade level and ethnicity could be statistically significant within the findings. The degrees of statistical significance for these groups are discussed within the Findings sections of this chapter.

Disability Status

There were two categories for disability status: No and Yes. These categories were not defined by the researcher but are the only categories denoted in the school records. The No category indicated which students did not have a disability designation, also known as SWOD (Students Without Disabilities). The Yes category indicated students who had the SWD (Students With Disabilities) designation. The SWD participants received speech services but did not receive academic-related services. Overall, there were two students who were classified as SWD. There were 203 students who were SWOD. Within the Yes category, there was a girl from seventh grade and a boy from sixth grade. When grouped by ethnicity, the SWD data indicated that the girl was Caucasian and the boy was African American. Given there were only two students designated as SWD, it is highly unlikely this factor will influence future statistical results.

ECD Status

There were 3 categories listed within the school records concerning ECD (economically disadvantaged) status: Free Lunch Status, Reduced Lunch Status, and Paid in Full. Students who had free lunch and reduced lunch status would be considered ECD by the school system. Out of the 205 participants, there were 133 who pay a full-priced lunch. This is approximately 64.9% of the total sample. There were 60 students who had free lunch status, which was approximately 29.3% of participants. Twelve students were eligible for reduced-priced lunch, which was approximately 5.9% of the sample.

Within the grade levels, eighth grade had the most students who received free or reduced-price lunch (n = 28). For the gender subgroups, there were more boys (n = 34) than girls (n = 26) within the free lunch category. More girls received reduced-price lunch (n = 10) than the boys (n = 2). Analysis for the disability status categories showed that there were 59 SWOD students who received free lunch. There were 12 who received reduced-price lunch, and 132 paid in full. One of the SWD students had free lunch status, and the other had paid in full status.

Furthermore, Caucasians (n = 44) and African Americans (n = 22) had more within the free lunch and reduced-priced lunch categories than the other ethnic groups. There were 2 students within each of the remaining categories who had either free lunch or reduced lunch status. In other words, there were 6 students with free or reduced lunch status who were categorized as either Asian, Hispanic, or Multiracial. During the analysis stage, there were ECD percentages calculated by grade level and ethnicity. These percentages are presented in Table 5.

Table 5

Subgroup Percentages for ECD by Grade Level and Ethnicity

Subgroup	Free Lunch	Reduced-Price	Paid in Full
Grade Level			
Sixth	26.3%	10.5%	63.2%
Seventh	26.6%	2.5%	70.9%
Eighth	34.8%	5.8%	59.4%
Ethnicity			
Asian	10.0%	10.0%	80.0%
Black	50.0%	7.9%	42.1%
Hispanic	20.0%	20.0%	60.0%
Multiracial	20.0%	20.0%	60.0%
White	25.9%	4.1%	70.1%

Note. Data includes general education and special education students.

Within the grade level subgroups, there was a consistent trend found where the majority of students were within the Paid in Full category. The second major category for all grade levels was the Free Lunch category. Seventh grade had a highest percentage of students within the Paid in Full category. The highest amount of difference found within this category was 11.5%, which was more than the differences within the other categories. The highest difference within the Reduced-Price category was 8.0%, and the highest difference within the Free Lunch category was 8.5%.

Even though these percentages did not show much of a difference within each category, there was a great amount of difference when comparing ECD across subgroups.

For instance, the smallest gap between the Free Lunch and Paid in Full group was 24.6%. This was found within eighth grade, indicating that these gaps were wider for sixth and seventh grade. The smallest gap between the Paid in Full group and the Reduced-Price group was 52.7%, which was found within sixth grade. These large gaps in the amount of students within each group can affect the amount of variability found within the results for each category.

When comparing ECD across ethnic groups, the percentages for Asians, Hispanics, and Multiracial students would not influence future results because they reflect a small amount of students within the sample (n = 20). In these instances, however, the majority of students were within the Paid in Full group. Caucasian students reflected this trend as well, but the majority of African American students were within the Free Lunch category. The gap between the Free Lunch and Paid in Full category was very small for African Americans at 7.9%, but there was a difference of at least 40 percentage points for the other ethnic groups. These differences within ECD and ethnicity could be influential to the results since a large amount of students are affected.

Another potential area of statistical influence concerns differences in the Reduced-Price category and the Paid in Full category for the ethnicities. The smallest difference was 34.2%, which was seen between African Americans who had Reduced-Price status and African Americans who had Paid in Full status. The largest difference was found within the Asian category at 70%. Large differences could affect the amount of dispersion seen within group means during statistical testing. More information about these differences are found within the Findings sections of this chapter.

Results for Screening

For the beginning of the quantitative analysis stage, multiple rounds of screening occurred. These rounds of screening took place to determine variable descriptors, missing data, outliers, violations in normality, and violations in homogeneity. Data included students' demographic and academic information. There were 3 rounds of screening in total, 1 for each major data collection period. In the first round, data collected before treatment was reviewed in order to determine preliminary results for the pretest scores. The second round involved the use of statistical data collected during the week that the token switch occurred for the treatment groups. The third round included data that were collected after the administration of the posttest. Data for students who did not have consent on file were removed before analyses were performed. Because of the screening processes used, not every member of the 205 participants were used in all analysis procedures. For instance, students who had missing data needed for a given analysis procedure were omitted from that specific procedure; however, they were included in the procedures where their information was available.

Round 1 Screening

Round 1 analyses contained nominal variables for grade level, class, gender, ethnicity, reduced lunch status, disability status, and classroom behavior. There were pretest scores measured on the interval scale, and there was referral information measured on the ratio scale. Referral information is based on referrals given the month before the pretest (i.e., December).

During analyses, missing data were found. There was 1 student in seventh grade who did not complete a pretest, but the participant did have complete information for the

other variables assessed. Therefore, a pairwise deletion was performed when available. Stem-and-leaf plots and boxplots did not show outliers for pretest data overall. The majority of students (n = 202) received no referrals for December. There were 6 students who did receive referrals. There were 208 participants at the time the screening procedures were completed for Round 1. This number changed during screenings for Rounds 2 and 3.

The Shapiro-Wilk normality test and Levene's homogeneity test were later performed. For the pretest scores, the Shapiro-Wilk test found that there was a non-normal distribution that was negatively skewed. A transformation was performed in order to correct the skewed shape of the distribution. Mertler and Vannatta (2013) list data transformation procedures for non-normal distributions. A reflection and square root procedure was performed on the pretest variable to create a new pretest variable. The equation is as follows: PretestNew = SQRT (K – Pretest), where K is the constant that gives the distribution a minimum of 1. The constant used for the equation was 106 because the maximum pretest score was 105. The transformed variable had a normal distribution.

The referral numbers had a non-normal distribution with a severe positive skew, but a transformation was not performed since most students had 0 referrals. There were no violations found within Levene's test. All quantitative variables will be used for parametric testing, even if they violate assumptions in normality and homogeneity. This is due to the fact that ANOVAs are highly robust against these violations as long as group sizes maintain statistical credibility (Mertler & Vannatta, 2013).

Round 2 Screening

The purpose of Round 2 screening was to analyze the data collected at the switch point for the treatment groups, which was approximately 4 weeks after the start of the study. Round 2 screening added nominal variables for token use and goal orientation that were not present in Round 1. Additional nominal variables were created for classroom behavior to account for referral amounts in January. This was done for the ratio variable for referrals as well.

By the time the treatments were switched, 1 student withdrew from the study. All data, including pretest scores, were removed for the student. This brought the total sample to 207. Missing data were found for the goal orientation questionnaire. There were 20 students who did not turn in a questionnaire. This was approximately 9.7% of the sample. The data for all these students were still included within statistical tests, and numbers that pertained to missing data were documented when the questionnaire results were analyzed. Because there was a student whose Round 1 pretest scores were removed, screening tests for the pretest variable were completed for a second time. Similar to Round 1, there were no outliers found for the pretest variable overall. There were thirteen students who received referrals for Round 2. Normality tests were repeated for the pretest and referral variables. Non-normal distributions were found again for pretest scores and referral numbers. Homogeneity of variance was violated within the pretest variable. Transformations were unable to correct the issue.

Round 3 Screening

The purpose of Round 3 screening was to assess data collected after treatment ended for the quasi-experiment. Round 3 screening included nominal variables for

classroom behavior, motivational preferences, and goal orientation that were not available in the previous rounds. Also, there were new dependent variables created for the posttest information that were non-existent in Rounds 1 and 2. During this stage of the study, there were 2 students who withdrew from the study, which brought the sample size to 205. Missing data were found for the motivational preferences check sheets and the goal orientation questionnaires. The other data that were turned in for these students were kept, and their missing numbers were mentioned during analyses with the check sheets and questionnaires. Approximately 163 check sheets (79.5%) were returned, and approximately 171 participants (83.4%) completed questionnaires.

Because there were students who withdrew from the study at this stage, statistical tests for the pretest in Round 1 were redone. No outliers were found for the pretest and posttest information. Statistical tests for normality and homogeneity were performed for the pretest and posttest information. The variables were found to be non-normal. The same transformation equation used for the pretest was used for the posttest variable, except that the constant changed to 109 since the maximum score was 108. The transformed variables showed a normal distribution. Violations for homogeneity were found for the pretest and posttest variables. Transformations did not correct all violations. Specific information about the violations are provided for the findings within Research Questions 1 and 2.

Research Questions

The remainder of the chapter presents the findings associated with the three research questions that are the focus of this study.

Research Question 1: To what degree, if any, are differences found in Math Nine Weeks Test scores among students who participate in a token program and students who do not?

Research Question 1a: To what degree are test scores influenced by the timing of token use in the classroom, i.e., students who receive tokens in the first half of the quarter versus students who receive tokens in the second half versus students who do not receive them at all?

Research Question 1a required the use of parametric tests. Specifically, one-way ANOVAs were used for parametric testing. The one-way ANOVAs were first performed to see if there were statistically significant differences in Nine Weeks Test scores when the points groups were compared to the control groups. Then, one-way ANOVAs were executed to determine if statistically significant differences were found when the coins groups were compared with the control groups. In order to answer this question, four different combinations of the variable known as token media use were created. These combinations are summarized within Table 6.

Table 6

Combinations for Research 1a Analyses

Groups	Points Timing (A1 X B2)	Coins Timing (B1 X A2)
Pretest Points (A1)	A1 X A1 X B2	A1 X B1 X A2
Pretest Coins (B1)	B1 X A1 X B2	B1 X B1 X A2
Pretest Control (C1)	C1 X A1 X B2	C1 X B1 X A2
Posttest Coins (A2)	A2 X A1 X B2	A2 X B1 X A2
Posttest Points (B2)	B2 X A1 X B2	B2 X B1 X A2
Posttest Control (C2)	C2 X A1 X B2	C2 X B1 X A2

Note. Relevant independent combinations for Research Question 1a are in black.

Each independent combination listed in the table was used to examine different aspects of timing within the treatment and control groups. Dependent combinations, which were written in gray, were excluded from analyses because the quasi-experiment

contained independent samples. The first and second combinations required comparisons between the timing of the points interventions for the treatment groups (i.e., A1 and B2) and the timing of instruction for the control groups (i.e., C1 and C2). The third and fourth combinations were used to compare the timing of the coins interventions (i.e., B1 and A2) as well as the timing of instruction for the control groups (i.e., C1 and C2). The pretest and posttest variables were rearranged in order to account for the new groups. For combinations of test scores that violated homogeneity, transformed versions were used. The test score values within the transformed combinations were the same values found in the transformations for the pretest and posttest variables. This was done in order to preserve the numbers within the transformations already performed. The transformed versions of the variables were able to correct violations in homogeneity. Because of this, non-parametric tests were not required for the combinations. Since a reflect and square root transformation was performed on the pretest and posttest, the transformed variables showed a reversal of the actual results. Therefore, data interpretation was based on the results of the untransformed variables (Mertler & Vannatta, 2013). Here is the overall finding based on treatment timing:

> *Finding 1.1: Students in the control group scored significantly higher than students who received tokens at any time during the study.*

In the first and second halves of the 9-week quarter, students who did not receive tokens outperformed those who received tokens. This was observed within pairwise comparisons for the timing combinations. In the token groups, those who were awarded points during the first half of the study scored significantly higher than students who were awarded points after the switch occurred. The coins groups had the opposite result, where students who were given coins in the second half had a significantly higher test

score than those who were given coins in the first half. Specific results for this finding are as follows:

- Result 1.1: Students who received no points for the first half of the quarter scored significantly higher than those who received points at any time during the study.

This result was observed within Treatment Combination 1. It contained three groups: A1, C1, and B2. Group A1 represented students who received points during the first half of the 9-week quarter. Participants in Group C1 were students within the control group who received no treatment for the first half. Students in Group B2 were participants who received points in the second half. Most participants were in C1 (n = 84). There were more students within B2 (n = 67) when compared to those in A1 (n = 54).

Group means for A1 and C1 were derived from data collected for the pretest. The group mean for B2 originated from posttest data. The mean for C1 (M = 77.82, SD = 15.93) was higher than the means for A1 (M = 69.69, SD = 16.73) and B2 (M = 63.63, SD = 21.10). For the one-way ANOVAs, homogeneity was violated within the untransformed version of Treatment Combination 1 (p = .010). This was corrected with the transformed version (p = .270).

Within the group means presented in this section, there was a difference of 6.06 points between Group A1 and Group B2. In other words, students who received points in the first half had a higher test score, on average, than students who received points in the second half. This difference of means was smaller than the difference found between Groups C1 and A1, which was 8.13. Students in C1 for the first half scored higher than students in A1 during their points intervention. A mean difference of 14.19 was found

between Groups C1 and B2. The control group scored higher than those who received points.

Group B2 had the widest spread of scores, which indicated a higher amount of variability than C1 and A1. The highest difference in terms of standard deviation was 5.17. This was found between B2 and C1. There was a wider range of scores for the points intervention during the second half of the quarter. A1 and C1 had a difference of 0.8, showing similar amounts of variability within these groups. The violation of homogeneity within the untransformed version of Treatment Combination 1 indicated an unequal amount of variability within the samples, but there were equal variances found within the transformed version.

The significance values for the one-way ANOVA procedures were less than .001 overall ($F(2, 202) = 11.83$, partial $\eta^2 = .105$). Because the overall results showed that significant differences existed between the groups, post-hoc testing with Tukey's HSD was required in order to clarify which groups showed significant differences in test scores. Within post-hoc testing, there was a p value of .027 between Groups C1 and A1. The control group for the first half of the quarter scored significantly higher on the pretest than those who had points in the first half. The p value was less than .001 between Group C1 and Group B2. This means that another significant difference was found within the results. Those who had the control in the first quarter scored significantly higher than those who had points in the second quarter.

There were different results found within the transformed and untransformed versions when it came to comparisons between Groups A1 and B2. The untransformed version did not display any significant differences within the pairwise comparisons (p =

.158). The transformed version, however, showed that the groups bordered on statistical significance (p = .042). Because the transformed version upholds the assumptions of normality and homogeneity, the significance level for this version would give a more accurate depiction of what occurred between the groups. Those who were awarded points in the first half of the quarter scored significantly higher than those who were awarded points in the second half. Even though a significant difference existed between these two groups, it is important to remember that both groups scored significantly lower than the control group from the first half of the quarter.

- Result 1.2: Students who received no points for the second half of the quarter scored significantly higher than those who received points at any time during the study.

This result pertained to Treatment Combination 2. The groups within this combination were A1, B2, and C2. The meanings for A1 and B2 were discussed within Result 1.1. Group C2 included participants who received no tokens for the second half of the quarter. They were the same participants within Group C1; hence, the amount of participants within C2 (n = 84) exceeded the amount of participants in Groups A1 (n = 54) and B2 (n = 67).

The mean for C2 (M = 81.70, SD = 13.15) was significantly higher than the group means for A1 and B2, which were discussed within the previous results section. Like B2, data for C2 were collected during the posttest. The one-way ANOVA for these groups showed that homogeneity was violated for the untransformed version of the combination (p < .001), but it was upheld for the transformed version. The p value for the transformed version was .052, which was slightly above the required .05 level that was stipulated for Levene's test.

There was a mean difference of 12.01 between Groups C2 and A1. The control group in the second half of the quarter scored higher than the participants who received points in the first half. A larger difference of 18.07 was seen between Groups C2 and B2. C2 scored higher than B2, which indicated that the control group scored higher than those who received points in the second half. When comparing the mean differences with Result 1.1, it was discovered that the group comparisons with the control groups show larger differences within Result 1.2. For instance, the mean difference between C2 and A1 was larger than the difference for C1 and A1. Also, the mean difference between C2 and B2 was larger than the difference for C1 and B2. From this information, it can be concluded that the average posttest score for C1 was higher than the average pretest score for the same group.

A difference of 3.58 was observed between the standard deviations of Groups C2 and A1. A slightly larger difference of 7.95 was seen between Groups C2 and B2. This information indicated that the test scores for the points groups in the first and second quarter showed more variability than the test scores assessed for C2. This trend was seen within Result 1.1 as well. Another similar pattern to Result 1.1 was the fact that there were unequal variances for the untransformed variable, but the transformed variable for Treatment Combination 2 showed equal variances.

Like Result 1.1, the one-way ANOVA for Treatment Combination 2 showed an overall significance level that was less than .001 ($F(2, 202) = 22.10$, partial $\eta^2 = .180$). Pairwise comparisons showed that there were significant differences when the points groups were compared to the control groups. The p value for C2 and A1 was less than .001. The control group in the second half of the quarter scored significantly higher than

the points group from the first half. Furthermore, the *p* value for C2 and B2 was less than .001. This means that the control group from the second half scored higher than points group from the second half.

The ANOVAs showed that the untransformed results lacked a significant difference between Groups A1 and B2 ($p = .129$), but the transformed results did show that there was a statistically significant difference between the groups ($p = .032$). The *p* values generated for this result differed from the ones in Result 1.1 despite the fact that these groups were used within both results. What was similar between the results was that the transformed results showed a statistically significant result when the untransformed results did not show this. Again, those who were awarded points in the first half scored significantly higher than those who were awarded points in the second half. Both groups scored significantly lower than the control groups.

- Result 1.3: Students who received no coins for the first half of the quarter scored significantly higher than those who received coins at any time during the study.

This result was discovered within Treatment Combination 3, which had three groups: A2, B1, and C1. Students within A2 were those who received coins during the second half of the quarter. Group B1 described students who received coins during the first half of the quarter. The third group, C1, included participants who were in the control group for the first half. The least amount of students were found in A2 ($n = 54$). Group B1 ($n = 66$) had a higher amount than A2, but it contained less participants than C1 ($n = 84$).

Data for B1 and C1 were collected from the pretest, and Group A2 contained posttest data. The missing score for the pretest was found within B1. The group mean

for students in C1 (M = 77.82, SD = 15.93) was higher than the means in the treatment groups. Students within A2 (M = 68.94, SD = 17.31) scored higher on average than students in B1 (M = 61.48, SD = 18.50). For the one-way ANOVAs, homogeneity was upheld for the untransformed version of the scores (p = .479). Since homogeneity was not violated in this instance, a transformed version of the variable was not needed for statistical analyses.

The difference of means was the smallest for A2 and B1 at 7.46, which indicated that the coins groups had the closest test score averages. Students who were awarded coins in the second half scored higher than students who were awarded coins in the first half. A slightly higher mean difference of 8.88 was found between C1 and A2. Like Result 1.1, the control group for the first half scored higher than the treatment group for the first half. The most amount of difference in the means was found between C1 and B1 at 16.34, which was almost twice the amount found between C1 and A2. Therefore, the control group from the first half scored higher than both coins groups.

Similar to the previous results, the group with the highest test score had the smallest amount of variability within the scores. Within Treatment Combination 3, this group was within Group C1. The differences between the standard deviations were small. There was a difference of 1.19 between Groups A2 and B1, a difference of 1.38 between Groups C1 and A2, and a difference of 2.57 between Groups C1 and B1. The most similarity in terms of range of scores would be found with Groups A2 and B1. Levene's test showed that the three groups had equal variances and that the differences within the standard deviations were not significant.

Within the one-way ANOVA for Treatment Combination 3, the overall results showed significant differences ($F(2, 201) = 16.94$, partial $\eta^2 = .144$). The p value within the procedure was less than .001. Unlike the untransformed versions of Treatment Combinations 1 and 2, the pairwise comparisons for the untransformed version of Treatment Combination 3 indicated that all pairs showed statistically significant differences. The difference between the scores within A2 and B1 was bordering on statistical significance ($p = .049$). For the coins treatment, students within the coins group for the second half scored significantly higher than students in the coins group for the first half. The p values for the other comparisons were much lower than the one discussed. The p value for groups A2 and C1 was .009, and the p value for groups B1 and C1 was less than .001. These results demonstrated that the control group in the first half scored significantly higher than both coins groups. This trend is seen with Result 1.1 with the points treatments.

- Result 1.4: Students who received no coins for the second half of the quarter scored significantly higher than those who received coins at any time during the study.

This result was included within Treatment Combination 4. This combination contained Groups A2, B1, and C2. A2 and B1 were the same groups used within Result 1.3. Group C2 involved those participants who were in the control group during the second half of the 9-week quarter. There were 84 students within C2, which was higher than A2 and B1. The amount of participants for A2 and B1 are provided in Result 1.3.

Data for Groups A2 and C2 were collected during the posttest, but data for B1 was collected within the pretest. In terms of group means, Group C2 ($M = 81.70$, $SD = 13.15$) scored higher than Group A2. Group C2 also had a higher group mean than

Group B1. These trends were similar to the ones in Result 1.2. For Result 1.2, Group C2 had higher scores on average than the treatment groups for the points intervention. For the one-way ANOVAs within Treatment Combination 4, homogeneity was violated for the untransformed variable ($p = .004$). This was corrected for the transformed version of the test scores ($p = .546$).

The highest difference of means was found between C2 and B1. There was a gap of 20.22 between the scores for these groups. Students in the control group for the second quarter scored higher than the students who received coins in the first half of the quarter. In addition, group C2 had a higher group mean than A2 by approximately 12.76 points. Students in the control group for the second half had a higher group mean than students who received coins during the second half of the quarter. The lowest mean difference was 7.46, which was between A2 and B1. The same amount of difference appeared within Result 1.3.

The group with the highest amount of variability was B1. When compared to the other groups, there was a wider range of scores for students who received coins at the first half of the quarter. The difference between the standard deviations of B1 and A2 was small at 1.19. More notable differences in variability were found between the treatment groups and the control group. There was a difference of 5.35 between C2 and B1, and there was a difference of 4.16 between C2 and A2. The differences in variability listed were significant within Levene's test when the untransformed version of the test scores was assessed. This was not the case with the transformed variable, which showed that equal variances could be assumed.

Furthermore, one-way ANOVA procedures showed an overall p value of less than .001 ($F(2, 201) = 30.08$, partial $\eta^2 = .230$) for Treatment Combination 4. Post-hoc testing with Tukey's HSD revealed a significant difference between C2 and A2 where p was less than .001. The control group in the second half scored significantly higher than the coins group in the second half. There was a significant difference between C2 and B1 that had a p value of less than .001. The control group from the second half had a significantly higher test score than the coins group from the first half as well.

Finally, the untransformed and transformed variables had different results when comparing A2 with B1. The untransformed version of the test scores revealed a significant difference between the groups ($p = .034$). The transformed version, however, showed that there was no significant difference between the groups ($p = .416$). In other words, the transformed version stated that the coins groups had similar performances on their tests since the difference between the means was too small to be of statistical importance. Result 1.3 displayed a significant difference for these pairs ($p = .049$), but the variable used for the test scores did not require transformation. This contradiction is further discussed within Result 1.6.

- Result 1.5: Students who receive points for the first half of the quarter scored significantly higher than those who received points for the second half of the quarter.

The result listed was applicable to Treatment Combinations 1 and 2. The groups that pertained to this result were A1 and B2. Since these groups are discussed within Results 1.1 and 1.2, this result elaborates on what is previously discussed. Again, A1 had 54 participants. Group B2 had 67 participants. This means that more students received points at the second half of the quarter.

The minimum and maximum scores were different for each group. The maximum test score for Group A1 was 105, which was higher than B2. The maximum score for students within B2 was 100. For the minimum, A1 had a score of 26, but Group B2 had a score of 21. Within A1, there were 24 students who scored below the mean. There were 30 who scored above the mean. Analysis for B2 showed that 31 students scored below their group mean. There were 36 who scored above the mean. Given the information provided, it was more likely that the points group for the first half would have a higher mean than the points group for the second half. Most students in A1 scored above their group mean ($M = 69.69$), which was higher than the group mean for B2 ($M = 63.63$).

The control group for the first half of the quarter had the same maximum score as B2, but the minimum was 40. The control group for the second half had a higher maximum score than all of the groups at 108. It also had the highest minimum score at 44. Consequently, both control groups had higher group means than the treatment groups. Regardless of timing, the control groups scored higher than the points groups. Percentile ranks for these four groups are presented in Table 7.

When the scores were configured according to percentile ranks, it was observed that students in C2 outperformed the other groups at each listed rank. Students in C2 who had a score within the 25th percentile met the requirements for passing exams, but the same cannot be said for the other groups. Students within Groups A1 and C1 had to score as well or better than 50% of the students within their groups in order to pass. For Group B2, a passing score was not seen for the 10th, 25th, or 50th percentiles. The

lowest scores were seen within B2 for the 10th, 25th, and 50th percentiles. Group A1 had the lowest scores for the 75th and 90th percentiles.

Table 7

Scores Delimiting Percentile Categories for Result 1.5

Groups	10th	25th	50th	75th	90th
A1	46.00	59.75	72.50	80.75	88.00
B2	36.00	48.00	64.00	81.00	94.00
C1	54.00	68.00	81.00	89.75	96.00
C2	63.50	75.00	82.00	91.75	100.00

Note. Percentiles were generated using the Explore procedure in SPSS.

For the one-way ANOVAs in both combinations, the untransformed versions of the test scores displayed no significant differences between A1 and B2 during post-hoc testing ($p > .05$), but the transformed versions displayed significant differences for the pair ($p < .05$). Since the transformed versions of the combinations upheld homogeneity, they depicted more accurate results than the untransformed versions in terms of significance. Based on these results, it can be concluded that students received a higher score on average when points were received in the first half of the quarter.

- Result 1.6: Students who received coins for the second half of the quarter scored significantly higher than those who received coins for the first half of the quarter.

This result pertained to what was observed within Treatment Combinations 3 and 4. The groups that were relevant to this result were A2 and B1, which were first written about within Results 1.3 and 1.4. This section provides more details about these previous results. A2 had the same number of students found within A1 ($n = 54$), but B1 ($n = 66$)

had one less student than B2. As far as the coins treatments were concerned, there were more students who received coins in the first half of the quarter.

Each group within Treatment Combinations 3 and 4 had its own minimum and maximum. Group A2 had a maximum of 98 and a minimum of 36. There were 23 students who scored below the group mean ($M = 68.94$); however, there were 31 students who scored above the mean. Group B1 had a higher maximum at 100, but the minimum was lower at 23. Thirty-two students had test scores below their group mean ($M = 61.48$). Thirty-four students scored above the mean. Group B1 had more students score below the mean than Group A2. The lower scores made it more likely for the mean to be lower within Group B1.

For the control group in the first half (i.e., C1), the maximum was the same as B1. The minimum was higher than the other 2 groups at 40. The control group for the second half, C2, had the highest minimum at 44 and the highest maximum at 108. The control groups scored higher than the other 2 groups regardless of timing. Percentile ranks are available within Table 8.

Similar to Result 1.5, the percentile ranks showed that C2 had the highest scores for each rank listed. In order to pass within Group C2, students had to score as well or better than 25% of the distribution. Students who scored within the 50th percentile passed if they were in Groups A2, C1, or C2. A passing score was not seen for B1 until the 75th percentile, which showed that a student had to score a 76 in order to do as well or better than 75% of the sample distribution. The lowest scores for the percentiles were found within B1.

Table 8

Scores Delimiting Percentile Categories for Result 1.6

Groups	10th	25th	50th	75th	90th
A2	44.00	54.00	70.00	81.75	94.00
B1	34.00	49.00	63.00	76.00	88.60
C1	54.00	68.00	81.00	89.75	96.00
C2	63.50	75.00	82.00	91.75	100.00

Note. Percentiles were generated using the Explore procedure in SPSS.

During the pairwise comparison procedures, the untransformed versions of these combinations displayed significant differences between A2 and B1 ($p < .05$). The transformed version of Treatment Combination 4 found no significant differences between the pair ($p > .05$). Out of the two combinations, Treatment Combination 3 provided a more accurate depiction of this pair because (a) the untransformed version of Treatment Combination 3 upheld homogeneity without requiring a transformation and (b) the direction of the untransformed mean differences coincided with the data that was collected for the study. From this information, it can be concluded that there was a significant difference between groups A2 and B1. Those who received coins in the second half of the quarter scored significantly higher than those who received coins in the first half of the quarter.

Interpretation of Finding 1.1. The fact that the control group consistently outperformed the token groups required further exploration. One possible reason for this was found after random assignment was completed for the groups. The control group had accelerated math instruction for 3 of the 4 classes. The other groups, which had 3

classes each, did not have accelerated math groups. On the pretest, all accelerated groups scored higher than the groups who received the tokens. The posttest showed that 2 out of the 3 accelerated groups scored higher than the other groups. The exception was found within seventh grade, where the accelerated group scored lower than the rest of the groups in the grade level.

Another possibility was found within trends for the grade levels. All 4 treatment combinations indicated that students in sixth and eighth grade scored the highest within the control group. For Treatment Combinations 1 and 2, students in seventh grade had the highest group mean within the points group for the second half. When the other grades were included in the calculations, the points group for the first half scored higher than the second half. The other 2 treatment combinations for this research question showed that seventh grade scored the highest in the coins group for the second half. If the seventh-grade scores were removed from these groups, a greater disparity between the treatment and control groups would have been seen.

A third variable that possibly contributed to the high performance of the control groups was ECD status. Within the combinations for points timing, students who had free lunch status scored lower than students who received reduced-priced lunch. The Free Lunch group scored lower than the Paid-in-Full group as well. All of these groups scored higher than 70% within the control group, but not all of the groups passed when using tokens. For instance, Free Lunch students did not pass when using the points at any time. The Paid-in-Full group did not pass in the second half, and the Reduced-Price students passed within all groups.

Data Mixing for Finding 1.1. The teachers often had higher expectations for the students who performed well and behaved well. As they did what they were supposed to do, they became models for other students. The sixth-grade teacher said that her sixth-grade control group was "more motivated" than the other groups. They were working for grades. There were ones who could have benefitted from the points and coins, but the teacher felt that nothing would have been gained from it if the control group had been selected for reinforcement. She stated that the ones who received the coins were the "neediest" group. The ones who received points were in the middle in terms of their need for reinforcement.

> Teacher 1 – "Those kids, they're just working for grades. I mean, overall, those kids are on it. They just…um…I got a few of them in there who I think the points or coins would probably be good, but overall that group…They're in it for themselves. They seemed to be more motivated. And then the kids that are getting my tokens…they're the ones who are probably my neediest group and need that reinforcement. And the ones with the points, they pretty much go either kind of way kind of class."

The seventh-grade teacher was of the opinion that the majority of students were "good kids." They were very well-behaved for her. She felt it was her responsibility to give them a lot of reinforcement because they were following the rules frequently and receiving good grades. This, in turn, provided more motivation for them than for the "bad kids." The "bad kids," who were few in number, received the proper punishment according to their actions. Before experiencing the PBIS system, she was more in favor of a system that would strictly "punish the negative behavior." Her experiences with PBIS changed her outlook to include reinforcement. She found that having tokens and rewards "raises morale" of the students. She had two control groups, and one was in the accelerated group.

> Teacher 2 – "Um, and that is probably…If you had asked me that a few years ago, I probably would have said to punish the negative behavior (briefly laughs.), but uh…you know, being on the PBIS team I definitely think…I mean, I see that as a—as a positive in our school. I think it raises morale of the kids to have, you know, to be noticed for doing the right thing. I think that it's both…And I enjoy rewarding our good kids. We have a lot of good kids. Um…it's just a tiny little percent of kids who cause behavior problems. So, I think it's a little bit of both. And I think it's better…I always think in teaching, better to be proactive than reactive."

The eighth-grade teacher found that behavior was consistent across classrooms. From her perspective, the students who were well-behaved were "intrinsically motivated" to continue their behavior. The other students lacked this type of motivation. She felt that it "doesn't matter" if reinforcement is received. This meant that she expected students who did "what they are supposed to do" to do the right thing without being prompted by her to do it. They "want" to behave and receive good grades without additional forms of reinforcement. Based on her statements, it was implied that those who misbehaved for her were more likely to be found in the treatment groups. Her control group had accelerated math, so they met general curriculum requirements at a faster pace than most students.

> Teacher 3, Comment 1 – Um…The ones that behave for me, they behave for the other teachers. The ones who misbehave for the other teachers, by the time they get to me…You know, I was telling one of 'em today. I said "The reason you get write-ups from me is because you've misbehaved all day long, and you're coming into my classroom with your other two teachers following you, saying, 'If he blinks wrong, write him up because he gave me a fit today. You know, and I warned him and warned him. Now, he doesn't deserve anym—' you know.
>
> Teacher 3, Comment 2 – "What I see is that, in general, students who…are intrinsically motivated…students who are going to do what they're supposed to do just because that's what they want to do. It really doesn't matter if I reinforce that or not."

Research Question 1b: To what degree are test scores influenced by the type of token used in the classroom, i.e., points-based system versus coin-based system versus no tokens used at all?

Research Question 1b required the use of parametric tests and non-parametric tests. This differed from Research Question 1a due to the fact that Research Question 1a only included parametric testing. Parametric tests were performed for the pretest and posttest. For Research Question 1b, a one-way ANOVA was executed to determine differences in pretest scores when grouped according to token type. After running the ANOVA for the pretest scores, separate ANOVA findings were generated for the posttest scores. This is also unlike Research Question 1a, which required the pretest and posttest scores to be analyzed together within different combinations.

Non-parametric tests were necessary because they addressed violations in homogeneity during procedures with the posttest scores. For non-parametric testing, Kruskal-Wallis *H* Tests were performed with the posttest scores. The Kruskal-Wallis *H* Tests were used for non-parametric testing due to the fact that it was the non-parametric equivalent to the one-way ANOVA (Cronk, 2012). The combinations for Research Question 1b looked at different token types within the same time frame of testing. Originally, there was only one combination available during the time of the treatment switch. It was labeled according to the guidelines found in Chapter 3. After the data collection phase ended, another combination was added. The first combination was relabeled. The resulting combinations that are required to answer this question are outlined within Table 9.

Table 9

Combinations for Research 1b Analyses

Groups	Pretest Tokens (A1 X B1)	Posttest Tokens (A2 X B2)
Pretest Points (A1)	A1 X A1 X B1	A1 X A2 X B2
Pretest Coins (B1)	B1 X A1 X B1	B1 X A2 X B2
Pretest Control (C1)	C1 X A1 X B1	C1 X A2 X B2
Posttest Coins (A2)	A2 X A1 X B1	A2 X A2 X B2
Posttest Points (B2)	B2 X A1 X B1	B2 X A2 X B2
Posttest Control (C2)	C2 X A1 X B1	C2 X A2 X B2

Note. Relevant independent combinations for Research Question 1b are in black. These combinations do not have timing conflicts.

The 2 combinations for Research Question 1b were labeled Treatment Combination 5 and Treatment Combination 6. The fifth combination addressed comparisons among token type that were available for the pretest. For the pretest scores, there were 2 different treatment groups and a control group. Categories A1 and B1 were labels used for the treatment groups, and category C1 was used for the control group. The sixth combination allowed for posttest comparisons between the coins treatment for Group 1, the points treatment for Group 2, and the control for Group 3. These groups were represented by the categories A2, B2, and C2. Here is the overall finding that pertains to token type:

> *Finding 1.2: Students who received no tokens scored significantly higher than the points and coins groups; moreover, the points groups scored higher than the coins groups.*

While Finding 1.1 focused on the timing aspect of the treatment, Finding 1.2 pertained to the different token types for the treatment. The control groups received the

highest scores within the treatment combinations for token type. The control group scored significantly higher on the pretest than the token groups, which received points and coins. After the switch point, there was a bigger gap seen between the control group and the token groups. Within the points and coins groups, a significant difference was seen for the pretest scores. The points group scored significantly higher than the coins group. This difference was not seen for the posttest scores.

- Result 1.7: Students who received no tokens in the first half scored significantly higher than students who were awarded tokens.

This significant finding was seen within the scores and categories for Treatment Combination 5. There were 3 groups assessed within this combination: A1, B1, and C1. Group A1 represented the students who received points for the first half of the nine-week quarter. Group B1 included students who received coins for the first half, and Group C1 applied to those students who were in the control group for the first half. Group C1 (n = 84) had the highest number of participants. Groups A1 (n = 54) and B1 (n = 66) did have enough students available in order for the study to maintain statistical credibility. In other words, there were over 187 participants who were able to provide test scores.

Groups A1, B1, and C1 were used to represent the students who had pretest scores for the first half of the quarter. Students in C1 (M = 77.82, SD = 15.93) had the highest group mean among these groups. Participants within A1 (M = 69.69, SD = 16.73) had a higher group mean than participants in B1 (M = 61.48, SD = 18.50). For the one-way ANOVA procedure, homogeneity was upheld within the untransformed variable (p = .458); hence, a correction via transformation was not needed.

The greatest difference of means was 16.34, and it was found between groups C1 and B1. Students who received the control for the first half of the quarter scored higher,

on average, than students who received coins for the first half. Another noteworthy difference was between C1 and A1 at 8.13. The third observed difference of means was 8.21, which was between A1 and B1.

The one-way ANOVA procedure yielded significant differences overall for the pretest scores ($F(2, 201) = 17.12$, $p < .001$, partial $\eta^2 = .146$). Upon closer inspection within pairwise comparisons, the output displayed significant differences for all pairs. Groups C1 and A1 had a p value of .018. The control group in the first half scored significantly higher than the points group within the same time frame. A lower p value that was less than .001 was seen between C1 and B1. The control group not only scored significantly higher than the points group during the pretest, but it also scored significantly higher than the coins group in the first half. The third pair, A1 and B1, had a p value of .025. The treatment groups had a significant difference whereby the points group for the first half scored significantly higher than the coins group for the first half.

- Result 1.8: Students who received no tokens in the second half scored significantly higher than students who were awarded tokens.

This finding was observed within Treatment Combination 6. There were three groups for this particular combination: A2, B2, and C2. Group A2 was utilized for those students who received coins during the second half of the quarter. The group known as B2 centered around students who were awarded points for the second half. C2, which was the third group, represented students who received instruction in the control group for the second half. Group C2 ($n = 84$) had the highest number of participants. The second highest amount was found in Group B2 ($n = 67$), and the lowest number for participants was in Group A2 ($n = 54$). The amount of participants who took the posttest did exceed the number necessary for statistical credibility (i.e., 187).

In contrast to the groups within Result 1.7, the groups for this finding included posttest data. The highest group mean was shown in Group C2 (M = 81.70, SD = 13.15). For the treatment groups, Group A2 (M = 68.94, SD = 17.31) had a higher group mean than Group B2 (M = 63.63, SD = 21.10). Overall, students in B2 and C2 improved their scores from their pretest (i.e., B1 and C1). A2 had a lower group mean for their posttest. During the one-way ANOVA procedures, homogeneity was violated for the untransformed posttest information ($p < .001$). This was not corrected for the transformed version ($p = .043$). As a result, non-parametric testing was required with the Kruskal-Wallis H Test.

Mean differences were calculated between the groups, and the largest gap was found between B2 and C2 at 18.07. Those who were within the control group for the posttest outperformed students who were awarded points for the second half. The mean difference of 12.76 between C2 and A2 indicated that the control group who scored higher, on average, than the points group scored higher than the coins group for the posttest. These 2 pairs, when compared with Result 1.7, had higher mean differences than their corresponding pretest pairs. For example, the mean difference between B2 and C2 was greater than the mean difference between B1 and C1. Moreover, the treatment groups A2 and B2 had a mean difference of 5.31. The coins group scored higher than the points group. The treatment group comparison had a much lower mean difference than the other comparisons. This mean difference was lower than the pretest difference for A1 and B1 in Result 1.7.

Overall, the one-way ANOVAs produced an overall p value that was less than .001 ($F(2, 202) = 22.14$, partial $\eta^2 = .180$) for posttest scores in Treatment Combination 6.

Within pairwise comparisons, there was a significant difference of less than .001 for A2 and C2. The control group from the second half scored significantly higher than the coins group from the same time period. B2 and C2 had the same *p* value in both versions ($p < .001$). Therefore, the control group from the second half scored significantly higher than the points group from the second half as well. Finding 1.7 had a similar trend within the pretest scores. For treatment groups A2 and B2, no significant differences were found ($p = .210$). This trend differed from the 1 seen for the pretest results since there were significant differences found for A1 and B1 ($p < .05$).

Kruskal-Wallis *H* Tests were required for the posttest. They were used in order to see if the significant differences found within the parametric tests still existed without the homogeneity assumption. The results for the untransformed and transformed variables were the same for these tests. Overall, there were significant differences for the groups ($H(2) = 33.48$, $p < .001$). Pairwise comparisons generated the same significant differences found in the ANOVAs. A *p* value that was less than .001 was shown for A2 and C2. The same value was observed for B2 and C2. The control group scored significantly higher than the points and coins groups. No significant differences were found for A2 and B2 ($p = .917$). For the posttest, the points and coins groups had a similar performance overall. This coincided with the results for the one-way ANOVA procedures.

- Result 1.9: Students who received points scored significantly higher than students who received coins.

This result focused on data discovered within Treatment Combination 5. The groups that were relevant to this finding were A1 and B1. These pretest categories were discussed in Result 1.7. Any information in this finding elaborates on what is previously

discussed, and it does include comparisons to what was discovered for Result 1.8. A1 contained less students (n = 54) than B1 (n = 66). There was a difference of 12 participants between the two groups. This means that there were more students who received coins than points within the first half of the 9-week quarter. For the posttest pair A2 (n = 54) and B2 (n = 67), there was a difference of 13 between the groups. There were more participants who received points than coins in the second half.

In terms of minima, Groups A1 and B1 had a difference of 3 between them. The lowest score within A1 was 26, which was more than the minimum of B1 at 23. For the maxima, Group A1 had the highest pretest score at 105. This was 5 points more than the maximum score of 100 for B1. Category A1 had 24 students who scored below the mean (M = 69.69) and 30 students who scored above the mean. Category B1 had different results where 32 students scored below the mean (M = 61.48) while 34 students scored above the mean. The higher maximum and minimum for A1 as well as the fact that less students scored below the mean in A1 increased the likelihood that those who received points would score higher than the ones who received coins.

Within the posttest, A2 had a higher minimum at 36 and a lower maximum at 98 to its counterpart A1. The group mean, however, was higher for A1. As far as the different quasi-experimental groups were concerned, students in these groups scored lower when they received coins in A2. Group B2 had the same maximum as B1, but the minimum was less than B1 at 21. The posttest mean for B2 was higher than B1. Students who participated in B1 and B2 scored higher when they received points in B2. The pretest control group, C1, had a highest minimum within the pretest results at 40. The maximum score within the control group was the same as Group B1. Category C2,

the posttest category for the control group, had the highest minimum overall at 44 and the highest maximum overall at 108. The group mean for C2 was higher than the other group means calculated for this study. Regardless of token type, the control groups scored higher on average than the treatment groups (see Table 10).

Table 10

Scores Delimiting Percentile Categories for Result 1.9

Groups	10th	25th	50th	75th	90th
Pretest					
A1	46.00	59.75	72.50	80.75	88.00
B1	34.00	49.00	63.00	76.00	88.60
C1	54.00	68.00	81.00	89.75	96.00
Posttest					
A2	44.00	54.00	70.00	81.75	94.00
B2	36.00	48.00	64.00	81.00	94.00
C2	63.50	75.00	82.00	91.75	100.00

Note. Percentiles were generated using the Explore procedure in SPSS.

During comparisons with the pretest and posttest categories, most of the percentiles were higher for the posttest categories with the exception of A2 at the 10th, 25th, and 50th percentiles along with B2 at the 25th percentile. For instance, a student in Group 1 needed to have a pretest score of 59.75 to do as well or better than 25% of the distribution. The posttest score for the same group was 54.00 in the same percentile rank. For the majority of comparisons among the ranks, students had to perform better during the posttest in order to match or beat the other participants' scores at that time. Passing

scores for the pretest appeared within the 50th percentiles for A1 and C1. Group B1 had a passing score beginning at the 75th percentile, indicating that students were least likely to pass within B1. Furthermore, the posttest showed that the same group (i.e., B2) did not have a passing score until the same percentile rank was reached. Even with a different token treatment, B2 students were less likely to pass than Groups A2 and C2. Students in C2 could have attained a passing score within the 25th percentile, which was an improvement over the pretest. Students in A2 could have passed by performing better or as well as 50% of the students. This passing rate was seen for A1 in the pretest scores, except that students needed to score 2.50 more for the pretest to make it to the 50th percentile.

Since homogeneity was not violated within the untransformed pretest scores for Treatment Combination 5 ($p = .458$), the amount of variability in the pretest was less than that found for Treatment Combination 6. The untransformed ($p < .001$) and transformed ($p = .043$) versions of the posttest scores did not correct the issue of unequal variances for Treatment Combination 6. The untransformed version of Treatment Combination 6 had the highest amount of variability. Post-hoc testing in the pretest produced a significant difference for the pair A1 and B1 ($p < .05$), but the posttest scores yielded no significant differences within A2 and B2 ($p > .05$). For this reason, it can be concluded that students during the points treatment in the first half scored significantly higher than students during the coins treatment in the first half. They scored similarly in the posttest after the treatments were switched.

Interpretation of Finding 1.2. Similar to Finding 1.1, the control groups outperformed those who had tokens. This finding added to this result the fact that the

pretest token groups had a significant difference, but the posttest token groups did not have this difference. One consideration, as stated within the previous finding, involved accelerated courses. For sixth grade, the accelerated group had the only passing scores for the pretest and posttest. All groups scored lower on the posttest. In seventh grade, the posttest scores were higher for three out of the four classes. The accelerated group scored the lowest on the posttest, even though they scored the highest on the pretest. Only the accelerated group passed the pretest, but all classes passed the posttest. In eighth grade, there were two classes that passed the pretest. One was the accelerated group, and one was the points group. Only the accelerated group passed the posttest.

Grade, excluding acceleration, was a second area of interest that possibly contributed to the differences found within the treatment and control groups. This was discussed in the previous finding as well. All grades had their highest scores within the control group when sorted by their pretest averages. Sixth and eighth grade had their lowest scores within the coins group, but seventh grade had the lowest scores within the points group. The posttest showed that the highest score for seventh grade changed from the control group to the points group. The lowest score for seventh was within the control group, whereas the lowest scores for the other grades involved the points tokens.

Another area of possible influence was gender. Boys and girls scored at least a 70 within the pretest control group. The boys had the same result within the points group as well, but the girls did not receive these scores within either token group. The lowest scores were found within the coins group. A significant influence was not found for the posttest; however, the only passing group means were found within the control group. The lowest posttest scores for both groups were found in the points group.

Data Mixing for Finding 1.2. One action that the teachers had in common was the fact that they said they gave out more coins than points. The control group was given no tokens and prizes; therefore, the points groups within each grade level were closer to the control groups in terms of frequency. The sixth-grade teacher stated that she felt the tangible tokens (i.e., coins) were a "distraction" for the students. She did not say this about the points. She explicitly stated that she gave "more tick marks" to those with the students who received coins, even though she found it easier to mark the points. This was because students liked the coins. Unfortunately, there were times when students did "pay attention" to the tokens to the point where it affected their focus during instruction.

> Teacher 1, Comment 1 – "Um, cause I seem to give more positive…More tick marks for the class that I've done the coins with…versus the one where the kids are supposed to be keeping up with the tally marks on a sheet of paper."
>
> Teacher 1, Comment 2 – "Um…Like I said…I don't really care for the tokens. Um, because kids start tapping them. They're becoming a distraction to some of them…with the ADD…that silver, shiny thing…Let's go pay attention to it and not the instruction. So, overall, it's not my favorite."

The seventh-grade teacher stated that she thought it was "easy" to forget about the points. Like the sixth-grade teacher, she gave out more coins than points. In other words, the amount of tokens given to the points group was more than the control group but less than the coins group. The ones in the coins group were more likely to be awarded when they exhibited good behavior. Although the students had a "harder" time keeping up with the coins, the teacher had a different experience. She kept the coins in a box on her desk, and she was able to "notice" them quickly during instruction.

> Teacher 2, Comment 1 – "And so…it was easy for me to forget about it. It has been easy for me to forget about the points more."
>
> Teacher 2, Comment 2 – "The coins…I would say…Like I say, they're a little harder to keep up with for the kids, but the good thing is that that box is on my desk and a lot of times I walk by and notice it and grab it and pick it up and

> start walkin' around, and…you know… (makes giving out gesture with hands, makes slight noise with table.) …giving out a couple coins. So, to me, that's just the kinda person I am. When I see something like that, it reminds me of it. So, it was just a good…I mean, that's the good thing for me with that."

The eighth-grade teacher stated that she did not give as much points as she did coins. This was stated after the switch had taken place. She said that the points group had serious "behavior issues", and this is why she did not give as many points. Despite this, the minimum and maximum amounts for each group were relatively close. The points group received anywhere from "11 to 24" points, whereas the coins groups were awarded approximately "11 to 27" coins. The close ranges were important to consider because the students' overall scores within the token groups had less of a difference within group comparisons. Performance gaps were wider when the token groups were compared to the control groups.

> Teacher 3 – "So, most of 'em that have a low number, um, have traded in. But, yeah, it's about the same. Um, maybe not quite as high with points (flipping around the log sheets.) even though they were, um…Yeah, not quite as high with points. Um, the behavior…The ones who are on points now, that's my last class of the day, and we have some serious behavior issues in there. Like five who have been kinda targeted and said if they mess up again this year, they're out of here. You know, so that last class of the day is tough. And um, so even though the points are easier for me to keep up with and easier for me to make them aware of on a daily basis, um, I just didn't give as many to that group because their behavior is not as good. But still…You know, the low one in that group that probably is just…um…Yeah, their low is like 11, and there's a high of 24 it looks like. Yeah, so 11 to 24 versus 11 to 27 maybe. So."

Research Question 2a: To what extent are the test scores influenced by classroom behavior referrals?

Classroom behavior referrals were assessed through the utilization of parametric tests and their non-parametric equivalents. One-way ANOVAs and *t* tests were run in

order to determine the degree to which test scores were affected by classroom behavior. Referral data were collected at the beginning, middle, and end of the quasi-experiment. From the referral information, a nominal variable for classroom behavior was created for each collection period. Each variable had separate procedures. The first classroom behavior variable had two groups for the pretest participants (i.e., No Referrals and Low Frequency), and the appropriate test for this variable was the *t* test for independent samples. The second variable had three groups: No Referrals, Low Frequency, and High Frequency. With the pretest information, a one-way ANOVA was performed to see if there were any significant differences among the three groups when they were grouped according to referral information provided at the switch point. The third variable was created after the posttest was completed by the participants. There were three groups in this variable, and they were the same groups used for the second variable. A one-way ANOVA was performed for this variable with the posttest scores; however, post-hoc testing was not completed because one of the groups had an insufficient number of students within it. Because of this, a *t* test was performed in order to compare the two groups that had a sufficient number of participants. Homogeneity was not violated within these tests, and transformed versions were not needed for them.

Furthermore, three ratio variables for referral numbers were created with the referral data collected during the study. These variables were used along with the nominal variables for classroom behavior in order to run additional tests. The ratio variables had a severe positive skew where the number of referrals given was very low. Assumptions of normality and homogeneity could not be validated given the nature of the variables. Therefore, non-parametric tests were performed to see whether the differences

in referral numbers were significant. The Mann-Whitney *U* Test was used as a non-parametric equivalent for the *t* test that compared the differences in pretest referral information among two groups. Cronk (2012) recommends the Mann-Whitney *U* when independent samples are used within research. The switch point referral amounts required the use of the Kruskal-Wallis *H* Test, which compared differences in pretest information among three groups. As outlined within Research Question 1b, this test is a non-parametric equivalent to the one-way ANOVA. The referral amounts for the posttest were compared using the Kruskal-Wallis *H* Test since the non-parametric requirements for group numbers were less strict than what was found within the one-way ANOVA. Here is the overall finding for the referral data:

> *Finding 2.1: Students who had no referrals showed significant differences in test scores and referral frequencies when compared to those who had referrals.*

Data for the referrals showed that students who had zero referrals in the month of February scored significantly higher, on average, than students who received a low frequency of referrals (i.e., one or two) during the same month. This was not seen for the months of December and January. For all months, there were significant differences in terms of referral amounts. The students who received referrals had a significantly higher amount than the students who received zero referrals for those months. There were no significant differences in referral amounts between the low frequency and high frequency (i.e., three or more) groups.

- Result 2.1: Students who received referrals in December, January, and February had a significantly higher amount of them than students who received no referrals.

Classroom behavior had three possible groups: No Referrals, Low Frequency, and High Frequency. As stated within Chapter 3, No Referrals includes those participants

who received zero referrals in a month. Low Frequency referred to participants who had one or two referrals in a month. The High Frequency category had participants who received three or more referrals in a month. In December, there were 199 students (97.1%) who received no referrals overall, and there were six students (2.9%) who were classified as Low Frequency. There were no students in the High Frequency category. Four students received one referral within a month's time. There were two who received two referrals.

For January, there were 192 students (93.7%) who received no referrals. This was a decrease of seven from the number in December. There were 11 students (5.4%) who belonged to the Low Frequency category, which meant that the category had five more students than the previous month. Ten students received one referral each, and one student received two referrals. There were two students (1.0%) who belonged to the High Frequency category. One student received three referrals for the month of January, and the other student received four referrals.

In February, there were 190 students (92.7%) with no referrals. This showed a decrease of two when compared to the month of January. There were 14 students (6.8%) with a low amount of referrals, and there was a student (0.5%) who had a high amount of referrals. The Low Frequency category had increased by three, and the High Frequency category decreased by one. In the Low Frequency category, there were 10 participants who received one referral within a month's time. Four received two referrals each. The High Frequency student received four referrals.

Since the December information contained only two groups, the Mann-Whitney *U* Test was used to see if significant differences existed for the referral data collected at that

time. Significant differences were observed for the untransformed (U = 1,194.00, p < .001) referral amounts for December. The Low Frequency group (m place = 202.50) received a significantly higher frequency of referrals when compared to the No Referrals group (m place = 100.00).

For January, the Kruskal-Wallis H Test was performed on three groups. Means for pretest and referral data were assessed during the procedures. There were significant differences found for the referral variable overall ($H(2) = 203.96$, $p < .001$). When the No Referrals group was compared to the Low Frequency group, a p value of less than .001 was displayed. This p value was displayed between the No Referrals group and the High Frequency group as well. The No Referrals group had significantly lower referral amounts than the Low Frequency and High Frequency groups. No significant differences were found between the Low Frequency and High Frequency groups ($p > .05$). This means that the referral amounts were statistically similar within these two groups.

February results required the use of Kruskal-Wallis H Tests to compare posttest means and referral numbers. The third group was counted regardless of size, which meant that Mann-Whitney U Test could not be used here. In other words, the non-parametric Kruskal-Wallis H Test removed the size requirement previously explained for the parametric one-way ANOVAs. Significant differences were found for the referral information overall ($H(2) = 203.81$, $p < .001$). Like the results in January, a p value less than .001 was found between the No Referrals group and the Low Frequency group. This same p value appeared between the No Referrals group and the High Frequency group. The No Referrals group had significantly less referrals than the other two groups. No

significant differences were found between the Low Frequency and High Frequency pair. The referral amounts for these groups were statistically similar to one another.

- Result 2.2: Students who received no referrals scored significantly higher on the posttest than students who received a low frequency of referrals.

As stated in the previous result, referral information was collected for December, January, and February. December and January referral variables were analyzed along with the pretest information, which was collected at the beginning of January. The student who was missing a score for the pretest had no referrals. February referral variables were applicable for the posttest information that was collected between the end of February and the beginning of March. For each month, most students were found within the No Referrals group. The lowest number found for this group was 190, which was in the February referral information. Means and standard deviations for the referral groups are provided in Table 11.

The highest score overall for December, which was 105, was found in the No Referrals group. The lowest score in December, which was 23, was found with the Low Frequency group. In January, the highest and lowest pretest scores (i.e., 105 and 23) were found within the No Referrals group. The minimum posttest score for February, which was 21, was found with the No Referrals group. The maximum score, which was 108, was found with this group as well.

Table 11

Average Test Scores by Referral Group

Group	*M (SD)*	95% CI	
		LL	*UL*
December			
No Referrals	70.75 (18.19)	68.20	73.30
Low Frequency	58.33 (19.78)	37.58	79.09
High Frequency	NA	NA	NA
January			
No Referrals	70.98 (18.24)	68.38	73.59
Low Frequency	64.09 (18.14)	51.91	76.27
High Frequency	47.50 (2.12)	28.44	66.56
February			
No Referrals	74.15 (18.07)	71.56	76.73
Low Frequency	51.43 (15.82)	42.30	60.56
High Frequency	41.00 (NA)	NA	NA

Note. CI = Confidence Interval; *LL* = Lower Limit; *UL* = Upper Limit; NA = Not Applicable.

A *t* test was required for the December referral groups because there were no High Frequency groups available for statistical comparisons. The procedure generated no significant differences ($t(202) = 1.64$, $p = .102$). Those who had no referrals scored similarly to those who received a low frequency of referrals. The one-way ANOVA for the January groups had a *p* value greater than .05 overall ($F(2, 201) = 2.35$, $p = .098$). Pairwise comparisons between the groups corroborated this lack of significance. All three referral groups scored similarly on the pretest, even when the referral information

had changed. The one-way ANOVA for the February referral groups, however, yielded a *p* value of less than .001 overall. Further inspection with the *t* test for independent samples was required because the one-way ANOVA was unable to calculate a group mean for the High Frequency category. The procedure displayed a *p* value of less than .001 between the No Referrals group and the Low Frequency group ($t(202) = 4.58$). When referral data was reported at the end of the study, students in the No Referral group had a significantly higher posttest score than students in the Low Frequency group.

Interpretation of Finding 2.1. Based on the trends within the referral data, possible explanations were found as to why there were significant differences within the referral numbers. Higher referral numbers toward the latter months would have been typical within the school system. This could be attributed to the fact that December had fewer days of school than the other months. The testing schedule of the school was a consideration as well. With the Georgia Milestones in April, there was increased pressure on the teachers and students to understand what was being taught in math. There was more instructional time spent on getting ready for the test and less flexibility in instructional choices. This lack of flexibility could have affected the behavior and performance of students over time.

In terms of the research question associated with Finding 2.1, the results tell us how the referral rates of students corresponded with pretest and posttest scores. It was revealed that differences in test scores were not significant for every month the study took place. This indicates that the dynamics of classroom behavior may change throughout the school year, which in turn could have an impact on performance and behavior within the classroom environment. Significant differences in referral rates do

not imply significant differences in test scores. When there are students who receive referrals in class, it cannot be assumed that they perform badly on tests. If that was the case, then all months would have shown statistically significant differences for the pretest and posttest results.

Data Mixing for Finding 2.1. The goals and behavior patterns of students varied for each grade level. In the focus group for sixth grade, there were a few students who had their behavior trackers in their agenda signed for misbehavior. These behavior trackers were used in the PBIS punishment system. Once the tracker was filled, the students were "written up" via referral. None of the students in the focus group had received referrals at the time. Student 5a commented that one person outside of the focus group had received a referral. There were two students, Students 2a and 5a, who said they needed to "pay attention more" in class. Student 5a had his agenda tracker signed the most out of the group. Student 2a preferred not to comment about whether or not his agenda had been signed. There were three students who had their agenda signed two or fewer times while having goals related to grades. Student 4a had his tracker signed two times, and he wanted to make a "90 or higher in math" during the 9-weeks period of the study. Student 3a had her tracker signed one time for "not paying attention." Her goal was to "study more." Student 6a did not have any instances where her tracker was signed, and he wanted to receive an "85 or higher" in math. Another student, Student 1a, did not have his tracker signed, but his goal was to "go easy" on reading during class time because it was a distraction.

> Student 5a – "There's only one person who got written up."
>
> Student 6a – "I haven't even got my tracker signed."

Student 4a – "I got my tracker signed…First nine weeks, I forgot my agenda on the second or first week we were there at the school. Second was…I left my…(whispers)…I left my book in the hallway. I…my bus was…like broke down…so we had to get another bus. I had to hurry to class."

Student 3a – "I had it signed once in the first nine weeks because I forgot my Social Studies binder. I wasn't even paying attention."

Student 5a – "I got mine signed like five times in the first nine weeks, and then at the second…seven. Now…well, they're not seven…It was like four…five."

Out of the students who participated in the seventh-grade focus group, there was one who had said he received a referral. According to Student 2b, he had at least one write-up for "fighting." He said that ever since he received coins, the amount of times his tracker was signed decreased. The other students tried to avoid write-ups as much as possible. Students 1b and 5b said that they did so because they were concerned about how their family members might act if they were to "find out" what happened. Student 1b added that he did not have a signed behavior tracker. Student 6b did make comments concerning her family, but it was unclear as to how her comments were connected to her behavior. Students 3b and 4b tried to avoid misbehavior because of their teacher. Student 3b was thought of as the "teacher's pet" of the group. All of the students who participated in the focus group stated personal goals concerning grades. Student 3b had the highest average, which was 100. Most students wanted to make or maintain an A in math. The exceptions were Students 4b and 2b. They had the lowest scores out of the group. Student 2b wanted at least a high B, but Student 4b just wanted to pass.

Student 2b – "Before coins, I always used to have my agenda signed."

Student 3b – "Yeah. Um, I kinda try to be a teacher's pet."

Student 1b – "I haven't got my tracker signed this year."

> Student 4b – "I try so hard not to get it, 'cause…I don't wanna get in trouble…like, I just can't. I will get deeply in trouble, so I don't do it."
>
> Student 5b – "I don't have any at the moment. And then, like, I don't ever get it signed because I'm the oldest in my house. Like, the oldest child in my house, so I have to set an example. I have a lot…Like, my dad he works on (place. occupation removed.). So, he's *really* strict with things like this. So, if you get in trouble, there's something coming for you."

The eighth graders had the most referral numbers overall. There were multiple students in the focus group who reported that they received referrals, and two had in-school suspension (ISS) during the school year. Student 1c reported that he received a referral for "fighting." He thought that boys were more likely to receive referrals than girls, even though there were girls who misbehaved in class. The other boys in the group, Students 2c and 4c, agreed that the treatment was "unfair." These two students received the most amount of referrals within the focus group. They both had around seven write-ups, and both of them experienced ISS. Both agreed that there were students who did "act better" because they were able to receive points and coins when they did behave. Student 5c, who had received no referrals, was of this opinion as well. Student 4c stated that the tokens helped him avoid ISS, but he also said that Student 2c had more ISS during the study. Like Student 5c, Students 3c and 6c did not report any referrals for themselves. All of the students had personal goals that pertained to grades. Student 2c stated that he wanted to make sure he was "not going to sleep" so that he could receive an A in the class. Student 3c did not "pay attention" in class either, and she wanted to restore her B in math. Student 1c wanted to maintain "a high average," which was at least an 89. Students 4c and 5c wanted to pass the class. Student 6c did not explicitly elaborate on her goals, but she did mention along with other students how she was looking at her grades over time.

> Student 2c – "Like, it—it makes some of the students act more, eh, like…It makes s—some of the students that were like terrible before the points and coins act better like when they start getting points and coins 'cause they—they'll actually strive to get *something*. Like—"
>
> Student 4c (talks while Student 2 talks.) – "Other than ISS."
>
> Student 2c – "I got…One for, uh, out of area. I got one for this—like, four for like, um, disrespect…I got…one…for, it's like I was joking on somebody, and I got in trouble for it. Uh, it's about…There was this one for …It was another one. I forgot what it was. I—I just got it. Like, last week. Like, seven."
>
> Student 3c – "For eighth…I've—I don't have any."
>
> Student 4c – "I—I basically…like, the way it was in that class, it was—it was kinda…uh, unfair. I mean, and—and it was kinda…like somewhat like all the guys were the only ones getting in trouble, and like…We—We were doing the same thing the girls were…"
>
> Student 1c – "Because, like, you don't see the girls nearly getting in as much trouble. So, for instance, we had like the assistant principal come down. And, he came down like three times. So, yeah, he talked strictly to the guys."

Research Question 2b: To what extent are the test scores influenced by classroom achievement goals?

The purpose of this question was to assess students' abilities to determine their achievement status in math, set math goals, and meet math goals. Students were grouped according to achievement goal status. The groups were determined by whether or not they passed the pretest and posttest for the study. Students who passed all tests (i.e., the pretest and posttest) met goals for math, and students who did not pass both tests failed to meet goals for math. If a student only passed one test, they did not meet the goals for math. Parametric testing was required for the test scores. A series of *t* tests were executed in order to see if there were significant differences in test scores between the groups. The pretest and posttest scores were analyzed within separate *t* tests.

In addition, students were grouped by goal orientation. The goal orientation questionnaire was disseminated to students in order to determine their goal orientations during the pretest and posttest. There were five items within the questionnaire (see Appendix B). It contained the following topics:

- Students' opinion about math class
- How well students did on math tests
- How students thought they would do for the quarter
- Personal goals for math class
- How long it would take to complete the goals listed

The last 2 items had 2 sub-items, which meant that students who filled out the questionnaire in its entirety had a total of 7 responses. These responses were then reduced in order to create nominal variables from the information. This was necessary in order to incorporate the information within statistical procedures. Because the same questionnaire was administered twice, there were 14 variables created from the information. Seven were from the time of the pretest, and the other 7 were from the posttest dissemination of the questionnaire.

After the data were collected from the questionnaire, one-way ANOVAs were performed on each item. Some items required the use of the Kruskal-Wallis *H* Test along with parametric testing because there were instances where test score data violated homogeneity even after transformations occurred. Specific groups for each variable were generated by the researcher in order for quantitative analyses to be performed. Furthermore, factorial ANOVAs were produced to determine if there were any significant main effects and interactions between achievement goals and classroom behavior. Any

influences found from these procedures are discussed within the interpretation section. The significant finding is as follows:

> *Finding 2.2: Students showed significant differences in test scores when grouped according to achievement goal status and various goal orientations.*

For the achievement goals variable, those who passed both the pretest and posttest for math consistently outperformed students who did not meet this criterion for math goals. Data from the goal orientation questionnaire variables displayed significant differences within the items pertaining to how well students did on tests, how well they thought they did in the quarter, what their first personal goal was, and how long it took to complete the first goal. Students who said they did great or good on tests often scored higher than those who said they did badly on tests. Those who said they would do great in the quarter scored higher on the pretest than those who said they would do okay. Students who said they wanted to make an A as their first goal scored higher on the pretest than students who said they wanted to make a B. On the posttest, those who gave a yearlong time frame for their first goal scored higher than those who gave themselves a month to complete it; however, they scored similarly to the rest of the groups within the data collected.

- Result 2.3: The students who did not meet goals for math scored significantly lower than the students who did meet goals for math.

This result involved the use of an achievement goals variable with two groups: Did Not Meet Goals for Math and Did Meet Goals for Math. The first group scored lower than a 70 on their pretest and posttest. Scoring a 70 or higher on both the pretest and posttest designated a student as Did Meet Goals for Math. Students who only passed

one test were placed in the Did Not Meet Goals group. There were more students within the first group (n = 118) when compared to the second group (n = 86).

The achievement goals variable was used with pretest and posttest information. The pretest showed that students who did not meet goals for math had a lower mean on the pretest (M = 58.94, SD = 14.62) than the posttest (M = 62.38, SD = 17.78). For those who did meet goals, their group mean for the pretest was similar (M = 86.08, SD = 8.66) to the posttest (M = 86.01, SD = 9.43). In either situation, those who met math goals had higher scores on average than those who did not meet math goals. Homogeneity was violated for the untransformed variables for both tests ($p < .001$). It was upheld for the transformed pretest ($p = .886$) and posttest ($p = .071$).

The t tests for the achievement goals showed similar trends for the untransformed and transformed variables; therefore, the untransformed versions of the variables were reported since they most closely resemble the data. For the pretest, significant differences were found between students who did meet goals for math and students who did not meet goals for math ($t(194.70) = -16.57$, $p < .001$). Those who did not pass both the pretest and posttest had pretest scores that were significantly lower than those who did. The posttest results between the groups indicated that significant differences still existed between them ($t(186.49) = -12.26$, $p < .001$).

- Result 2.4: The students were able to accurately determine how well they would do on the pretest and posttest.

The goal questionnaire contained an item that asked students how well they did on math tests. For the pretest, four response groups were generated: Great, Good, Okay, and Bad. There were 21 students who did not complete the questionnaire for the pretest, and there were 35 who did not complete it for the posttest. The Good group had the most

student responses (n = 74) and the least amount of responses were found within the Bad group (n = 20). The posttest information showed that the Good group was still in the lead in terms of responses (n = 67). The lowest amount was found in the Bad group (n = 27).

The highest group mean for the pretest was in the Great group (M = 78.56, SD = 16.47), and the lowest group mean was in the Bad group (M = 56.77, SD = 20.39). This means that students who thought they did great on math tests scored the highest on average, but students who thought they did badly on math tests scored the lowest on average. Those within the Good group (M = 71.81, SD = 17.46) only scored lower than the Great group. The Okay group (M = 68.93, SD = 16.94) scored above the Bad group, but it did not score above the other groups. Therefore, students' test scores increased as their expectations for performance increased.

For the posttest, the Great group (M = 77.64, SD = 11.41) had the highest group mean. Students who gave a very positive response about how well they did on tests scored higher than the other groups. The Bad group still had the lowest group mean (M = 58.89, SD = 20.84). Students who thought they did bad on tests, even after treatment, scored lower than the other students. Again, the Good group (M = 74.52, SD = 17.39) scored higher than all groups except the Great group. The Okay group (M = 68.63, SD = 18.69) only scored higher than the Bad group. After treatment, students who expected to do well on tests were still accurate about their expectations. Homogeneity was not violated by the pretest (p = .448), but it was violated by the transformed and untransformed versions of the posttest scores ($p < .05$). This means that the variances were assumed to be equal for the pretest means, but this was not the case for the posttest means.

For the pretest, one-way ANOVAs for the data showed significant results overall ($F(3, 180) = 7.80$, $p < .001$, partial $\eta^2 = .115$). Pairwise comparisons showed that the Great group ($p < .001$) and the Good group ($p = .003$) scored significantly higher than the Bad group. Those who thought they did great on tests had a significantly higher score than the ones who thought they did badly on tests. Other notable differences had to do with the Okay group. The Great group scored higher than the Okay group ($p = .05$), but the Bad group scored lower than the Okay group ($p = .04$). One-way ANOVAs for the posttest resulted in significant differences between the groups overall ($F(3, 166) = 7.16$, $p < .001$, partial $\eta^2 = .115$). Within post-hoc testing, there were two groups that scored significantly higher than the Bad group. Students in the Good and Great groups outperformed the Bad group ($p < .001$). This means that for the pretest and posttest, students were accurate judges of how well they would do on tests.

Transformed data showed the same trend for the scores; however, homogeneity was still violated for the posttest after the transformation. Therefore, Kruskal-Wallis *H* Tests were performed for the data. The same significant differences in the ANOVAs for the posttest were observed in the non-parametric procedures for the untransformed and transformed variables. In sum, the pretest and posttest data showed that students who had higher expectations about performance scored significantly higher than those who responded negatively about their expectations ($H(3) = 16.69$, $p < .001$).

- Result 2.5: Students who thought they would do great on the math tests for the quarter scored significantly higher than students who thought they would do okay.

This result included groups from the third questionnaire item, which pertained to how well students thought they would do for the quarter. There were six response groups

for this item: I Don't Know, Same, Great, Good, Okay, and Bad. For the pretest, there were 21 students who were missing data for the calculations. The largest amount of students was in the Good group (n = 112). The least amount was observed within the Bad group (n = 4). The missing data amount for the posttest was higher at 37. Students had the most responses within the Good group (n = 105), which was similar to what the pretest results showed. There was a difference in terms of which group had the lowest number of responses. For the posttest, that was the I Don't Know group (n = 4).

The highest mean found in the pretest was in the Great group (M = 76.15, SD = 18.92), and the lowest group mean was in the I Don't Know group (M = 54.40, SD = 15.69). Students who thought they would do great for the 9 weeks of the study performed better on average than those who were not sure about how they would do. The difference of these means was at 21.75, which was a large amount of difference between the groups. On the posttest, the highest group mean was in the Same group (M = 84.80, SD = 7.98), but the lowest mean was still found in the I Don't Know group (M = 57.50, SD = 24.84). Students who thought they would perform about the same at the end of the 9-week period did better on average than those who were unsure about how their performance would be in the future. Compared to the pretest mean difference, the posttest was higher at 27.30. Homogeneity was not violated for the pretest (p = .221) or the posttest (p = .107).

One-way ANOVAs for the pretest information indicated significant differences between the groups ($F(5, 178) = 2.37$, p = .042, partial η^2 = .062). Within pairwise comparisons, a meaningful difference was found specifically between the Okay and Great groups (p = .079). Those who thought they would do okay in the quarter scored lower

than those who thought they would do great for the quarter. For the posttest information, no significant differences were found among the groups (p = .260). This means they scored similarly for the posttest.

- Result 2.6: Students who wanted to make an A for their first goal scored significantly higher on the pretest than students who wanted to make a B.

The questionnaire also had items that pertained to personal goals. Students had to list 2 personal goals for the pretest and posttest data. The first goal had 21 students who did not respond in the pretest and 36 students who did not respond in the posttest. For the second goal, there were 27 students who had incomplete or missing responses in the pretest, and there were 37 students who did not respond in the posttest. The groups for these goals were as follows: Make an A, Make a B, Make Good Grades, Pass, Do Better, Pay More Attention, Study More, and Understand Math. For the first goal, the pretest showed that Make an A (n = 75) and Understand Math (n = 2) contained the highest and lowest student amounts, respectively. The posttest indicated that Make Good Grades (n = 77) had the most, and Understand Math still had the least amount (n = 3). The second goal showed that Make Good Grades had the highest amount of students (n = 41), whereas Study More had the least amount of students (n = 9). The posttest data for the same goal had Make Good Grades as the group with the most students (n = 50), but the lowest group for these test scores was the Make a B group (n = 4).

On the pretest for the first goal, the highest average was found within the Make an A group (M = 76.71, SD = 16.41). Students scored the lowest within the Make a B group (M = 59.26, SD = 18.28). This was seen in the posttest scores for the Make an A (M = 77.45, SD = 14.96) and Make a B (M = 64.30, SD = 15.25) groups. Therefore, students who had Make an A as their first goal had a higher score than the students who had Make

a B as their first goal. For the second goal, the Make a B group had the highest mean score on the pretest (M = 89.75, SD = 4.92) while the Do Better group had the lowest group mean (M = 64.53, SD = 26.53). The posttest had a different result whereby the Understand Math group had the highest group mean (M = 76.78, SD = 15.10) and the Study More group had the lowest group mean (M = 61.44, SD = 18.01). Homogeneity of variances was violated for the first goal on the posttest results (p = .006) and the second goal on the pretest results (p = .013). Transformations did not correct this issue.

One-way ANOVAs indicated that there was a significant difference overall when the pretest was sorted by the groups for the first goal (F(7, 176) = 2.78, p = .009, partial η^2 = .099). Within pairwise comparisons, those who made an A scored significantly higher than those who made a B (p = .001). No significant differences were found for the posttest, which required one-way ANOVAs as well (p = .152). This means the test scores were similar for the groups.

For the second goal, one-way ANOVAs showed no significant differences for the pretest (p = .233) and posttest (p = .166). Students within the goal groups had group means that were not statistically different from one another. The pretest required the Kruskal-Wallis H Test to be performed, and those results (H(6) = 8.62, p = .196) corroborated the lack of significance found during parametric testing. Overall, those who made specific goals about grades scored higher, on average, when their expectations were higher.

- Result 2.7: Students who said it would take a year to complete their first goal scored higher on the posttest than students who said it would take a month to complete it.

This result pertained to the fifth questionnaire item, where students described the amount of time it would take to complete their goals. Since there were 2 goals, each goal had a separate variable that dealt with time. Each variable, however, ended up having the same groups so that sufficient comparisons could be made. On the time variables for the first goal, there were 21 students who had missing information on the pretest, and there were 35 students who had missing information for the posttest. The time variables for the second goal had 29 students who did not respond to the pretest and 39 students who did not respond to the posttest. There were nine groups for each time variable: Other, I Don't Know, Not Long, A While, Week, Few Weeks, Month, Nine Weeks, and Year. For the pretest, the first goal had the highest amount of students within the Nine Weeks group (n = 49). The least amount of students were found in the A While group (n = 3). The posttest showed the highest amount of students in the Not Long group (n = 43), but the lowest amount of students was still in the A While group (n = 2). The second goal had the most students in the Not Long group (n = 58) during the pretest. The least amount was found within the A While group (n = 4). The posttest showed that the most students were still in the Not Long group (n = 49), but the least amount of students were in the Week group (n = 5).

The first goal had the highest pretest average in the A While group (M = 85.33, SD = 9.07) despite it being the smallest group. The lowest group mean was within the Week group (M = 52.83, SD = 26.98). Students who said it would take a longer, more indefinite amount of time to complete a goal had a higher score than those who specifically said it would take a short, definite amount of time. The posttest scores for the same goal listed the Year group as the one with the highest group mean (M = 75.91,

SD = 15.82), but the A While group had the lowest group mean (*M* = 42.00, *SD* = 2.83). Those who thought their goal would specifically take a year to complete scored higher than students who thought their goal would take a while.

Like the first goal, the second goal had the Week group scoring the lowest for the pretest (*M* = 55.45, *SD* = 17.35). The highest group mean was found in the I Don't Know group (*M* = 76.89, *SD* = 14.60). Again, an indefinite time frame scored higher than the more definite time frame of a week. The posttest showed that the I Don't Know group scored the highest again (*M* = 83.89, *SD* = 10.02), and the Week group scored the lowest for a second time (*M* = 62.00, *SD* = 26.53). The same trend found for the pretest scores for the second goal was found in the posttest for the same goal. Homogeneity was not violated for the variables in this result ($p > .05$).

One-way ANOVAs for the first goal indicated that no significant differences were found for the pretest ($p = .132$). The groups scored similarly during the first half of the study. There was a significant difference between the groups for the posttest ($F(8, 161) = 2.09$, $p = .040$, partial $\eta^2 = .094$). During post-hoc testing, there was a meaningful difference between students in the Year and Month groups ($p = .071$). Students who said it would take a month to complete their goals had a lower score than those said it would take a year.

Data for the second goal did not show any significant differences for the pretest ($p = .200$) and the posttest ($p = .361$). The groups, when sorted for the second goal, were statistically alike in terms of their test scores. This means that a difference existed for the first goal that was not found in the second goal.

Interpretation of Finding 2.2. There are aspects of the learning environment that may help to explain the results for Finding 2.2. Students received instruction based on the Georgia Performance Standards (GPS), and this did not change for the school year. They did very well on tests when they had high math goals and personal goals based on these academic standards. Throughout the study, a positive orientation about math was encouraged within the learning environment. The students' test performance reflected this. Students who make goals that last 9 weeks or more would be setting mastery goals. As stated in Chapter 1, mastery goals are long-term goals about how students will use their abilities over time. The commitment required to achieve these goals would have been more sustainable than what was required for performance goals, which were short-term achievement goals that were set during the academic year. This would provide a reason for why students who had yearlong goals did better on the posttest than students who had goals that lasted approximately a month.

As far as the research question is concerned, the results pertain to how GPS achievement goals and personal goals may help to affect pretest and posttest performance. The tests did not measure how much students have learned as a whole, but they did help to measure how many questions students got right given the parameters set by the Common Core GPS standards in math. From these standards, teachers created classroom goals concerning achievement status. Students who did not do well within the GPS Framework had a great amount of trouble with the standardized tests given to them. Furthermore, there were items on the goal orientation questionnaire that showed important statistical differences. From these noteworthy differences, it can be concluded that students were able to successfully (a) identify where they were in terms of how well

they did on math tests (b) match their pretest performance with their responses about their quarter performance and their first personal goal, and (c) match their posttest performance with their responses about how long it would take to complete their first personal goal.

Data Mixing for Finding 2.2. As mentioned in Finding 2.1, students had different goals. Most of them within the focus groups pertained to grades and academic performance. The students were able to assess their math performance, math ability, and testing preferences for themselves. In sixth grade, there were two students who named the specific grade they would like to have for the 9-week period of the study. Student 4a wanted a "90 or higher" in math. He kept receiving an 89. Student 6a wanted to have "at least an 85 or higher." She received anywhere from an 80-82 at the end of each marking period; however, she stated that her math grade was "getting better" than the grades she had for other subjects. On a coordinate graph assessment, she stated she was the second to finish. Finishing quickly was a good thing for her in terms of performance, and she added that she found some of the math tests to be "fun." Students 1a, 2a, 3a, and 5a did not discuss their overall average for math. Student 5a did say that he scored anywhere from an 85-90 on class activities, but he did not share his overall average. His goal was to pay attention more because he only paid attention one-third of the class instructional time. This was more than the amount of minutes Student 2a paid attention in class. His goal was the same as Student 5a, and he lost focus after approximately 20 minutes of instructional time. Student 1a had an attention-related goal where he wanted to lessen the amount "reading" he did that did not pertain to math instruction. Student 3a wanted to "study more," which was less specific than the grade-related goals that were mentioned

by Students 4a and 6a. She felt there were "consequences" in terms of school performance when students did not study.

> Student 1a – "Um…Well, I read a lot in class, so I want to stop doing that." (everybody laughs.)
>
> Student 2a – "Pay attention more to the lesson."
>
> Student 3a – "To study more because there's times where I'd rather do other stuff than study. But there's always consequences for when you're not studying about it, especially if you're struggling on it. So, I have to study more."
>
> Student 4a – "I'd like to get a 90 or higher in math. I keep getting an 89."
>
> Student 6a – "I'd probably say my grades 'cause I keep getting like an 80, an 81, or an 82 every time. But, my math actually is getting better than my other grades."

All the students in the seventh-grade focus group reported their overall grades for math. Their goals identified grades that they wanted to attain or keep during the study. The students scored As and Bs for overall grades. Student 1b said he was "terrible at math" with an 88 average, yet he also said that he was good at determining information about triangles. He felt that a good grade would be "at least above a 90." He said that this goal would take until the end of the year to complete. Students 2b and 4b had Bs too. Student 2b had a "middle B," and he wanted at least a high B moving forward. He felt he made "pretty good grades" overall. Student 4b said that she had an overall average between 83 and 84. As long as she passed the class, she was happy with her grade. She said that measuring triangle angles with the protractor was the "fun" part about math instruction. The other students agreed with her. She said that she did not like that she had to use Scantrons during the tests. Students 3b, 5b, and 6b had As for their overall grade. Student 3b had the highest average with 100, and he consequently got along well

with the math teacher. His goal was to have "above a 96" in math, which he had already met by the time he was interviewed. Student 5b wanted to maintain her A, which was "above a 95." As long as it was higher than a 95, she was happy with the grade. Like Student 5b, Student 6b had the goal that her grade would be higher than a 95. She had a 98; therefore, she had met her goal by the time she was interviewed. The students' tests were mostly multiple-choice, but they had to make sure to show their work in order to get full credit for their answers.

> Student 1b – "I'm, like…terrible at math, so I want to at least improve."
>
> Student 2b – "I'm at a B. Like a…middle B."
>
> Student 4b – "I got an 83 (Student 5 laughs.)…or like…an 84. I just wanna pass, okay!" (everyone laughs.)
>
> Student 5b – "Uh, stay above a 95 average…so just keep the grades up."
>
> Student 6b – "I want to get above 95…'cause I get something if I do."

In eighth grade, most of the students listed specific grades that they wanted to receive. Student 1c wanted to "maintain a high average." For him, this meant an overall grade of 89 or higher. He was not sure of his overall average at the time of the focus group. He stated that the tests they took were mainly from USA Test Prep, which was computer-based. Student 2c added that there were "written tests" they received as well. In other words, there were multiple-choice tests and open-ended tests that they took during the school year. Student 2c said that his goal was "not going to sleep" so that he could get an A instead of a B. Student 3c also said she did not "pay attention" during class. At the time of the interview, she had a 77 average. Her goal was to get a B, which she had in the past. She added that students did better on the open-ended written tests. Students 4c and 5c had the same goal: to pass the class. Student 4c felt he was "bad" at

math. He thought that the tests given were "hard ones," especially the tests within USA Test Prep. He said that many people failed the ones they had to take on USA Test Prep. He had failed a test, and it dropped his average from an 89 to a 74. Student 5c agreed with him about the USA Test Prep assessments, but she then stated that she liked the "written tests" because she knew that the material had been covered during instruction. Student 5c said she was "not very good" at math. She said her performance in math was similar to Student 4c. She did not know her overall average. Student 6c had an overall grade of 77, which was the same math average given to Student 3c. She said she performed about the same as Student 5c. Her comments were grade related, but she did not go into detail about the specific goals she set at the beginning of the study.

> Student 1c – "To maintain a high average."
>
> Student 2c – "Uh, not going to sleep. Like, stay awake so I can actually like get a...an A in that class. I have a B in that class."
>
> Student 3c – "I have a 77 right now because…of the lectures and everything…and the work that she just gives us after the lecture. I don't even know what she said."
>
> Student 4c – "Um…to pass the class (laughs nervously.)."
>
> Student 5c – "Um…to pass. I'm not really good at math like him (refers to Student 4.)."

Research Question 2c: To what extent are the test scores influenced by students' motivational preferences?

For motivational preferences, a nominal variable was created from responses within the motivational preferences check sheet. In this check sheet, students selected what they liked as a source of motivation: Points, Coins, Both, or Neither. It was administered during the posttest (see Appendix D). Parametric testing was required for the motivational preferences variable, which was discussed within Chapter 3. One-way

ANOVAs were run in order to determine if significant differences existed for the pretest and posttest. This involved ANOVAs for scores within all six Treatment Combinations. After these procedures were conducted, two-way and three-way factorial ANOVAs were performed in order to (a) observe if significant main effects and interactions were found when nominal variables for the treatment combinations were paired with motivational preferences, (b) determine if classroom behavior and achievement goals had any significant main effects and interactions on test scores when paired with motivational preferences, and (c) investigate the extent to which demographic variables had any main effects or interactions on test scores when paired with motivational preferences. The results for the treatment combinations are reported within the results sections along with the one-way ANOVAs. The possible influences found for motivational preferences are reported in the interpretation section. Transformed versions of variables were not necessary since homogeneity was not violated for the untransformed versions of the variables. The finding available within this section is as follows:

> *Finding 2.3: Students who preferred coins displayed significant differences in test scores within the coins groups for the first and second half.*

Within one-way ANOVA procedures, there were no significant differences found in test scores when grouped solely by motivational preferences. This changed when the information was grouped by motivational preferences and the treatment combinations. A significant interaction was found between students' preferences and their arrangement during the token treatments. Specifically, this was found for those who chose coins as a motivational preference. They had the lowest test score within the coins group for the second half (i.e., A2), but they had the highest test score in the control group for the second half (i.e., C2).

- Result 2.8: Students who preferred coins scored the lowest in the coins group for the second half, and they scored highest in the control group.

This result included the motivational preferences variable, which had four groups: Points, Coins, Both, Neither. There were a total of 42 students who did not turn in a check sheet. Students preferred points the most ($n = 55$) with the Both preference coming in second ($n = 46$). The smallest group preferred neither ($n = 23$). The pretest results showed that those with Neither had a higher mean ($M = 73.87$, $SD = 15.02$) than the other groups. The posttest showed that the Both group scored the highest ($M = 72.09$, $SD = 21.18$). The lowest mean for the pretest was within the Points group ($M = 66.22$, $SD = 18.34$), and the lowest mean for the posttest was in the Coins group ($M = 70.46$, $SD = 17.62$). Homogeneity was not violated for the test scores within this result ($p > .05$).

One-way ANOVA procedures for motivational preferences and the pretest scores showed no significant differences overall among the groups ($F(3, 158) = 1.75$, $p = .159$, partial $\eta^2 = .032$). Similar scores were found for all groups. In addition, this was seen for the posttest scores ($F(3, 159) = .07$, $p = .978$, partial $\eta^2 = .001$). No group had major differences in scores within pairwise comparisons. For this reason, two-way factorial ANOVAs were executed to look at possible significant main effects and interactions between the treatment combination groups and motivational preferences. For all procedures with the nominal variable for the treatment groups, there were significant main effects ($p < .001$). These main effects were discussed within Findings 1.1 and 1.2. A significant interaction was found between motivational preferences and the groups within Treatment Combination 4 ($p = .037$). Within the Coins group for the second half, students who chose coins had lower scores than all students in the group. When students were grouped by preference, those students who had coins as a preference scored the

lowest within the coins group for the second half. They scored highest in the control group. This same trend was observed in Treatment Combination 3, even though the interaction with motivational preferences was not statistically significant ($p = .092$).

Interpretation of Finding 2.3. There are aspects of the school environment that can possibly explain the results generated in Finding 2.3. Before and during the study, students received PBIS rewards and feedback strategies that were outside of the treatment intervention. The other systems used were previously documented in Chapter 3. This variety, along with the options given to the treatment groups, gave students the chance to experience different types of motivation regardless of their personal opinions at the time. Since students had a variety of options available to them outside of treatment, it took away the amount of reinforcement value that could have been placed on the tokens themselves. This means that the token system was not as fully integrated into the school environment as it could have been.

In terms of the research question, the results help to determine the extent to which motivational preferences play a role in test performance. Students' preferences for a particular reward or instructional strategy did not guarantee success in terms of pretest and posttest performance. Conversely, their dislike for a reward or instructional strategy did not imply failure in correctly completing tasks during the pretest and posttest. It is possible that the actual options available to students within the treatment may have been different from students' initial expectations at the beginning of the study, or students became more indifferent to the treatment strategies as a result of their exposure to them over time. Because of the significant interaction found within the results, it is also

possible that students who really liked the tokens did not perform as well when they already received what they wanted.

Data Mixing for Finding 2.3. Because the teachers informed their groups about whether or not they could receive tokens during the study, the students developed their perceptions based on that information. Two out of three teachers stated that there were participants in the control group who wanted the tokens. All teachers stated that even though there were students who wanted the tokens in the treatment groups, there was no guarantee that performance and behavior improved for every student who wanted them. The sixth-grade teacher stated that the majority of students in the "nothing group" (i.e., control group) were "more motivated" and "in it for themselves." There were a few students who may have benefitted from the use of tokens, but she felt the majority were mainly focused on "grades." There were students from the control group, however, who complained to her about not "getting anything." She was of the opinion that the students in the treatment groups "like the coins." She saw instances where the students were helped and harmed by this preference. On the one hand, there were students who regularly did "nothing" in the treatment groups who worked hard to get the tokens. As previously stated within Finding 1.2, the downside was that students with short attention spans used the coins as a "distraction" to avoid paying attention to instruction. She did expect students to lose the coins over time.

> Teacher 1, Comment 1 – "Yeah, the nothing still stays the same. Yeah, they're like 'Why aren't we getting anything?' And I have heard complaints on that."

> Teacher 1, Comment 2 – "But, you know, we have kids that come in and do absolutely nothing. And a few of them have started working to get the tokens. For some of them, that's just not going to help."

> Teacher 1, Comment 3 – "Like I said…we've only been doing this for…this is…going into our third week. I do anticipate them losing the coins and not being able to keep up with them until they get to their goal of 20. And I think, I mean I have told them to put them in their pencil pouch. I anticipate them losing them. Those who are not keeping up with their sticker or something like that. Yeah…'Cause I know when we did the passports and we were punching…kids would lose their passports. So, you know…kids lose stuff. So I know they're…they're going to. But that's the…the big problem that I see with it."

The seventh-grade teacher wished that she could have used the tokens with all of her students, but she made sure to follow study protocols. Even though the control groups knew they were the "nothing group," they still were "hoping" to receive them during the study. The teacher used strategies such as "praise" and the PBIS reinforcement systems already in place to mitigate the fact that they were not being awarded coins and points. She perceived that the control groups were the "best ones" as far as behavior was concerned. She felt that the students in the treatment groups "tend to like the coins." Even though they preferred the coins, they were "harder to keep up with" over time. This trait within the coins system did not stop the points group from wanting to try out the coins after the switch. There were students who had lost coins and reported it to her. She stated that token systems could be used with different PBIS tiers for "motivation purposes."

> Teacher 2, Comment 1 – "We've had kids before that didn't care if they got their trackers signed, but these kids do for the most part. Especially the classes that are doing the nothing. They actually both are pretty well-behaved classes. Um, probably two of my best ones…And they don't like to get their agendas signed."

> Teacher 2, Comment 2 – "But, the coins…a lot of them…the boys especially don't have like a pencil pouch or things like that, so they kinda—they keep them in their locker. But they're kinda harder to keep up with. I've had a couple of boys tell me that they've lost some."

> Teacher 2, Comment 3 – "Um, you could use it…I think that maybe your Title teachers. We have Title teachers that teach Title math and reading…and that's your Tier 2 kids…um…I think, and some of those are Tier 3 too that fall in there. But they could use it in their classrooms because I think, like for motivation purposes they might could use it to try to get the kids to be motivated. That, to me, would be a place where a ticket system—or, token system—might very…might work well as far as small group. But otherwise, I think large group…Tier 1."

At the beginning of the study, the eighth-grade teacher noticed that the treatment groups found the token systems "very motivating." This motivation started to decrease after the third week of the study. She saw that students went back to their typical behavior and performance patterns, especially at the latter part of the study. She felt that the control groups were "intrinsically motivated," so they did not have a strong preference in terms of the tokens. Overall, she found there was not a big change in terms of how the students acted. Those who were "well-behaved" students acted the same for their other teachers. She did state that points were more "efficient" for the eighth-grade classroom, but coins could be used for small groups. She also indicated that both systems could be "more meaningful" when used with small groups. There were students during the study who lost track of their points and coins totals. Because she kept track of the totals, she was able to remind students about how many tokens they actually had.

> Teacher 3, Comment 1 – "Um, and at the beginning it was very motivating. Closer to the end, they were kinda like, 'Yeah, we'd rather talk.'"

> Teacher 3, Comment 2 – "So then I get to do the write-ups at the end of the day, which is the same for them too. You know, they're…by the end of the day, those kids who have been misbehaving all day…The end-of-the-day teacher is generally the one that writes those kids up. So…But yes, I would say (sighs.)… I real—Yeah, I really didn't see a big change in behavior between the classes."

> Teacher 3, Comment 3 – "But even…I was gonna say with a small group, the coins might be…more realistic with a small group. And they might be more meaningful in a small group, because…kids would…They would be more aware of, 'Oh, well, she just got a point for answering a question. Even though she didn't get the question right, she at least tried and gave a good reason for it.' or, you know, tried to justify her answer or 'She got a pen—a token for coming to class with her pencil and her book today.' you know. I think they would be more aware. So, the coins might would work better, or better than they did in the class of 30 or 31. With a small group, but I still think points, as far as teachers go…as far as my preference…Just making a little tally mark when they earn a point, telling 'em what it's for. I think that's just more efficient."

Research Question 3a: What are the perceived experiences of students involved with the use of token-based systems in their classrooms?

For this study, three focus group interviews were conducted for students who participated in the study. Each focus group had 6 students. All of the interviews were based on the questions provided in the interview guide. The first interview occurred at the beginning of the study before the switching of the treatments occurred. Students in sixth grade participated in the discussion. The second interview took place around the switch point with seventh-grade students. The last interview, which had eighth graders, was conducted at the end of the study. After each interview was recorded, the researcher transcribed them and briefly read over the contents during the data collection stage. Then, the focus groups were sent to the students' parents for member checking. The interview guides were given to teachers for their records. The analysis stage required multiple readings as well as multiple coding formats. Each focus group required by-hand coding and electronic coding in *R* statistical software. As previously stated in Chapter 3, the steps for analysis were as follows:

- Through the epoche process, set aside preconceived ideas about what is said in the text.

- Determine codes and categories through bracketing.
- Identify patterns seen within the text.
- Organize the data into meaningful clusters so that themes are easily located.
- Create a structural description of the overall experiences of the participants.
- Synthesize what is revealed from the textual and structural analyses.

The first 4 steps were followed for both the by-hand and electronic versions of the transcripts. The last 2 steps were completed on the same password-protected computer used for making the electronic transcripts. A hardcopy of each focus group was available for by-hand analysis. Important key words and themes were marked in the passages, and colored pencils were used for color coding. The development of text codes and categories was an ongoing process during analysis that emerged as all the transcripts were thoroughly read. A list of codes and categories was generated from the information, and it was used to help the researcher label important experiences found in the text. The list was used for all focus groups so that consistent patterns would be easier to find.

Once by-hand analysis was completed, the information was transferred to *R* statistical software. The electronic versions had some additional markups, text codes, and color codes that were used to better organize the information. All of the transcript analyses were saved in a project file. From the information, data matrices and outlines were created in order to determine how to approach the structural descriptions of each focus group. Three structural descriptions were finalized for each grade level. The synthesized finding for the students' focus groups is as follows:

Finding 3.1: Students' experiences were affected by (a) their personal attitudes about tokens and rewards, (b) how they were awarded tokens during instruction, and (c) their perceptions about the management of tokens over time.

Noticeable trends emerged from the focus group data. Firstly, personal beliefs about tokens helped to shape present perceptions about tokens and the study procedures. These beliefs, which were mostly positive and neutral, developed over time from past reinforcement experience and present instructional experiences. Many students had a more positive past experience. Secondly, the classroom experiences of the students depended on how they received tokens. Teachers did not award behavior and grades in the exact same way. Even though the ratios were fixed, the teachers had to choose what behaviors deserved a token. Students who felt the system was unfair believed that certain students received advantages over others because (a) they performed well in math with little effort and (b) they were already liked by the teachers. Thirdly, students had different preferences in terms of token management. These preferences were not the same for each grade level. The older students tended to prefer points. Their preferences were stronger as they experienced the token systems over time. Not all students received tokens at the same rate. Their ability to keep up with tokens depended on their personal attitudes and habits.

- Result 3.1: The majority of students had either positive or neutral reactions about their experiences with tokens and prizes.

All students had experiences with token systems and other types of rewards systems. In sixth grade, the students had to recall their elementary school experiences. The experiences included receiving prizes from a prize box in kindergarten, getting candy in first grade, getting rewards in homeschool, moving clips for behavior in second and fifth grade, having cash drawings in third grade, and using a Class Dojo points system in

fourth grade. Later on, they discussed what they thought about the tokens they received for the present. When polled, four students out of the six indicated that getting the tokens mattered to them. Two students said it did not matter. This means that no student said they disliked the tokens. When asked how others felt about tokens and prizes for the study, more ambiguous answers were presented. Some figured that those who did not get much awarded to them for their behavior may see more value in it than those who received tokens and prizes more often.

Seventh grade had different past experiences with tokens, and most of them were in elementary school. In general, their experiences involved turning in fish tokens to get prizes, trading paper money for prizes, and using coins to get other items. The students had a wider range of responses. Three students stated that there were different reactions in their past experiences. Two liked their past experiences and thought that others would agree with them. There was one student who disliked her experience because she did not receive much; however, she did say that other students probably liked it because they got more. For present experiences with tokens, five students noticed positive changes in behavior when students knew they could receive tokens during the study. The noise level in the room went down, and students paid more attention in class.

Eighth graders had varied experiences as well. Overall, students had past experiences with receiving stickers as prizes, using Accelerated Reading (AR) points as tokens, using colored popsicle sticks as tokens, having cards as rewards, and using coins as tokens. Five students out of the six enjoyed their experiences. One student did not care for his experience, which included the cards. For the token experiences in the study, all of the eighth-grade students had an indifferent reaction overall. They did have their

preferences, but they felt their behavior was not affected by whether they received tokens or prizes. This was observed moreso within this grade level than the rest.

Themes for Result 3.1. One major theme found in the focus group information was that the students thought of what they received as a "good" thing. There were a few instances where students "didn't care." In sixth grade, students generally expressed positive views in terms of past experiences. Students 1a and 5a felt that being able to receive items was "rewarding" in itself. Student 1a was unsure how other students felt, and Student 5a did not have a response for how others felt within his experience. Student 2a briefly said that his experiences with moving clips was "good" and "great" overall, except when the students got in trouble. He noted that there were some students who felt the same way he did. Students 3a and 6a presented the additional view that their experiences were "good" because they felt they "accomplished" something important for the school year. Student 3a said that some students liked it and others did not care about the cash they received; however, Student 6a stated that students in her experience did not care about the treats they could receive from a system that involved, a ruler, clips, and a Class Dojo system. Yet another participant, Student 4a, felt "good" because he was being noticed for his behavior. He later recalled in a past experience with a nine-weeks rewards system that the other students were "excited" during that time, and he felt "accomplished" because he was able to receive prizes for being a top reader.

> Student 3a – "I felt like it was pretty good. Like, yay, I accomplished something. I don't have to be like…no you didn't do any of this. It's just good to know that you can do it and you can push yourself to be the best you can be."

> Student 4a – "I felt good that somebody noticed how good…my good behavior…how well I'm doing."

> Student 5a – "I mean, it was kind of rewarding, to get like a little tiny toy at the end of the week. It felt like, I got these points and I just used them on this, and with this little toy in like a day."
>
> Student 6a – "Um…It's kinda like when you go to the dentist…If you're younger and you get a prize you feel so good about it because you passed the dental…the dentist check. It's just very good in how you go through the school year and see what kind prizes you get for accomplishing things."

In terms of receiving tokens and prizes in the present, there was a wide range of reactions for the sixth-grade students (see Appendix I for prize list). The overall themes shown within these present experiences further emphasized the ideas that students "liked" and "didn't care" about the systems. These themes were similar to what was seen for the past experiences, except that students had more positive reactions in the past. Student 1a said he "liked" receiving tokens because it made the teacher mad when he banged them on the desk. He said "Yaay!" when the topic of receiving tokens and prizes came up. He especially liked the homework pass. In terms of what others students thought, he stated they did not think much of it. Student 2a felt that it "doesn't matter" unless the prizes are expensive. In other words, he felt that the tokens and prizes had to have a lot of worth attached to them in order for him to care about them. He was not sure how others felt about the tokens and prizes. Student 3a mentioned different feelings about the tokens and prizes. On the one hand, she felt like they were an "accomplishment" that she could be proud about in school. This was the same word she used to describe her past experiences. She also felt at the same time that they were just "okay." She stated that students who usually did not receive much felt "proud" of themselves, but others did not care much about it. Student 4a said that it did not change his behavior if he got tokens or not. This was because he was generally a well-behaved student. He did try his best to get the prizes for the most part, but the sticker prize was the one that he wanted to avoid. He

thought that students who were not well-behaved would celebrate if they received something. Student 5a, like Student 4a, said that he tried to get the prizes. He felt he was most likely to receive a small prize because of the small amount of tokens he was likely to receive. Similar to Student 4a, he thought that students who did not get much cared more about it than students who received a good amount of rewards already. Student 6a said that she did not care about what was received, even though she was eligible to receive all of the prizes except the headphones. Her class did not receive that many tokens and prizes because of bad behavior. Those who did receive prizes had mixed reactions. At first, they liked that they received a prize, but over time they cared less about it.

> Student 3a – "I say that some people that don't get a lot would be proud of themselves, but others would be like 'I could care less if I get this. It wouldn't really matter to me.'"
>
> Student 4a – "I think that the ones who aren't well-behaved and they do get it...I think they'd be celebrating."
>
> Student 5a – "I think that those people who get a bunch don't really care about it, and the people who have so little are like 'Yay.' I don't know."
>
> Student 6a – "Um…I don't think there's too many prizes in my classes because there are people who talk a lot…They just don't pay attention in class. When they get a prize, it's like 'Yay. I got a prize.' Then it's like 'Okay. Whatever.' Afterwords, like, they don't really care for it."

The past experiences for the seventh graders mostly supported the "good" and "don't care" feelings of the previous grade. Students 1b and 5b shared a mix of reactions about their past experiences. Student 1b indicated that his experience with Fish tokens started out fine in third grade, but the tokens became "boring" to the point where students "didn't care" about it anymore. The system lacked variety in terms of prizes, and other

students thought there was too much bias as far as who received the tokens and prizes. Student 5b stated that her past experience with a paper money system was "okay." Students who did what they were supposed to do did well, but the rest did not get a chance to earn as much. The teacher only changed prizes for one class, so that particular class had more "fun" with the experience. Students 2b, 3b, and 4b shared that they "liked" their experiences. For Student 2b, his past experience with cash was "fun" because he could receive prizes from a store and bid on items. He said that everyone liked the system because they all had a chance to receive what was there. Student 3b liked the coins system she experienced, but she did acknowledge that some students did not like it because they were not noticed as much as they would have liked. They could not trade in as frequently as others. Student 4b did mention that she had "fun." Like Student 2b, she liked the prizes she received for being good in class. She stated that all students thought it was a "good" thing. Student 6b, who was the exception to most of the reactions, stated that she did not like her Fish experience because she was not picked much during it. She did say that some students probably liked it when they received a lot from it. Those who did not, like her, felt similar to how she felt.

> Student 1b – "It was…like when we got to fourth grade, it was just getting boring, like nothing good about them anymore since…We had like a different library teacher--(name deleted.)--that just came in, so she didn't like…she didn't get anything new for us. So, it was just getting boring, so we just stopped. We didn't care about the fish anymore, but we still were on our good behavior…We just didn't turn in our fish."

> Student 2b – "It was fun 'cause we had a little store, and at the end of the year, we had to spend…They were like bids like they (Student 3b and Student 5b) did, but it was like…some like, books and stuff like that. Some store-bought toys and stuff like that, that we had to bid on. And whoever had the most, they could bid on it. So, it was pretty cool."

> Student 3b – "Yeah, I liked it, but some of the kids didn't because…they—they weren't bad, they just weren't noticed, I guess."

When compared to their past experiences, the seventh graders showed even more positive reactions for the token systems during the study. Although they did share that the prizes needed improvement overall, the main themes seen for their present experiences were that students were "quiet" and "excited" when they had chances to get tokens and prizes. Students 1b, 2b, and 3b described the "quiet" aspect of the students' reactions. Student 1b said that the students went "quiet" when the teacher talked about giving out points and coins in math class. He pointed out that the highest prize, which involved the use of headphones during independent work, was what had them stay "quiet" in class. The students wanted to use their technology for music because they rarely had the chance to do so without the tokens. Student 2b agreed that everyone was "quiet" when the teacher said she was going to give out tokens. They usually talked during class, but there was less noise when the teacher said she was giving out tokens. Student 3b agreed with the other two students about students getting "quiet" in class, and he added that he thought it was "exciting" to get a point. He helped to shift the theme within the discussion from "quiet" to being "excited". Student 1b did agree it was "exciting" because it reminded him of his childhood. Student 6b mainly said she was "excited" because she received a prize for her tokens. Student 4b expressed that she felt she needed prizes, and she said she liked that a lot. Student 5b was the only one who expressed that the general reaction to the tokens changed over time. She stated that everyone was excited at first, but over time that decreased. She also indicated that the teacher did not say when she gave out tokens in her class.

Student 1b – "Like when she says…when she says 'I'm giving out points today.' Like when we're--we're loud, she says 'I'm giving out points today.' Then everybody just…"

Student 2b – "Yeah. Then everyone's quiet."

Student 3b – "Everybody gets quiet."

Student 2b – "Like when everybody's talking in the room and she says that, everyone gets quiet."

Student 1b – "Like, it goes from 50 to zero."

Student 3b – "Yeah, it's…like, I guess it's really exciting whenever you get a point 'cause…"

Student 1b – "It makes us go back to our childhood." (everyone laughs.)

In eighth grade, most students recalled past experiences where they "liked" the tokens and rewards they received. Others showed they "didn't care" about them overall. All students except Student 4c had positive responses about what they experienced. Student 1c felt "good" and "great" about receiving stickers as rewards for answering questions correctly. In general, he thought that the other students "liked" it. There were a few unhappy students who did not receive stickers, but the majority of students had the same feelings as Student 1c. Student 2c "liked" his experience with AR points because he was able to collect them easily. He was a top reader in elementary school. He noticed that the smart students "liked" it, but those who were not successful were "frustrated" about the system. Student 3c expressed that she "liked" and "loved" the popsicle token system she experienced because she was able to have parties as a prize. She added that the experience made her feel "accomplished," and most of the students "liked" the experience. Similar to Students 1c and 2c, she saw that there were a few students who were mad because they did not receive much. Student 5c "liked" the treats she received when she exchanged coins for prizes. She believed that the other students had similar

feelings. Even though Student 6c had an experience that was different from Student 5c, she "liked" the prizes she received for collecting AR points. She also thought that the other students in her class felt the same way. Unlike the other students, Student 4c felt that his experience with receiving behavior cards was not a big deal. The people who did well in the class "liked" it, but those who misbehaved "didn't care."

> Student 2c – "I liked it because in elementary school I was reading like—like thick, thick books like…thirty or forty AR points, and I would have, like, the most in the grade. And I'd just go and cash and I would buy people stuff like…like so I—I liked it because I was like the best at it like…" (trails off.)
>
> Student 3c – "Uh, I liked it because I loved going to the little party things, and that made me feel like I accomplished something."
>
> Student 4c – "I mean, it wasn't a really big deal to me just 'cause it was a little card." (everyone laughs.)
>
> Student 5c – "I liked it. I liked it 'cause it was like, you get the little treats or whatever."
>
> Student 6c – "Basically what she said."

For the eighth graders, the overall reaction seen within the present experiences was that of being "indifferent" to the points, coins, and prizes. These responses were more neutral compared to the reactions given to the past experiences. The expression of indifference was not the same for all students. Student 1c felt that the tokens "don't influence" his behavior. In his opinion, the prizes were not appropriate for his grade level. Students 3c and 5c both stated that the tokens and prizes "didn't change" their behavior. Student 3c said that even though she liked the headphones time as a prize, it did not make her do anything different than what she usually did. Student 5c stated that maybe if the prizes were better, then the systems might have had more of an impact. Students 2c and 4c stated that they felt "indifferent" about the whole thing. Student 2c

further elaborated that he was "indifferent" to the systems because he could go to a store and get a better deal with the use of real money. Student 4c said he would be more "enthusiastic" if the prizes were better. At first they seemed to be good, but then when he saw what the prizes were, they did not meet his expectations. His statement was in agreement with what was said for both Students 2c and 4c. Student 6c shared that she agreed with what the other students said, but she did not provide any further elaboration about her feelings of indifference.

> Student 2c – "Like fifty-cents. You can go to the store. I want, instead of like…Now, the listening to music during class, that's kind of cool, but, then again…I mean, I can wait 'till the end of the day 'cause the class is only like an hour. I can wait. So, it's not worth, like…It makes me indifferent 'cause it's not like I'm getting beat if I—if I act bad that day. It's not like I'm getting a Lamborghini because I—I did perfect at the next day. So, it's like, indifferent to me. And candy's not that much, so…I mean."
>
> Student 4c – "You get like two pieces that are like (indicates small size with fingers.) this big."
>
> Student 3c – "It—It didn't really change anything bec—I just do what I usually do. I mean, I like the listening to music part, but I mean…I'm not gonna like change anything different to do. It's not a big thing."
>
> Student 4c – "I mean, I kinda…At first, it was kind of—It—It seemed pretty cool. But then, like, once we started realizing what, like, the prizes were, we're like (showing indifference.) 'Eh. It—It's kinda cool but not really.'"
>
> Student 5c – "Pretty much what he said. Like, I mean I—It doesn't really change because like…If the prizes were better…like…they're just like…little pieces or whatever. They're really nothing…well one or two, but."

Interpretation for Result 3.1. The information documented in Result 3.1 talked about the past and present experiences of students who participated in the focus groups. There were positive and neutral reactions expressed within both time frames. There were more students who had positive reactions with their past experiences when compared to

the present. As students get older, their wants and needs can change. What they liked in the past may not be the same as what they liked within their school environment during the study. There were differences present when the reactions were sorted by grade level. Overall, the seventh-grade students appeared to have had the most positive outlook about the points and coins systems during the study, but the eighth-grade students had the most indifferent outlook about them. Good feelings about tokens had variations according to perspective. Some students liked having the tokens because of the prizes they could get for them. Others felt that getting tokens said something about what they were able to accomplish as a student. They were further perceived as a nice distraction from the typical classroom routines. The students' feelings of indifference toward the token systems contained variations as well. Some students did not care about the systems because they felt their behavior did not change dramatically as a result of them. Other students wanted better options for the tokens and the prizes.

There are possible reasons that help to identify why there were different reactions within and between grade levels. One aspect of student learning that needs to be considered is that not every student perceived their environment in the same way. Similarities were found in past and present experiences, but this does not negate the fact that differences were found. Another consideration is that school culture can change from school to school. Each classroom can have its own culture as well. The cultural dynamics of each classroom can change over time.

This result helps to show how culture and identity can be factors in the experiences of students. Each individual student has their own identity, beliefs, and background. The students interviewed had differences in terms of behavior and

performance. This helped the researcher to gather a wider range of experiences and perceptions. It was found that a student who thrived in the school culture of their elementary school may have had a different outlook than a student who felt that the culture was unfair. Similarly, a student who performs and behaves well in middle school may be more accepting of school policies, procedures, and instructional strategies.

- Result 3.2: Students agreed that they mainly received tokens and prizes for good behavior, classroom participation, and following the rules.

According to the students within the focus groups, all of the teachers used the tokens to award good behavior during instruction. Students in the treatment groups were awarded tokens during individual tasks or group activities. The sixth and seventh graders also mentioned that the tokens were used for good grades. The sixth graders stated that the following behaviors would receive tokens: helping others, answering questions, turning in papers on time, and demonstrating good behavior for substitutes. The seventh graders received tokens for answering questions and other forms of class participation. They mentioned five additional behaviors worth noting: working hard, completing tasks with effort, sitting quietly, and not skipping class. For grades, there were a few times where students were recognized for receiving 100 on tests. The fact that students were required to complete tasks with effort in order to receive tokens was not mentioned in the sixth-grade focus group. The seventh-grade students did say that they liked to talk a lot in class, but it was important that they were quiet when they were supposed to be quiet.

Similar to the other 2 grades, students in eighth grade said that their teacher awarded tokens for behavior; however, they did not state any instance where they were awarded for grades. Specifically, they stated that they received tokens for class participation, doing what they were supposed to do, and following the rules. Class

participation meant that students were answering questions during discussions and doing their work. Following the rules and doing what they were supposed to do meant completing work on time, completing enough work within ten minutes, not acting out in class, sitting quietly, and working quietly.

Themes for Result 3.2. The emergent themes for this result included the fact that the participants held the perception that those in the treatment groups received tokens when they were doing what they were "supposed to do." This involved the need to be "good" and willing to "help." How students defined these themes changed slightly from focus group to focus group. In sixth grade, Students 1a, 4a, and 5a said that tokens were received for "good behavior" and "good grades." Student 1a added that "helping" someone pick up dropped papers was an example of something that could receive tokens. Student 4a did not use a specific example, but he did state that students could be awarded for different things. Student 5a elaborated that doing good things, making good grades, and mastering what was on Khan Academy would be eligible for tokens. Student 2a, like Students 1a and 5a, gave an example of when tokens could be awarded for behavior. He said that students could answer questions to receive them; however, he also contradicted himself by later stating that his class did not get tokens at that time. Student 6a discussed multiple instances where students could receive tokens, and this supported the examples already given. She said students could get tokens by "helping" others, turning in papers on time, and behaving well. Student 3a brought up that the advanced math classes in the study did not receive tokens.

Student 1a – "Umm…We get it for like good behavior and good grades."

Student 2a – "We sometimes get them by answering questions."

> Student 5a (smiles.) – "Oh yeah!...We just get them for like…I guess doing good things but also like making good grades and like…stuff on Khan Academy. Mastering those things."
>
> Student 6a (clears throat) – "Um…You could also get stuff…You can get the tokens by…like if you help a new kid that comes in, or if you turn in your papers on time. Kinda like what we did. We turned in our papers, for this, on time. And if you turned them in on time you got one token. Or if you behaved very well while the substitute was here, you got a token. And so far, I've only gotten two."

In seventh grade, Student 1b was the most vocal about how students earned points and coins. He said that the teacher did a "good job" from what he witnessed. Students were required to do what they were "supposed to do," which included students who completed work, stayed on task, were quiet, and asked reasonable questions. Students 1b and 2b both observed that students who were "getting involved" in class received tokens. Student 1b clarified this by saying that students who participated in class could get them too. Students 2b, 3b, and 4b discussed that students had to be "good." Student 2b stated that students could get tokens for "working good together" in groups. Student 3b said that students had to receive a "really good grade" on a test to get tokens. Student 4b went even further by saying that students needed to be "good all around." This included being really quiet during instructional activities that involved technology. Furthermore, Students 1b, 2b, 4b, and 5b agreed that students needed to put effort into their work to receive tokens. Student 6b did not have much to say after the other students stated their perspectives. She did say she agreed with what Student 5b said because they were in the same class.

> Student 1b – "(Name deleted.)…I think she does a good job of it. Sh—She walks around and look at kids who are supposed t—that are doing what they are supposed to do. So, it was easy for us to be on our best behavior in her class. But then, like once we get out of her class it's just…(trails off.)"

> Student 3b – "But, like…whenever we're ready…whenever she says to stop and get ready, we get ready and we work. And so she—she pays attention to that, and she usually likes it when kids, like, give her the answers whenever she asks for them. Like....sometimes I think she did this once or twice…If you got like 100 or a really good grade on a test, then she's give you a coin."
>
> Student 4b – "Just being good all around…Like, we…I was…We were all on technology today. Sooo, we were really, really quiet, and she (students briefly talk at once about being awarded.)…Yeah, a point…We got…We switched, and so we got points today, like everyone…except the people, like, who don't get points…like, who didn't sign up for it."

The eighth-grade group had similar themes as the sixth and seventh graders. Student 1c stated that those in eighth grade received tokens for "good behavior." He gave two scenarios to support the idea. First, he linked good behavior with answering a question correctly during a class discussion. Later in the interview, he emphasized that the teacher gave tokens when students "got a good bit done" during timed activities. On the other hand, Student 2c thought of answering questions as a form of "participation." He stated that students received tokens for "participation and behavior." He did not go into detail about what he thought behavior included as a separate category, but he did say that the teacher gave "participation and behavior" points when students worked hard and answered questions. He added that students did not get points if they were "moping" and decided to "act out" in class. Unlike the previous participants, Students 4c and 6c both said that students were given tokens when they "follow the rules." Student 6c added that students who do not talk at appropriate times get tokens as well. Student 5c did agree that the teacher gave students tokens for following rules. She also listed additional actions that qualified for tokens: "participating" in class and doing work quietly. Student 3c agreed that students were awarded tokens for being quiet, but she specifically stated that they had to sit quietly. Students might read while sitting quietly, and they still were

able to get tokens for that behavior. As far as when the teacher gave out tokens during instruction, Student 2c said that the teacher did not tell them when she would give out tokens during instruction. Students 3c and 4c provided more information about this. At first, Student 4c informed the researcher that the teacher would mark tallies at the end of the day to indicate when students earned points. Student 3c followed his comment by saying that she used tallies at the beginning of the day to mark students' progress, and then she marked more at the end of the day.

> Student 1c – "Basically, if you…had a good behavior in the class or…if like, so say that nobody could really figure out a question, and it's more of like a class discussion type of thing. If you got it right, and like, then you would get a point or whatever."
>
> Student 2c – "Uh, participation in class, like…If you…If you're just sitting there and you know, moping and having a bad day and you act out in class then you're not gonna get a point. But if you like…If she asks a question and you answer it, and (mumbles.) —if you do work and stuff, then she'll give you a point and, a point on behavior too, so like…It basic—It's basically participation and behavior."
>
> Student 4c – "B—Basically if you just follow the rules, you get a point."
>
> Student 3c – "It's just like, at the beginning of class, she'll look at us and she'll put little tallies down but that's about it. And then, like he (refers to Student 4c.) said, that we—at the end of the day she'll go back and give us some more."

Interpretation for Result 3.2. The information in this particular result pointed out that students received tokens for behaviors and performance habits that teachers found desirable. This depended largely on the ability for teachers to appropriately judge what behaviors needed improvement and what behaviors needed to be expected for all students. Students were awarded for a variety of behaviors. These behaviors helped to support the overall culture of each classroom environment. Most of the behaviors reinforced, in general, had to do with being actively involved in the classroom activities.

For the sixth graders, the focus was about doing good things independently and in groups. Good performance and behavior were expected of the students. The seventh graders were awarded for activities that stressed the need for good behavior, good performance, meeting requirements made by the teacher, and exceeding those requirements. Eighth grade students had to make sure to follow the routines and procedures that were explicitly stated and implied during instruction. From the students' responses, it was apparent that they did not have much personal input as far as the behaviors were concerned. The fairness of that depended on students' personal perceptions.

One explanation as to why students in all grades primarily received tokens for classroom participation, doing what they were supposed to do, and following the rules is because all teachers found those three aspects of behavior to be a priority. Not all teachers found the rewarding of grades to be a priority during instruction, even though the token treatments could be used for both performance and behavior. Also, the purpose of the token reinforcement was to focus on behaviors and tasks where students needed assistance. The PBIS framework required teachers to focus on behavior and how behavior can be managed, monitored, and improved according to students' needs.

In terms of the research question, the result helps to determine how the points and coins systems actually worked within PBIS. Students recalled how they received tokens, how they felt about what they received, and how they would improve these systems over time. For all grade levels, teachers had to determine what would be a priority when awarding tokens. They had to make judgements about who needed the tokens the most and why. Students who had good behavior would be more likely to receive tokens; however, the idea of fairness is subject to the ones giving and receiving

tokens. When priorities were established, they were based on the personal experiences and beliefs of the teachers. The amount of tokens students received was dependent on factors that were within and outside of the PBIS framework.

- Result 3.3: The higher the grade level, the more likely students preferred to keep up with points over coins.

During the focus groups, students shared their preferences in terms of tokens. They confirmed that the tokens used in the study were points and coins. The points were recorded by the teacher and the student. The teachers used their own versions of the Reinforcement Log form to record the amount of tokens students received. The students used notebook paper and sticky notes to keep up with the points. For the coins, the teachers had fake coins to give to the students. For sixth grade, most students in the interview said they liked tokens. Within the context of this particular focus group, students initially thought of tokens as coins. Later in the interview, most of them did not make that distinction. This may have to do with the fact that some students slightly changed their perspective as they heard others discuss the tokens. When compared to the sixth graders, the seventh graders had stronger preferences. Four out of the six students said they liked points, whereas two out of the six indicated that they liked coins better. Even though most students liked the points better, they noticed that the teacher gave out coins more frequently. Out of the 3 grade levels, the eighth graders had the most students who preferred points. There were 5 students who said that they liked the points. There was one of the five who did state he liked that they had more responsibility when using the coins, but ultimately the points were better for him. There was 1 student in the focus group who was indifferent to both tokens. For the eighth graders, the coins were easier to forget.

Themes for Result 3.3. The ability to view tokens as a large or small "matter" is another important theme that emerged from the data. Students had to "keep up" with the coins and points they had. There were students who had trouble remembering token amounts. Some students developed a "doesn't matter" attitude, whereas others showed more conflicting feelings about receiving them. Others did say that either coins or points were "easier" in terms of remembering. The sixth graders were interviewed at the beginning of the study before the switch point. They were just starting in terms of the collection of points and coins. Student 1a gave a response of "I don't know" in terms of how many tokens he had received at the time of the study. Even though he liked the coins he received, he said that there was "not much" to say from other students about it. There were two students, Students 2a and 3a, who said they had not received tokens. Student 2a said that he would like to receive twenty tokens despite the fact he said that the prizes needed to be "expensive" in order to "matter." Student 3a did not give a preference as to how many she would like to receive. When students talked about the prizes, she said that the homework pass was "not necessary" and that it "doesn't matter" to her. She did say in the interview that others may have the opposite opinion about tokens and prizes. Student 6a, who had two tokens, agreed about the homework passes by saying that she never used the ones she received. She had an "I don't know" response as to her reactions for the prizes overall. Student 1a disagreed with them when he indicated that they should have given him their passes because he had "other stuff to do." Student 4a had the most tokens out of the group, which was six. Even though he said getting the tokens "doesn't matter," he did express dislike for marking the points himself. He also stated that he wanted to try to get the prizes, but it depended on what it was.

Student 5a, who had around four tokens, agreed that she attempted to get the prizes as well.

> Student 4a – "When we get points, we have to multiply that score by 4. That's the part I don't like."
>
> Student 3a – "The homework pass…It's not necessary…You can always do your homework. It doesn't matter to me."
>
> Student 1a – "But I got other stuff to do."
>
> Student 4a – "I'd do my best trying to get that prize. Depending on what it is. If it's a sticker, then no."
>
> Student 5a – "Um, well I would try to get it, but it would probably be something really small, like maybe like a little Skittle."

Because the seventh-grade students had more experience with the tokens at the time of their focus group, more specific information could be gained about students' perceptions of the points and coins systems. Students 1b, 2b, 4b shared their experiences about how their teacher would "keep up" with the tokens. They also expressed which of the two token systems they preferred during the study. Student 1b said that the teacher "keeps up" with the coins, but some of the students chose to give them away to others. He said that she has a "roster" for the points, and the students made tally marks in their folders for the points. He preferred keeping up with the points because of this. Student 2b agreed that the teacher "keeps up" with the points and coins. Out of the 2 options, he said the points were "easier" for him. The coins were "hard to keep up with" in his opinion. Student 4b did not like that students wanted to "give them away." She said she used a "pouch" for the coins. Even though she did like both options, she thought the coins were "kind of better." Students 1b, 2b, and 4b all said that they had lost coins during the study. Student 5b said the teacher kept the coins in a box on her desk. She

added that the teacher "forgot" about the points moreso than the coins. Out of the two systems, she thought the coins were better because she was able to "hold onto" them without "forgetting" about them. Students 3b and 6b stated they preferred points, which was in agreement with Students 1b and 2b. Student 3b said the students received sticky notes for tallying the points, and he liked to place it in a notebook. He said sometimes it got "annoying" to manage when his bags were already packed at school because he would have to unpack everything in order to put down tally marks in his notebook. Overall, it was a better system for him than the coins. Student 6b simply stated that she liked points because she lost items easily. She did not go into detail as to whether or not she actually lost the coins during the study. As far as the amount of tokens are concerned, the highest number of tokens mentioned during discussions was 15.

> Student 1b – "She has a roster for…for points. That's why I like the tally marks that we do in the back of our folder. But you can't give away those."
>
> Student 2b – "I think—I think she keeps up with them…I think she keeps up with those and the coins. And the coins are hard to keep up with too. The points are easier."
>
> Student 4b – "I guess that I like coins kind of better, but I also like points. The coins, I dunno, I just have this little pouch and I just loved it so much…and I kept up with it after that. But the points, they're okay."
>
> Student 5b – "Overall, it made everybody behave a lot better. (all students agree) It like controlled them. But, like, coins I think they're better because you actually get to keep onto something and hold onto them, and know that you have it and like are not forgetting about something. And then, like points…they're good and everything, but in our class she forgot about points a lot. But now, like with coins, she gives them out more often because she has the big old box on her desk with them."

The eighth graders within the last focus group had experienced both token systems for the full-time span of the project. Even though Result 3.1 indicated that the general perception for eighth grade was indifference to the tokens, most students had a

preference in terms of which one worked best for them during the study. Between the two systems, the only student to say it "doesn't matter" was Student 4c. Even though he did agree with the others that the teacher marked which students received tokens, he also stated that the students would "keep up" with the points moreso than the coins. For the coins, he decided to throw them away when he first received them. He had 31 tokens altogether. He felt that the systems were "unfair" for people who put effort into their work but were not the top performers. The other students preferred the points. Like Student 4c, Student 1c felt that those who performed well were likely to "stand out" and get awarded tokens. He first stated that he liked the coins in the sense that the teacher relied on the students to "keep up" with them more; however, he concluded that the points were "less to worry about" and were the preferred choice. He said he had anywhere from 24 to 26 tokens. Student 2c liked the points overall because he felt that keeping up with it was "not a responsibility." In other words, the material aspect of the coins made it harder for him to keep track of the tokens. He liked the coins, but having fake coins instead of real coins did not make much sense to him. He did say at times he had trouble remembering the points. He had 13 tokens at the time. Student 3c said she preferred the points because she could not "keep up" with the coins. She frequently lost the coins, and she noted that the teacher would run out of them. It was hard for her to remember the coins. She said she had 24 tokens. Student 5c said she liked the points because she did not like "keeping up" with the coins. She had collected around 23 tokens during the study. Student 6c said she liked the points because she did not have to do much with them.

> Student 1c – "I like the coins because…I feel like she, relies on you more to keep up with the coins. But like, for instance, what if she really needs 'em?

> But I don't really have 'em like, 'cause I leave some of mine at home I know. So, like, I feel like the points…I really wouldn't have to keep up with them, so that's like less to worry about with me."
>
> Student 2c – "The points all the way because it's, you know, it's not a responsibility to keep up with points, like, it's not a material thing."
>
> Student 3c – "I'd say points because, uh, I can't keep up with coins. I—I lose 'em. She gave me three of 'em, and I lost them the next day."
>
> Student 4c – "I said, uh…I mean, it didn't matter. It doesn't matter."

Interpretation of Result 3.3. At the beginning of the study, responses about the tokens systems themselves were ambiguous at best. It was not until the middle and the end of the quasi-experiment that students were able to accurately form opinions about each type. From those opinions, they were able to determine which token medium was most appropriate for their current educational needs. When faced with a choice between one token type and another in the practical sense, some students had different answers than their general reactions within Result 3.1. For instance, a student could say that tokens do not matter to them in general; however, they could also have the perception that points are more appropriate for them when they do decide to receive tokens during instruction. This showed that students' opinions were flexible, and different feelings could arise within different situational contexts. It was possible for a student to prefer a token type due to ease of use, specific physical characteristics, and the storage options available for it.

One possible reason as to why more students preferred points, particularly within the higher grades, was that there may have been more of a negative perception of coins for the older students within school culture. Some students associated coins with being more childlike since they were fake gold coins. Another possible explanation had to do

with how others handled the points and coins systems. They watched their teachers and peers keep up with the systems, and this helped to formulate their opinions about their experiences. Finally, the symbolic nature of the points could be more reflective of the abstract concepts and processes that older students had to use to solve problems in the learning environment. Because of this, their learning goals and beliefs would be more aligned with the use of points during instruction.

This result describes how students thought about tokens in their experiences, which is important to know in order to sufficiently answer Research Question 3. Most students thought of the coins as a less practical option for their activities. The points were easier to manage along with all of their academic and personal responsibilities. There were students, however, who did like the coins because of their tangible nature. They did like the responsibility of keeping up with them, and they had to find inventive ways to store them. They also liked the fact they could make noises with them and use them during instruction. For the classes involved in the study, points may be a more viable option for Tier 1 supports. Coins could be more suitable for small group and individual reinforcement systems, which would be found within PBIS Tiers 2, 3, and 4.

Research Question 3b: What are the perceived experiences of teachers involved with the use of token-based systems in their classrooms?

Individual interviews were conducted for each teacher who implemented the token systems in their classrooms. As stated in Chapter 3, there were three math teachers who directly used the tokens with the student participants. Each one represented a different grade level. Like the student focus groups, the interviews took place at different times during the study. The first interview occurred at the beginning of the study with the sixth-grade teacher. The second interview, which was around the switch point,

involved participation from the seventh-grade teacher. The last interview was with the eighth-grade teacher, and it was scheduled around the end of the study.

Each interview was recorded with the same device used for the focus groups. Because the teacher interviews were coordinated in the same manner as the focus groups, the researcher was able to transcribe one interview and one focus group for every major data collection point within the study. The teachers did receive a copy of the interview guide strictly created for the individual interviews. For each teacher interview, the contents of the transcripts were read over by the researcher. They were then e-mailed to each teacher for a validity check, and all three approved of their transcripts.

The analysis stage for the interviews followed the same steps required with the focus groups. The major differences were in terms of individual coding and categorization. A separate coding and categorization system was created for the teacher interviews. This means that new text codes and color codes were required for the information. This system was first developed within the hardcopies, and the result was transferred to *R* statistical software. The synthesized finding for the teacher interviews is as follows:

> *Finding 3.2: Teachers' experiences were affected by (a) the reactions of students when the systems were in place (b) the amount of effort that was required for the activities within the token systems, and (c) their beliefs about reinforcement and punishment systems.*

The teachers had reactions that were based on their past and present experiences. All of the teachers observed that PBIS could be used in the middle school setting, but the amount of relevance the tokens had during instruction depended on personal perceptions of the systems. The amount of ease in managing the coins and points helped to determine the teachers' preferences and opinions on tokens. The students received coins and points

according to the behaviors that exemplified that they were doing what they were supposed to be doing. The teachers had different preferences about reinforcement and punishment. There were two teachers who preferred punishment, but they preferred the points when they actually decided to use reinforcement. There was one teacher who preferred both, yet she preferred coins when using the reinforcement systems.

- Result 3.4: Teachers had positive and negative reactions about their experiences with tokens and prizes.

Unlike the students, not all teachers had past experiences with tokens. According to the sixth-grade teacher, she had not used tokens as a student or a teacher, but she had heard of it as a teacher. She did mention a passport punishment system for wrong behavior that was used in the past during her teaching years. It was her opinion that there were improvements and challenges when using the token systems. The seventh-grade teacher used tokens as a student and as a teacher. She bought prizes with tickets as a middle schooler. Her field experience involved the use of cash and stamps. She had a positive reaction about her experiences with tokens in childhood, but she followed the rules regardless of what she received. She felt there were good and bad aspects about using it as a teacher. The eighth-grade teacher received behavior tokens in kindergarten. She later received bonus points in college for reinforcement. As a teacher, she used a sticker system and a ticket system with elementary students. In general, she felt that the tokens worked best for those in elementary school, but they would not be appropriate for middle schoolers.

Themes for Result 3.4. All of the teachers were familiar with the concept of tokens before the study began in their school. Overall, teachers mentioned the points and coins systems had "good" and "bad" features. Because the sixth-grade teacher did not

state any past experiences with tokens, her reactions about them stemmed from her present experiences. Like some of her students, she perceived that coins were tokens and that points were a separate form of reinforcement. In her classroom, she noticed that both were a "good thing" for a few of her students who had behavior problems. She said that they were "really working" for the points and coins. This was an "improvement" from her experiences with them in the past. Out of the students who received points and coins, the "high achievers" consistently had the most tokens. She did not find it surprising that those who had performed well wanted to get the tokens. They were "teacher pleasers" who wanted to do their best. She observed that most students "liked" the coins because they were "tangible." In term of the prizes, the majority of students were "really working" for the independent headphones time with music. They were also "excited" about the bathroom pass, and they "wanted" the candy. They were "not too excited" about the sticker. Those in the control group complained because they wanted to receive tokens and prizes too. One drawback of the points and coins for her was that she felt they were not going to affect the students who were "a little more mature." She felt that age, along with maturity, did play a role in how students reacted to the systems. Another issue had to do with the recordkeeping aspect of the project. She thought that keeping track of students' behaviors was "overwhelming." She stated that the other teachers on her team were "glad" that they did not have to directly deal with this aspect of the study. Based on her experiences during the study, she recommended tokens for students in Tier 2 and higher.

> Teacher 1, Comment 1 – "Um, not…Well there are kids who will consistently get more points than others. Those are your high achievers, you know, the teacher pleasers. I've seen a difference though with the kids who I've had behavior problems with in the past--wanting to be off task--and, you know, now they're

like ‘Uh, (Teacher’s name deleted.) I did something good. Where’s my point?’ or ‘Do I get a token?’ and all so. I’ve seen a change in a few of those kids.”

Teacher 1, Comment 2 – “Yeah, I mean, there’s probably been one or two that it’s not going to affect them either way. But overall, it’s been an improvement with those that I’ve had some issues with in the past.”

Teacher 1, Comment 3 – “They’re excited about the bathroom pass. And…um…I don’t give that much homework. I mean, I give homework, but it’s not an overwhelming amount. It’s…maybe eight, twelve problems over a two-day period that they can complete. They’re really working for the headphones. They want to use their technology.”

Teacher 1, Comment 4 – “Yeah, they’re…Actually, I mean…’Cause I know you’d given the form…that other form where I’m supposed to write down what each kid…(we smile at each other.) Yeah that one. That’s really overwhelming. That’s very overwhelming to sit there and write each time…If you’re giving a lot of points to a kid…I mean, that’s just a lot.”

Teacher 1, Comment 5 – “I dunno, I mean…’Cause they’re getting at the age now, I mean, that…It doesn’t work with all of them. Some of them are a little more mature. It’s just not going to happen for them. But, you know, we have kids that come in and do absolutely nothing. And a few of them have started working to get the tokens. For some of them, that’s just not going to help.”

In contrast to the sixth-grade teacher, the seventh-grade teacher did have experiences with tokens and prizes as a student. She said she “enjoyed” the ticket system she had as a student. It was “good” for her because she was a “rule follower.” She wanted to earn the tokens, but she did behave regardless of the amount she received. In her past experiences as a teacher, she noticed that there were “good and bad things” about token systems. She stated she was the co-chair of the PBIS program at the school. The PBIS program outside of the study did not have token systems, and past idea about tokens were “abandoned” because they were “tedious” as far as implementation was concerned. Within the quasi-experiment, she noticed that the treatment groups “liked” the token systems. They had “good” behavior before the study, and this behavior continued with the token systems. According to her, they especially “liked” the coins because of their

"tangible" nature. The control groups did want to be in the points and coins classes, and they did display feelings of being "jealous" over the fact that they were not going to receive the tokens and prizes. Those who had points the first half of the study were "excited" to receive coins for the second half. She felt that the students had a "childlike" mentality where they "loved" to receive the gold coins. Similar to the sixth-grade class, there were students who wanted to earn the headphones time with music. It made the students who received the prize look "cool." As she stated with her past experiences, the present experiences involved components that were "tedious." She felt that it was "difficult" to implement the systems effectively because she was busy with other educational tasks. She said that the boys had more trouble with storing the tokens. Students did change their minds about what prizes they wanted, and she felt that the "contrary" and "fickle" behavior was normal for seventh-grade students. The other seventh-grade teachers on her team were "glad" she was doing it instead of them, but they were "supportive" about any strategy that could address behavioral needs. She suggested that tokens should be used as a Tier 1 support for the majority of students.

> Teacher 2, Comment 1 – "Um…I enjoyed it…as a student. I mean, it was good. I always got my tokens 'cause I was a rule follower, so I always tried to…earn my tokens. But, I mean, I behaved whether I got the tokens or not."
>
> Teacher 2, Comment 2 – "Ummm…I…do not love it because I feel like, um…I think there are some good things about it and some bad things about it. I think that it's tedious to keep up with, and that's why we abandoned it with PBIS and kinda went towards some different directions."
>
> Teacher 2, Comment 3 – "I think they like it. Um, I think that the (brief laugh midspeech.) ones who got the nothing class, they don't like it (We both laugh.). They wish they were in the points and coins class. They like it…I think some of 'em do…I wish that ALL of my students would have brought back their, um, permission slips…like the whole class in each class because I think that would have made it better. I got a few who didn't, you know, and,

> um…it just kinda makes it more tedious for me to have to look back and see who gets points, who gets coins, who's not participating, you know."
>
> Teacher 2, Comment 4 – "I think the ones who…I will say that I think that the ones who got the coins this time were excited to trade in their points for coins. Um, there's something about the tangible coin, I think, compared to the points. I really do. I feel like they like getting that."
>
> Teacher 2, Comment 5 – "I think they're glad that I'm doing it (both of us laugh.). That's what I think. Um…I mean, they're supportive of it. They're definitely supportive of it. I have a good team, and we're all on the same page discipline-wise. And so, you know…They're supportive of anything that helps the overall team behavior."

The eighth-grade teacher had a few experiences with tokens and rewards as a student, but it was hard for her to remember her reactions because it was "so long ago." From her perspective, the majority of her childhood was "old school" in the sense that students did their work without getting many rewards for it. She had worked in elementary school before working in the middle school setting. She felt that the past token systems and reward systems she saw as a teacher worked better for elementary school students. At first, she was "not sure" if the present token systems made much of a difference for her eighth-grade students. Later on, she stated that the systems "didn't affect" the behavior that the students had in their classes. Those who were "well-behaved" continued to act that way, and those who misbehaved were not interested for the most part. She did say that students were "intrigued" and found it "very motivating" at the beginning. As time went by in the study, they preferred to go back to what they usually did. Unfortunately, there were some students who broke the coins or left them on the floor. She had to frequently remind students to trade in their points and coins for prizes. Most students thought the independent headphones time was "exciting," but there were some who became discouraged because of the high point value. There were

students who traded in for this prize, the candy, and the bathroom pass. Most of the trades were completed at the beginning and the end of the study. The eighth-grade teacher did express some conflict about using tokens and rewards as a form of motivation. She first felt that well-behaved students were "intrinsically motivated," so it "doesn't matter" if she gave them tokens and rewards for their behavior. On the other hand, she later expressed that those who were well-behaved should receive some form of "recognition" for that behavior. She did not like the recordkeeping aspect of the project. She felt it was "a lot" for the teachers to handle. This reaction was in agreement with the other teachers who participated in the study. She thought the behavior tracking she had to do was "extreme" because she already knew her students and the general reasons for why they received tokens and prizes. Her other team members were "glad" that they did not have to directly implement the systems. She agreed with the sixth-grade teacher in the fact that tokens were most appropriate for Tier 2 and higher.

> Teacher 3, Comment 1 – "Um…I think that it's very…um, for the younger students that are still real concrete, I think it um, worked well with them. I'm not so sure with these middle schoolers that it made a whole lot of difference."

> Teacher 3, Comment 2 – "And I mean, the thing—the truth of the matter is because it was a behavior…um…incentive, those kids who were generally well behaved…They didn't start misbehaving because of points or coins. They continued to. And a lot of those kids who are not well behaved…they…really weren't interested in poi—I mean, at this point in the game, by the ti—you know, and that's sad to say, but…Their behavior's kind of established by this point. And they've decided whether they want to…veer right or veer left."

> Teacher 3, Comment 3 – "There were a few of them that when they got close to 20 points, which was our top prize, and that meant they could listen to music while they did their independent work…um, a few of 'em. And some of 'em are—they still aren't there, but they're close. And um…you know, that was more exciting to them. But it seemed like, you know, 20 points was a

> long ways away when they were two or three weeks into the study. And so some of them were like, 'Huh…Whatever, I'll never get there. So, let me just have fun cuttin' up in class.' "
>
> Teacher 3, Comment 4 – "And, like I said, so many of 'em, they were just sitting there breaking 'em and saying, you know, 'Can I have half a token?' You know…It just became a…you know, fifteen minutes of passing out tokens and then the behavior and then I'm like 'I want to take that token back!' (we laugh.) 'Cause now you're acting crazy, and I just gave you tokens for good behavior."
>
> Teacher 3, Comment 5 – "Um…I think that the record keeping that you gave us as far as listing the students' names and how many points they earned and what their points were earned for…That was just…I think that was extreme."

Interpretation for Result 3.4. According to the teachers' perspectives, the token systems did not work for everyone. There were instances where students who behaved well and performed well wanted tokens and prizes. There were other situations in the classroom where students who behaved and performed poorly did better with the token systems. Then, there were examples where the systems did not matter to the students. Not all of the teachers' beliefs were in agreement with the students' beliefs. For instance, the teacher in sixth grade mentioned that the high achievers wanted to get the tokens more. There were students in the sixth-grade focus group who felt that students who often did not get much recognition wanted the tokens more. The seventh-grade teacher felt that students liked the coins more, but most of the students in the seventh-grade focus group said that the points were the preferred option between the 2 systems. The eighth-grade teacher felt that her decisions about tokens were fair to all students, but not everyone in the eighth-grade focus group felt that the rules of the systems were fair. All of the teachers agreed that in terms of their students, the younger and less mature ones were more likely to respond well to tokens. They also were of the opinion that the

systems they had outside of the token systems were the most appropriate for what they needed to do. Although there were students who benefited from the tokens, the teachers preferred to use other supports.

A possible explanation for the fact that teachers and students were in conflict may have to do with their past experiences. Teachers' past experiences helped to shape their beliefs about the present. The sixth-grade teacher had no past educational experiences with tokens when she was a middle school student, and she generally felt that tokens and rewards were not necessary to get students to do what they were supposed to do. The eighth-grade teacher had elementary school experiences with tokens, and she thought they worked best in the elementary setting. The seventh-grade teacher did have a middle school experience with tokens, and she felt that there was a need for reinforcement in some capacity. The reason for the dislike of tokens in terms of management may have to do with the fact that the system was not fully integrated with the policies and procedures of the school environment. The paperwork required for tokens seemed more of hassle than a form of support, and the teachers perceived it to be something that was separate from their instructional priorities. They saw the tokens as something that was not practical for their classroom cultures, regardless of students' opinions about them.

As far as the research question is concerned, this result assisted in establishing how teachers felt about the token systems of the past and present. It also discussed how teachers perceived the actions and reactions of their students. The experiences were not the same for each grade level. The seventh-grade teacher and her students seemed to have the most positive experience with tokens. Their reactions were more in favor of using tokens and rewards in the school setting. Past exposure to tokens and rewards in

the school setting may have helped the teacher and her students keep an open-mind about token systems in general. The eighth-grade teacher and her students seemed to exhibit more disinterest about the system. They perceived a great disparity between expectation and reality, and this perception may have played a part in their overall disinterest in the systems.

- Result 3.5: Teachers agreed that students received tokens for doing what they were supposed to do.

The points and coins were systems used for noticing students' when they did well. The sixth-grade teacher gave out tokens for students who actively participated in class, worked in groups, came back from the bathroom in a timely manner, received good grades, and displayed behavior outside of expectations. She did give them out every day class was in session, and all students had a chance to earn the tokens. They could earn at least 3 a day. The seventh-grade teacher gave tokens and prizes for a combination of actions during instruction. Like the sixth-grade teacher, the tokens were given out during individual and group activities. The seventh-grade teacher mentioned the following actions that were eligible to receive tokens and prizes: behaving above and beyond expectations, participating in class, scoring 100 on tests, sitting quietly when appropriate, helping other students, and following directions the first time. She used the tokens in class every day, and she recognized anywhere from 4 to 10 students per day. She gave all the students in the treatment groups a chance to receive the tokens, and she frequently gave them out to her students because they were well-behaved in general. She did like to give students extras when they showed that they could maintain behavior near dismissal time. The eighth-grade teacher kept track of who received tokens on a daily basis, but she would give out the tangible tokens and prizes at the end of each week. The good

behaviors that she noted were associated with participation, sitting quietly, and following class rules. The tokens were received during class discussions and independent work. The lowest amount of tokens that were received at the time of the interview was 11. The highest amount of tokens received by a student was 27. As previously stated in Result 3.4, she often had to remind students of their token amounts because of the general lack of interest over time.

Themes for Result 3.5. The main themes involved with the teachers' interviews were that students had to "participate," "answer," and "do what they are supposed to do" in order to receive tokens and prizes. The sixth-grade teacher stated that students needed to "actively participate" and "respond" in class in order to receive tokens during the study. As far as participation was concerned, she especially looked at how students were "working in groups." For example, she observed students during a "reciprocal learning activity" that was forty minutes before the interview was scheduled. She gave groups points for "staying on task." She also liked when students gave "creative" and "outside the box" answers during instruction. She specifically said she wanted them "doing what they were supposed to do" to get the tokens. She saw that tokens could be applied to purposes within and outside of the study. This included using them for "mastery of the standards," "training" situations during childhood, "behavior charts," and "math tournaments."

> Teacher 1, Comment 1 – "Students who respond in class, um, and actively participate, um, if they're working in groups, and doing what they are supposed to then I might give 'em points. Um, I even will reward them for going to the bathroom. We have to take them, as sixth grade, we have to take them to the bathroom, and you can't just go during class change or whatever. And if they go and come back in a quick and timely manner, then they might get a point. Um, good grades, um…(Thinking.) what were some of the other things…That's

> primarily the big things. If you give me an answer that's outside the box, you know, something creative, you know, you get a point."
>
> Teacher 1, Comment 2 – "Uh, the kids were doing a, um, partner activity…a reciprocal learning activity. And the groups that were on task, I'd given them the point. And that was just…40 minutes ago!"
>
> Teacher 1, Comment 3 – "I mean, besides behavior and mastery of standards, guess you could do…hmm…(thinking.)…Some kids…trying to think…We've got some that just…the behavior charts…you could do it in place of behavior charts. With those kids, the few kids that are Tier 2, Tier 3. You could do it then."

The seventh-grade teacher continued with the aforementioned themes when discussing how she awarded tokens in the classroom. She said that she noticed students who were "really doing what they were supposed to be doing." They usually went "above and beyond" what was expected of them. They were "participating in class" and "getting things done on the first directions." For instance, she shared that students were having trouble during class. They needed to settle down and get ready for dismissal. She gave extra coins to students who were "doing what they were supposed to be doing." This took place sometime on the day of her interview. She especially liked to award points during group work, when she would walk by and see if students were "doing what they were supposed to be doing." When asked about how token systems could support PBIS, she said that the school used to have a stamp token system before their current PBIS system. If students had good behavior for a week, then they received a stamp. The students later exchanged the stamps for rewards. The current system does not involve tokens. It is a reward system that gives items to students every few weeks for no referrals and good behavior.

> Teacher 2, Comment 1 – "And I try to notice the kids who are going kind of above and beyond what the normal expected behavior is, you know…They are participating in class and they are getting things done on the first directions.

> That kind of thing, you know, notice those types of behaviors where they're really doing what they are supposed to be doing. Kind of the model for everybody else. I usually like, quietly…I put it on their desk. Like today, we…um, I went and put a couple extras on somebody's desk at the end of class because we ha...we were kinda having trouble like, settling down and getting ready to go. So, I just went to the kids who were doing what they were supposed to be doing and put a coin on their desk. And, all the kids were like looking around because they wanted a coin too, you know."
>
> Teacher 2, Comment 2 – "Um…We do rewards for the students who have not have referrals and for the students who have not had behavior issues. Um, and we do that, you know, every few weeks. We reward those students and so…I like the way we do it now just because it's less on the teachers. And to get teachers to buy into the PBIS program since I'm the co-chair of that…we have to make it easy for them, or they don't want to do it."
>
> Teacher 2, Comment 3 – "I mean, you know, you could go back to…you could do a token system. I think what they did before here is they used…um…the stamplike thing, and if you had had good behavior for a week, you got a stamp. And if you got so many stamps, then you got a reward at the end of the nine weeks…or something is how they did it. You could do something like that. Um… (taps table.) You know, I think it supports it in just the fact that it is…supporting positive behavior."

During the interview, the eighth-grade teacher generally stated that she used tokens to award "behaviors." She said she "rarely" awarded students for "good grades." She looked through her totals for the students during the discussion, and she specified what students did to get high token amounts in her classroom. Among the behaviors she listed, she said that they tried to "answer every question," made sure to "participate" during class activities, and did what they were "supposed to do." She said that she told students what she awarded the tokens for during class. For instance, she stated to the participants that they could get points for "sitting quietly" during class and being "ready" for class to start. She felt that for PBIS, tokens had to be "strictly behavior" oriented. She was unsure as to how tokens could be used for grades. She said that the major

problem with using token systems for her school was that they became a "management issue" that was hard to afford over time.

> Teacher 3, Comment 1 – "Um, I basically…I mean, at the very beginning of the study, um, when I told them what was going on…they would, you know, they would walk in my room and even though they were their normal…I would just say 'Okay, points start now.' And they would (slaps hands together.) sit down, get busy, you know. And—And they wanted to be noticed 'cause a lot of times I would say 'If you were sitting quietly, ready for class within a minute of walking in the room, you would earn a point,' you know. So, it was kind of that thing."
>
> Teacher 3, Comment 2 – "Most of them have made, alright—probably the lowest I have from anybody at this point is like 11. And, some of them have earned up to 27. Um…but they're, you know, these kids who have earned like up to the 27s are the ones who…they pay attention in class, they try to answer every question I ask, they are on task, they're doing their work, if they don't understand it they're getting help, they're…if I put 'em in a group, they get along with everybody, they participate, they do what they're supposed to. I don't need to tell you a lot about how the one with only 11 acts." (we laugh.)
>
> Teacher 3, Comment 3 – "Well, PBIS is strictly behavior. So, as far as performance goes, I don't know that there could be a tie in there. There used to be…now this was before my time here…but they used to have a token system here, or a card punch system or something. And they just said the management of it got…just too out of hand. Having to punch cards or having to give out tokens or whatever it might be that…it just became too much of a management issue. And I think it also became, um, just the money involved."

Interpretation for Result 3.5. The teachers within the study intended for the students to meet and exceed their expectations. They used tokens to show students how they wanted them to act. Two of the teachers did include performance outcomes as a qualification for tokens. Students who had good behavior and made good grades had a higher chance of receiving tokens in the treatment groups. This is because their actions were in accordance with the rules and routines of the classroom environment. This is not to say that students with bad behavior received no tokens. According to the teachers, they received tokens when they actually did what they were supposed to do. Teachers

specifically liked when students got involved in what was happening in the classroom. They paid attention to how well students performed individually and in groups. What they said in terms of how they awarded tokens is very similar to what the students had to say about it.

The teachers' decisions about how they award tokens may have to do with the support systems in place at the time of the study. Teachers had to focus on PBIS and the Georgia Performance Standards. Those areas of concentration were normal aspects of the school environment that were often addressed during math instruction. Also, they may have had to take into consideration the behaviors that were necessary for effective classroom management. In these token systems, the target behaviors were what students needed in order to stay focused on instructional tasks. Students who were well-behaved helped to establish order and control in the classroom setting. The teachers may have targeted those students in particular as a way to show that the rules actually had relevance to what was happening in their present experiences.

For the research question, this result provides answers concerning (a) how teachers awarded tokens during instruction, (b) why they gave tokens to students, and (c) how tokens can be incorporated into educational experiences. The teachers did have choices in terms of their expectations. They made sure to specify their expectations for students. Students who answered questions, participated in instructional activities, and were least likely to disrupt instructional tasks set examples that the teachers wanted other students to recognize. By calling attention to good behavior, the teachers stressed to their students what they found to be important within the classroom environment. Not all students agreed with the teachers' opinions, but the ones who did demonstrated that

through their actions. The token systems did help with classroom management to a certain extent, but they posed a challenge for teachers because they had to be used along other management strategies.

- Result 3.6: Two out of the three teachers preferred punishment over reinforcement.

From the interview transcripts, all teachers thought punishment was realistic to use with students. One teacher thought reinforcement was alright as well. During the interview, the teachers did not distinguish between the concept of tokens and the concept of rewards. There were times when tokens were referred to as rewards. In terms of punishment, there was a referral system outside of the study that was used by all teachers. Students who misbehaved in class would receive a verbal reprimand. If the behavior continued, they had a behavior tracker in their agenda that would get signed. When they were disrespectful in class or problem behavior continued to the point where instruction was halted because of it, it was an automatic referral.

There were also reinforcement systems in place besides the tokens systems within the study. Students who got zero referrals during the nine weeks received prizes that included movie time or extra free time. If they did not get their trackers signed, they could receive extra technology time or extra gym time. They could receive praise in the classroom or small food items. Other nine-weeks rewards like gym time, game time, and refreshments were used. The teachers, when creating the prize list for the token systems, included prizes that were affordable but unique to the systems in the study (see Appendix I). The teachers felt that the systems they had before the study were appropriate for students' needs; however, the teachers still had differences of opinion as to what worked in the school system.

The sixth-grade teacher felt that punishment was easier to implement in the classroom setting. Some of the strategies she used in addition to the ones listed above were the following: giving students mean looks, having silent lunch, removing team time, and taking students out to the hall. In her opinion, it was harder to remember to notice students for good behavior and give them something for it. She felt that students who were well-behaved displayed a pattern of behavior that was already established without rewards. She recommended that a certain amount of tokens should be given to students at the start of the system, and teachers could take away tokens when students misbehave. This type of system would be a response cost system. Information about response cost was provided in Chapter 2.

The seventh-grade teacher, on the other hand, felt that both reinforcement and punishment were appropriate in the school system. The way she enforced the rules was based on what she learned in the past from other teachers. Along with the referrals systems already in place, she preferred to walk around the classroom in order to monitor behavior. The close distance between her and the students made it less likely that students would misbehave. She also liked to talk to students privately outside of class about their behavior. She thought that punishment and reinforcement should be mixed. She had very little behavior problems in her class overall and she said that overall her students were good this year. She felt that when students did exceptionally well during the study, and it was important to give them recognition for it.

The eighth-grade teacher found punishment to be easier to implement than reinforcement. She referred to herself as a laid back type of teacher who was very patient with students. When she did get mad, the students knew it. In addition to the referral

system in the school environment, she did use certain looks to get students' attention. She sent them to another class if they were very disruptive. Within her teaching philosophy, she felt that students should be recognized for doing well. She did like reinforcement, in theory. She wanted students to be noticed for their good behavior. She did feel, however, that using a reinforcement system along with a punishment system was too much to handle. In terms of what actually worked in the school system, punishment was the preferred choice. She also suggested that a response cost system would be more appropriate for the school system.

Themes for Result 3.6. Themes emerged from the interview that addressed how teachers perceived reinforcement and punishment. They thought of all forms of reinforcement as something "positive" in order to "reward" good behavior. They thought of all forms of punishment as something "negative" for "bad" behavior. The sixth-grade teacher felt that it was "easier" to punish the "negative" behavior. This was due to the fact that it was less on her in terms of management. It was hard for her to "reward the positive" because it was difficult to remember in class. She said that she was raised where students just did what they were "supposed to do" without the use of tokens and prizes. She did state that there was a student who was raised liked that in her classroom. For her, it would have been "easier" if the tokens could have been removed instead of presented to students.

> Teacher 1, Comment 1 – "It's—It's easier to reward the neg—to deal with the negative than it is to reward the positive."
>
> Teacher 1, Comment 2 – "It's just easier when you're trying to do 500 other things in class. You don't remember to reward that positive behavior. When I was brought up, that's what you're supposed to be doing. So, that's what you do. (Points to a student.) That's how that one was brought up too."

> Teacher 1, Comment 3 – "That would probably be easier to remember. Yeah, that would be easier. To let 'em start out with…10 tokens or whatever—or however many tokens—and then remove them. That would be easier to remember."

In terms of preferences, the seventh-grade teacher stated that she preferred "a mix" of reinforcement and punishment as a result of her experiences in the PBIS program. She said she enjoyed "rewarding" the good students. She added that students who do what they are "supposed to do" should get a lot in terms of rewards. She stated that not many students displayed bad behavior, and it was important to be "proactive" in order to prevent potential problems with students' behavior. Punishments such as agenda signing and referrals were only issued by her for "major" behavior problems. From her perspective, she felt that tokens did help to support "positive behavior". She said that in parent meetings, they encouraged parents to set up "rewards" systems, but something like a token system created a "struggle" for them.

> Teacher 2, Comment 1 – "I mean…I can't think of the last time that I had a referral that was straight from classroom behavior. Not this year. I mean, I've written a couple…but it was because of things that, were like, out in the hallways or…um…you know, getting into an argument or inciting a fight. That kind of thing. Not anything in the classroom. Now, we do build up, and they can…in that agenda tracker…they can get a referral at the bottom. Like if they get that thing filled up, they will get a referral."

> Teacher 2, Comment 2 – "They get…So they have their…I mean it's…If you do what you're supposed to do, I think you get a lot. I mean, and that's the way we want it to be. We want to reward the kids who are always doing what they are supposed to be doing, so."

> Teacher 2, Comment 3 – "A lot of times when we have a parent meeting and we have a kid who has been really having a lot of disciplinary problems…we'll talk to them about setting up some rewards for that child to…like something that, if you can do this, you can have this. You know, that kind of a…Like, if you cannot have a referral for two weeks, then we will (laughs.), you know, take you skating, or whatever. Like, not just to make it about taking things away, but to add something back when they do what they're supposed to do…so."

The eighth-grade teacher did admit that ideally she liked to "reward" good behavior. She liked that she was more "aware" of their behavior. From a practical standpoint, however, it was not "logical" for her to use positive reinforcement for "good choices." "Bad" behavior, if ignored, would get "out of hand" without a punishment component. For her, there needed to be clear consequences for "bad choices." She felt that it did not matter if she "reinforced" students who "do what they are supposed to do." They were "intrinsically motivated" to be well-behaved. This meant that their "reward" was learning and staying out of trouble. Consequences, in her opinion, were only necessary for those who made "bad choices."

> Teacher 3, Comment 1 – "When I start trying to…and it's sad, that they kind of get pushed to the side—and they don't really get pushed to the side, but when I start trying to reward good behavior…as well as still continue to…penalize those with bad behavior…then it just becomes more…cumbersome for me. More, 'Oh, my gosh! Oh, my gosh! You're doing so good. Here, let me give you points, points, points…Here's a token. Eh…Let me sign that tracker.' You know. 'You don't earn a point, you don't get, you know…but I still have to sign that tracker.' So, I think it was just the—just the combination of both of them. Um, I prefer to reward good behavior. That's what I would like to do. Um, but if I—if I try to—to ignore the bad behavior and just deal—you know, just do positive reinforcement, then the bad behavior gets so out of hand that it's really not…logical to do it that way."

> Teacher 3, Comment 2 – "Um…(long pause.)…That's a hard one because…I do think kids who…do what they're supposed to…deserve some recognition for that. But as far as…I think there have to be consequences for bad choices. So, I think that our PBIS system that we have set up…um…You get some chances, but eventually if you get to this point, you've made enough bad choices that you have to suffer a consequence for it."

> Teacher 3, Comment 3 – "Um…for the kids who are behaving…sometimes I'm torn about that, because" I say "Why are we rewarding kids for doing what they are supposed to do?" Their reward is they're not getting in trouble. Their reward is they're learning because they're paying attention in class and doing what they're supposed to do. That's their reward. We have consequences for those who don't make those good choices. So…It's—It

> goes along with my philosophy in that it makes me more aware of who those students are that are really trying. And then when I'm…grading something or whatever…I can say, with all…you know…not feel bad at all about the fact that, okay, he really didn't get this answer—like an explanation or something—like, he really didn't explain it clearly."

Interpretation for Result 3.6. The teachers in sixth grade and eighth grade preferred to use punishment in the classroom. They had a similar teaching philosophy where they felt students who did what they were supposed to do were already motivated to maintain that behavior over time. According to them, the motivation either came from inside the person or came from how they were raised. Reinforcement was something optional for good behavior, but punishment was a requirement. The seventh-grade teacher preferred to use both. In her philosophy, teachers should desire that the good students get the most rewards. It was practical that teachers noticed those who were doing what they were supposed to do. They needed to use them as models for good behavior. She felt there were more good students than bad, and the good students needed recognition. The bad students needed punishment. None of the teachers were of the opinion that positive reinforcement was enough by itself to sustain the demands of the classroom environment. In addition, there were perceptual conflicts that existed for all teachers. Sometimes teachers thought of tokens as rewards while referring to them as reinforcement. They thought of reinforcement as a positive action, and they thought of punishment as a negative action.

The strong preference for punishment could have originated as a result of teachers' personal and professional experiences over time. They saw what worked for themselves and for their students. For the two teachers who preferred punishment systems, they stated that their past experiences with reinforcement as students were

limited. The sixth-grade teacher had not implemented a standardized token system until the time of the study. The seventh-grade teacher who preferred a mix of both systems based her decisions on her student experience with tokens, her teaching experience with different systems, and her co-chairing experience on her school's PBIS team. Conflicts in the teachers' perceptions of tokens and prizes may have derived from the educational language of the school system in which they were employed. The communication that teachers used to implement the systems was based on their understanding of the PBIS system and their knowledge of tokens as a whole. The educational knowledge needed for the classroom may not have been the same as what was required from textbooks or research studies.

The information provided from Result 3.6 provided answers for what type of reinforcement and punishment strategies worked during the teachers' experiences. They did prefer what they already had in terms of punishment strategies. The reinforcement strategies that were in place, which included the token systems during the study, needed to be improved or eliminated in order for the teachers to be content with what was occurring in their classrooms. Tokens are a form of positive reinforcement, but this is not their only use. The teachers may benefit from a reassessment of the prizes and token values associated with the study. This would allow for meaningful changes in the reinforcement system. If the teachers decide to make improvements based on their preferences and those of their students, they would be more likely to integrate tokens within their PBIS system in the future.

- Result 3.7: Two out of the three teachers preferred points over coins.

In terms of the coins and points systems, the teachers had different preferences. They did state their opinion about what the students liked. The sixth-grade teacher found that out of the 2 systems, the points were the most appropriate choice for her. She said that the students preferred the coins because they were tangible objects. The teacher felt they were a distraction. She noticed that the students banged them during instruction. The focus group did have an overall preference for coins.

The seventh-grade teacher did see good and bad to both systems, but she had an overall preference for the coins. Like the sixth-grade teacher, she noticed that students liked the tangible nature of the coins. Because the seventh-grade teacher was interviewed around the time of the switch point, she shared that the students who had the points were excited to trade them in for the coins. This contrasts with the overall focus group preference of points. The seventh-grade teacher mentioned that the eighth-grade teacher liked the points.

The eighth-grade teacher did state that she liked the points during her interview. Students were more likely to mishandle the coins than the points. For instance, the students would break the coins during class. The teacher did not like this, and she wished she could have taken the tokens away. Token reinforcement did not involve the removal of tokens because a separate system would have been required for it. The eighth-grade focus group preferred the use of points too. According to her, the students' behavior did not change from what it normally was. The teacher described in detail the students' reactions to the systems overall. She said that her students were excited about receiving tokens at first, but then this changed after a few weeks.

Themes for Result 3.7. The main themes in terms of reinforcement preferences were that their choices were based on how "easy" it was to "keep up," "mark," and "tally" the tokens. This is why they would "like" one over the other. The sixth-grade teacher said that she did "like" points. She did not feel as though it was difficult in terms of "marking". In her interview, she said she did not "really care for" tokens; however, she specifically said the reason for that was because the students tapped the coins. She found that the coins were a "distraction" for students during instruction. Ironically, she observed that she gave out more coins than points. If she had the option, she would just "keep up" with everything herself instead of having the students "keep up" with it too.

> Teacher 1, Comment 1 – "Um, well I think the points…Well, I like the points. I think the kids like the coins."
>
> Teacher 1, Comment 2 – " 'Cause they like the tangible. You know, something in their hands. Um, cause I seem to give more positive…More tick marks for the class that I've done the coins with…versus the one where the kids are supposed to be keeping up with the tally marks on a sheet of paper."
>
> Teacher 1, Comment 3 – "For me personally, I would do away with the kids keeping up with it completely, and I would just keep up with it. And I mean…like I've got my book, and…I mean…they can look at my sheet anytime to see where they are you know. But I would keep up with it versus asking for them to keep up with it or having that tangible. Although the kids like the tangible and they work for those little gold coins."
>
> Teacher 1, Comment 4 – "Um…Like I said…I don't really care for the tokens. Um, because kids start tapping them. They're becoming a distraction to some of them…with the ADD…that silver, shiny thing…Let's go pay attention to it and not the instruction. So, overall, it's not my favorite."
>
> Teacher 1, Comment 5 – "I prefer the points, but there again I don't feel like I'm marking it."

The seventh-grade teacher saw "good and bad" with both systems, but she had more positive comments about the coins. When asked about the tokens, she thought it

was good because of the "tangible" aspect of the systems. This was especially noted about the coins. She stated that her students "like" the coins because of this "tangible" quality. Even though it was "harder to keep up with," it was not as easy to "forget" when compared to the points. She gave out the coins more in class. If she were to implement the system again, she would just want a "tangible" form of reinforcement that was "easy" for her to grab and use on a daily basis.

> Teacher 2, Comment 1 – "Um…I think it's good. I think, um, definitely something tangible. Like, I think that it's better. Now, I know (name deleted.), that's the eighth-grade teacher, she said that hers…She likes the points better. Mine tend to like the coins. They still are a little bit childlike, and they like to get that coin in their hand. They love to get that gold coin. I don't know why. I mean, they just like it."
>
> Teacher 2, Comment 2 – "The coins…I would say…Like I say, they're a little harder to keep up with for the kids, but the good thing is that that box is on my desk and a lot of times I walk by and notice it and grab it and pick it up and start walkin' around, and…you know…(makes giving out gesture with hands, makes slight noise with table.) …giving out a couple coins. So, to me, that's just the kinda person I am. When I see something like that, it reminds me of it. So, it was just a good…I mean, that's the good thing for me with that."
>
> Teacher 2, Comment 3 – "Um…but that was easy, just to have a sticky note in the back with their name on it, and they just keep the points there. That'll be the good thing about the points to me. Um, the bad thing is that it's kinda outta sight, outta mind."
>
> Teacher 2, Comment 4 – "And so…it was easy for me to forget about it. It has been easy for me to forget about the points more."
>
> Teacher 2, Comment 5 – "I mean, you know, all these things that are going on, and um…that just adds another thing to the plate that I would say makes it difficult sometimes…Like, just too busy to do it effectively I feel like sometimes. Um…If I were gonna do it, I think I would definitely have everybody doing like something tangible that I could just go grab out and just hand to them. And I do that sometimes with my class with candy."

Similar to the sixth-grade teacher, the eighth-grade teacher commented that she did "prefer" the points over the coins. They were "easier to keep up with" for her when compared to the coins. She used a clipboard and her roster to "keep up" with both systems each day. Even though she did "prefer" the points, she noticed that she gave out more coins. This trend was seen with the other teachers. She felt that a points system was more practical for her students. The eighth-grade teacher did have perceptual conflicts in terms of how she viewed tokens within the study. There were times when the concept of tokens was strictly associated with the coins system.

> Teacher 3, Comment 1 – "And, with the points…I found that giving points, for me…It was a lot easier to keep up with points than tokens."
>
> Teacher 3, Comment 2 – "Um…Simply because I had my clipboard, and all I would have to do is say 'You sure are doing a nice job working together. Y'all each earn a point.' You know, and I would give them a point at that time. And then at the end of the…period, I would divide by…It took four behaviors. So, it was easier to do the points for me than the tokens because I had a hard time saying 'Okay, Good 1, Good 2, Good 3, Good 4. Now, you get a token.' That was—That was harder for me."
>
> Teacher 3, Comment 3 – "I prefer the points, and that's…I told you a while ago. It was just so much easier for me to do that, and I was good about doing the points on a daily basis. 'You earned four points today, you earned three points today.' You know, that kind of thing. Um…I prefer the points. I just…counting out those coins…you know, and not having enough tokens. And then trying to come up with 'Okay, you've got 22. Why don't you trade in 20 of them 'cause you can't get a prize for more than 20. So, go ahead and give me back 20 and I'll keep on record that you still have two left over to start adding again.' Um, it was just easier for me with the points. It just seemed faster and more efficient."
>
> Teacher 3, Comment 4 – "Like five who have been kinda targeted and said if they mess up again this year, they're out of here. You know, so that last class of the day is tough. And um, so even though the points are easier for me to keep up with and easier for me to make them aware of on a daily basis, um, I just didn't give as many to that group because their behavior is not as good."
>
> Teacher 3, Comment 5 – "With the small group…Well with 8th graders I would probably do points."

Interpretation for Result 3.7. The information from Result 3.7 indicated that the teachers who preferred punishment over reinforcement did make a choice when using reinforcement. In other words, they had preferences in the event that they wanted to utilize reinforcement. They both liked points over coins because the points system was easier for them to manage in the classroom. The teacher in seventh grade who liked a combination of reinforcement and punishment had a preference, but she was not as clear in her preference as the sixth and eighth grade teachers. She mentioned good and bad aspects of both systems, but her comments overall were more supportive of the coins system. It is important to note that all of them said they gave out more tallies for coins when they were marking the systems. Given the time of each interview, it was known that the teachers in sixth and seventh said this before and around the switch point, respectively. The eighth-grade teacher stated this after the switch point. The participants in the focus groups did not always have the same preferences as their teachers. In fact, the only students who agreed with their teacher as a whole were in eighth grade.

Each preference may have had to do with (a) how the teachers saw the students react to the systems and (b) how easy they thought the systems were compared to other teaching strategies available during the study. They possibly saw what worked for them and their students, and they formulated opinions based on what they witnessed. Conflicts in perceptions as far as tokens were concerned could have had to do with the language used within each classroom environment. Classrooms have different students and different requirements. It is possible that teachers can have perceptions about the classroom that conflict with their students. Regardless of any perception conflicts, all teachers said that tokens can be used in the PBIS system as a teaching strategy.

The data presented in Result 3.7 helped to answer Research Question 3b by stating teachers' preferences within the reinforcement systems they experienced. Not all of the teachers had the same preferences, but their points of view emphasized that token systems could be utilized for all grade levels. Two of the teachers did not see the need for mandatory reinforcement, but they liked having reinforcement options available to them. Because classroom environments are in constant flux, the instructional needs of students can change from day to day. This means that preferences were based on personal perceptions at certain periods of time. It is uncertain as to whether these preferences may change in the future.

Summary

This chapter presented the results of this study, which involved a convergent parallel mixed methods design. The data collected for this study were instrumental in answering three research questions. These questions focused on different aspects of token reinforcement, classroom behavior, achievement goals, motivational preferences, and classroom intervention experiences. From the analyses and interpretations required for the chapter, the following can be said about each research question:

Research Question 1

To what degree, if any, are differences found in Math Nine Weeks Test scores among students who participate in a token program and students who do not? Quantitative analyses showed there were significant differences found between the control groups and treatment groups. In terms of timing, the control groups outperformed the points and coins groups. It was observed that students who received coins in the second half scored significantly higher than students who received coins in the first half.

The opposite occurred with points, where students scored significantly higher in the first half. For token type, students in the control group did outperform the token groups on both tests. The points group outperformed the coins group on the pretest.

Differences in test scores were most likely influenced by the placement and grade level of the groups. When grouped according to whether or not students received accelerated math, it was found that three out of the four classes in the control group were accelerated groups. The sixth and eighth grade students were more likely to do well within the control groups. Seventh-grade students scored the highest within the control group during the pretest, but this same group scored the lowest for the posttest. Students who had free lunch scored lower than the paid-in-full and reduced-priced groups in terms of timing. Boys were more likely to pass the pretests when the groups were sorted by token type and gender.

The qualitative data supported these results in three ways. Firstly, the teachers expected more from the control groups because they tended to have less behavior and performance issues overall. They felt that those students demonstrated more intrinsic motivation and higher performance even though they were unable to receive tokens during the study. They further stated that students' needs could be met with the reinforcement and punishment systems already in place. Secondly, teachers stated that they gave out the coins and points in different frequencies. All students in the treatment groups had a chance to receive tokens, but teachers did not give out tokens in the same way for everyone. Thirdly, teachers in the sixth and eighth grades had similar perceptions about reinforcement and punishment systems. They felt that reinforcement was optional for students who had a mature attitude in regards to their school

responsibilities. They agreed that punishment systems were necessary for classroom management and academic success.

Research Question 2

To what extent are the test scores influenced by classroom behavior referrals, classroom achievement goals, and students' motivational preferences? For classroom behavior, there were significant differences found. Those in the No Referrals group scored significantly higher than those in the Low Frequency group, particularly in the month of February. The referral amounts for this group were significantly lower than the Low Frequency and High Frequency groups. Data for achievement goals showed that there were significantly high scores for students who (a) met goals for math achievement, (b) accurately assessed how well they did in math, (c) accurately stated at the beginning of the study that they would do great in the quarter, (d) had high expectations for test grades, and (e) had yearlong goals rather than monthly goals for the posttest. No significant differences were found solely for motivational preferences, but they were found when treatment combinations were added as a factor during analysis. Students who preferred coins as motivation had the lowest scores when they received what they wanted. They had the highest scores in the control group.

The practices and standards found within the school environment most likely played a role within the results. Classroom behavior could have been affected by the school-day schedule, instructional schedule, and the testing schedule of the school in which the study took place. The inclusion of the GPS standards in math helped to create a classroom environment that encouraged personal goals and achievements of this nature. The preferences of the students were determined by the wide variety of motivational

choices that were available to them at the time of the study. It was possible that students' perceptions of these choices had an effect as well.

The qualitative data supported the results about classroom behavior in the fact that there were apparent differences seen in classroom behavior and referral amounts between the focus groups. There were more referrals reported in the eighth-grade focus group at the end of the study. In addition, the data showed that most students could set accurate achievement goals and keep track of them over time. Most students had goals pertaining to grades, and they stated how well they usually performed on tests. Qualitative data on preferences included the implication that personal preferences did not guarantee student success, but the teachers did note that the control groups were more likely to be on their best behavior than the treatment groups.

Research Question 3

Important trends were found within the focus group data. The students had positive and neutral reactions about token experiences. The seventh graders seemed to benefit most from tokens and rewards in the present, but the sixth and eighth graders seemed to benefit from them more within their past experiences. The students received tokens according to how they well they behaved and how active they were within the classroom. Based on the students' responses, it was determined that the teachers' decisions were based on personal preferences as far as participation, performance, and behavior were concerned. The students tended to like points as the grade level increased. There were stronger preferences in terms of keeping up with them within the seventh and eighth grades. They preferred to keep up with the points. The sixth graders preferred the coins, but they had mixed feelings about their perceptions of tokens.

Additional trends were found within the individual teacher interviews. The teachers had positive and negative reactions about the token systems. The sixth and seventh grade teachers noticed positive reactions for several students, but there were ones who did not care for the systems. The eighth-grade teacher noticed more positive feelings at the beginning of the study, but most students went back to their old behaviors after the third week. They found that the amount of management involved in terms of monitoring overshadowed the benefits of the systems. The teachers did agree with the student participants in stating that they awarded tokens for appropriate behaviors and participation. The teachers each had their own reasons for awarding tokens, and they were based on students' needs. The majority of teachers preferred punishment over reinforcement, saying that it was easier to implement. They also preferred to use points over coins.

In terms of influences, one was that the culture of the classroom changed from day to day. Teachers' and students' perceptions did as well, and they formulated these perceptions based on past and present experiences. Also, the PBIS system was one that helped to encourage appropriate behaviors over time. Those who were well-behaved and did well on classroom activities tended to receive more tokens within the various treatment groups. Finally, the lack of full integration of the token systems within the school environment could have had a substantial effect on the dynamics of the school environment during everyday events. Not all students felt that the systems were compatible for their needs. Teachers did feel that the token systems for this study were overwhelming for them because they had to use other systems along with the ones required for it.

Chapter V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

The need for a wide range of instructional strategies within the current educational landscape is an ever present one. This includes the use of incentives. The advent of tokens within the 1960s educational landscape allowed for more opportunities to tie in the idea of reinforcement with different instructional techniques (Gaughan, 1985; Hackenberg, 2009). Since that time, there have been a variety of research studies about token economies; however, conflicts in results and gaps within the research literature persist (Maag, 2001; Maggin et al., 2011; Marinak & Gambrell, 2008; McDonald et al., 2014; Wolfe et al., 2003). Some of the more prominent issues found were a lack of standardization within token systems, lack of consistency with scheduling, and the need to incorporate achievement goals along with the reinforcement measures.

In order to address issues within past literature, a convergent parallel study was created that included procedures for standardization, reinforcement scheduling, and goal assessment. The overall purpose of this study, as explained in Chapter 1, was to ascertain the extent to which tokens influenced test scores for Nine Weeks Tests in math. It was developed to collect data pertaining to classroom behavior, goal orientation, and motivational preferences. The study involved 3 research questions:

1. To what degree, if any, are differences found in Math Nine Weeks Test scores among students who participate in a token program and students who do not?

a) To what degree are test scores influenced by the timing of token use in the classroom, i.e., students who receive tokens in the first half of the quarter versus students who receive tokens in the second half versus students who do not receive them at all?

b) To what degree are test scores influenced by the type of token used in the classroom, i.e., points-based system versus coin-based system versus no tokens used at all?

2. To what extent are the test scores influenced by (a) classroom behavior referrals, (b) classroom achievement goals, and (c) students' motivational preferences?

3. What are the perceived experiences of (a) students and (b) teachers involved with the use of token-based systems in their classrooms?

After the preliminary screenings were completed, there were 205 students who participated in the token intervention. All activities completed were voluntary in nature and were in agreement with IRB protocols. The convergent parallel design contained a quasi-experiment for 3 groups of math students: 2 treatment groups and 1 control group. Teachers received training for the token interventions. All grade levels (i.e., sixth, seventh, and eighth) had participants for each group. Group determination involved random assignment by drawing class names out of a hat. The duration of the study was approximately 9 weeks, with the treatment groups switching token treatments around the fourth week. During the study, the two treatment groups received points and coins according to behavior and test performance. Tokens were awarded according to the 4:1 ratio recommended for PBIS by Knoster (2014). Teachers developed a prize list, monitored referral amounts, and kept track of the tokens they awarded to students. Students also filled out goal orientation questionnaires and motivational preference check sheets. Teachers turned in their data according to the information presented in the Data Collection Timetable (see Appendix A).

For the qualitative strand of the study, three student focus groups and three teacher interviews were conducted in order to gather data about perceived experiences with the token interventions. The student focus groups had six students within each group. The students were asked questions according to a three-part interview guide (see Appendix H). Their responses were recorded and later transcribed. Parents received copies of the students' focus group transcripts. The teachers' interviews had a separate guide that asked them about their experiences with tokens. Three math teachers were interviewed for this study. Each teacher had a separate interview that was audio recorded. These recordings were later transcribed. Transcripts were later disseminated to the teachers for validation purposes. All data were analyzed according to the procedures presented in Chapter 3. Data mixing occurred during the interpretation stage of the study.

Conceptual Framework

The findings of each research question can be applied to the concepts found within Skinner's (1938) theory of operant conditioning along with Diener and Dweck's (1978) framework for goal theory. Research Question 1 addressed whether or not the token systems fulfilled their role as positive reinforcers as defined in operant conditioning. This question also can be used to address intrinsic motivation, extrinsic motivation, and goal achievement within achievement goal theory. As outlined in Chapter 1, a reinforcer was a stimulus that increased a target behavior after it was presented. The overall intention of the token systems was to increase the good performance and good behavior of the treatment groups over time. Research Question 2 was created in order to determine how classroom behavior referrals, achievement goals,

and motivational preferences played a role within the study. Research Question 3 discussed the perceived experiences of students and teachers.

Summary of Findings for Research Question 1

In terms of overall timing, increases in the average performance score on the Math Nine Weeks Tests were seen with the coins group and the control group. The coins group scored better when coins were given in the second half of the study. For the coins, the tokens did serve as positive reinforcement and extrinsic motivation. The control group performed better overall when knowing about the tokens and not receiving them. It can be argued that they were vicariously reinforced because they knew reinforcers were being used with other students. This knowledge could not be avoided within the practical school setting. The tokens were not directly presented to the control group, but they could have had an influence on any intrinsic motivators that helped the participants to succeed. The points group had a decrease in the overall group mean during the second half of the quarter. This means that the points actually had a positive punishment effect. Group performance decreased after tokens were presented over time.

Token type results in Research Question 1 indicated that the control group scored the highest overall. For the pretest, students who received points scored higher than the coins group. The coins group scored higher than the points group on the posttest. The fact the control group had accelerated math classes meant that the highest test scores were already found within that group. The control group participants did know the tokens were being given to the treatment groups. This may have helped them maintain performance in the sixth and eighth grades. It was found that a token combination could serve as a positive reinforcer if coins were given to students in the first half and points

were given to students in the second half. Students who received points before the switch and coins after the switch had a decrease in performance, resulting in a positive punishment effect. In terms of achievement goals, those in the control group passed the pretest and posttest. This was the only group that scored at least 70 on both assessments. It could be said that more intrinsic motivation was required because they did not receive tokens; however, it must be noted that they were receiving praise and rewards according to the PBIS intervention systems that were already in place.

Summary of Findings for Research Question 2

Within Research Question 2, the results about classroom behavior were able to determine the possible reinforcement capabilities of the token systems. Students who had good behavior (i.e., no referrals) performed better over time. The highest mean for these students was in February. In terms of operant conditioning, this demonstrated that the token interventions may have helped well-behaved students perform better through positive reinforcement. There were students who received no referrals who were from the control group. The knowledge of the tokens as well as the use of other interventions could have contributed to better performance. This cannot be said for those who received referrals. The highest mean for the Low Frequency students was in January, and the High Frequency group also performed the best in January. This happened to be the month where the first half of the study occurred. The amount of students who misbehaved, however, increased over time. In other words, the tokens seemed to have a positive effect on performance during the first half of the study, but any positive effects were not maintained by the end of the study.

In terms of motivation, those with no referrals seemed to be more extrinsically and intrinsically motivated than the other students overall. In terms of extrinsic motivation, the control groups were receiving acknowledgements for their achievements and good behavior through the use of positive feedback strategies that were separate from token systems. The students who behaved in the treatment groups were more likely to be reinforced with tokens. All students had to demonstrate intrinsic motivation by setting and assessing their own personal goals. Intrinsic motivation was present moreso for the control group because they did not receive acknowledgements as frequently as the treatment groups received their tokens. The teachers placed higher expectations on the control group in terms of self-regulation and discipline. The majority of students received no referrals, so both the treatment and control groups were able to show performance gains when sorted according to behavior. The Low Frequency and High Frequency groups were more indifferent about the treatment, especially after the switch point. As the referral frequency increased, the test score means decreased.

The results on achievement goals in Research Question 2 showed that students who did have a positive orientation within their personal goals were able to accurately judge how well they did on tests and how well they would do for the quarter. Those who thought they performed badly on tests ended up doing badly over the 9-week period. There were instances where students with high expectations were able to outperform other students, but this was not always the case. In addition, those with yearlong goals outperformed students with monthly goals during the posttest. These results showed that certain students who tended to set mastery goals, which were longer than 9 weeks,

scored higher on the posttest than students who had performance goals that lasted a month long.

Within the results for motivational preferences, the students scored similarly when sorted by personal preference only. If students had a personal preference in terms of token reinforcement, that preference did not guarantee high math performance. The token treatments, when factored in along with the motivational preferences, indicated a trend within students who preferred to receive coins. Students who received coins performed worse when their preference was coins. Those in the control group who preferred coins performed better when they did not receive that particular motivator. In terms of operant conditioning, this would mean that students may not perform well if they receive exactly what they want for long periods of time. The effect of a particular token can change over time, or it could have different effects on different students. For instance, it could play the role of reinforcer for one student while playing the role of punisher for another. It is possible that a student may be reinforced by the token in the first half of the study, but then the token could become a neutral operant that does not have any effect on students' behavior. The results on motivational preferences does have implications for goal theory. Students may perform similarly regardless of their motivational preferences, but other factors need to be taken into consideration in order to find any potential influences within the school environment. The fact that students received other motivators in addition to the ones developed for this study should not be overlooked. This may have contributed to the fact that most students scored about the same regardless of their preferences.

Summary of Findings for Research Question 3

The results for Research Question 3 showed that both teachers and students had positive reactions to tokens. They saw that there were students who responded well to them and improved their behavior. Along with the positive effects were neutral and negative effects on students. There were students who continued their normal behavior, and there were students who received more referrals over time. These observations can be applied to the theory of operant conditioning because the actual role of a particular consequence was not determined until after the participants were monitored over time. Ideally, the teachers would have liked to see all of the students increase the frequency of their good behavior. They wanted everyone to have higher test scores as the study progressed. This was the intention behind the token systems. The actual outcome was that certain students had behaviors that were positively reinforced, and others did not have the same experience. There were well-behaved students who responded well to the token systems, and there were students with behavior problems who could use them as well. It was clear that the teachers and students agreed on the fact that the token systems were used in a way that would encourage students to maintain and increase good behaviors.

Discussions pertaining to goal theory were relevant for the instances when (a) students reported their goals and preferences during the interviews and (b) teachers reported their preferences concerning reinforcement. The students were able to develop personal goals that were based on their school culture and their personal assessment of what they would be able to do. In the focus groups, most students had grade-related performance goals that were meaningful for their particular classroom environments.

Others had goals pertaining to how well they paid attention during instruction. Students had mixed reactions to the tokens, but a strong preference for points was apparent in the focus groups for seventh and eighth grade. The seventh-grade teacher's responses indicated that students preferred coins overall, but the eighth-grade teacher agreed with the interviewed students' assessment of their preferences. Teachers and students felt their preferences made a difference in how manageable the systems were. Students felt that their motivational preferences determined how interested they were in instructional activities, the prizes on the prize list, the ability to receive tokens, and their performance in the class. This does contradict what was found in the quantitative results. Teachers' reinforcement and punishment preferences indicated that tokens as a support system left much to be desired. Out of the two motivators, the points appeared to be preferred for whole-class token systems. It is possible that coins could be used as a support for small groups; however, the most appropriate form of extrinsic motivation for students was dependent upon the characteristics and perceptions of the students themselves.

Conclusions

There were three conclusions reached about this study: (a) tokens were not beneficial for everyone, (b) educators needed to account for the school environment when implementing token systems in their schools, and (c) token preferences were not strong predictors for performance and behavior. The first conclusion was derived from information within Research Question 1 and Research Question 3. The findings for these questions showed that students in the treatment group responded differently to the token intervention. There were students in the treatment groups who worked for the tokens, and there were students who did not work for them. The reactions were different for the students in the focus groups. Some felt good about the tokens, whereas others felt

indifferent to them. Reactions of indifference were most prevalent in eighth grade. Feelings of excitement were observed the most within the sixth and seventh grades. In addition, the teachers within their interviews expressed that those within the control group had different feelings about the fact that they did not receive tokens. The control group did outperform the other groups during the study, but there were students who would have liked to receive tokens. The first conclusion is important because there are studies that show conflicting results. Not all students may respond well to fixed ratio schedules or certain token types. Future reinforcement systems can incorporate different decision frameworks, different incentive types, and different incentive schedules in order to adequately address the academic and behavioral needs of all students.

The conclusion above corroborates the findings of Hayenga and Corpus (2010) where they found that high performers were not as dependent on external stimuli for performance. The study by Wulfert et al. (2002) also supported the fact that students with little to no problem behaviors were more motivated to complete tasks regardless of whether external stimuli were used. The results found further emphasize that the effectiveness of tokens are dependent upon the behavior characteristics and needs of the students in the classroom (Lovitt & Esveldt, 1970; Wulfert et al., 2002; Yager, 2008). Future research on tokens can include different combinations of ratio and interval schedules. For instance, some students could benefit more from variable ratio and variable interval schedules due to their flexibility. They may include the use of other token types as well. A response cost system, which intends to punish behavior, may be a more appropriate alternative in future designs.

The second conclusion was generated from data collected for all research questions. The groups for this study did not perform in the exact same way, and it was essential that other possible influences besides the token treatment were discovered within the data. Classroom behavior referrals tended to increase in the latter months of the study. Overall, those who received no referrals passed the pretest and posttest measures used for this study. This cannot be said for those who received referrals. The teachers in the school system incorporated the GPS standards within the instructional framework. This possibly influenced students' personal goals. There were goal orientations that changed over time. For instance, there were eighth-grade students in the focus groups who felt they performed well at the beginning of the study, but their math performance declined over time. Their expectations for their performance had lowered as the study progressed. There were sixth, seventh, and eighth grade students who felt that their performance and behavior were not influenced by token use. During the study, school policies and practices helped to determine the extent to which token systems were relevant to what was occurring within math classrooms. Each classroom and grade level had its own culture, and this in turn determined if tokens were beneficial for those involved.

There were past studies that agreed with the argument that the school environment can be an important factor to consider in research. Eccles et al. (1993) did point out that ability grouping and relationships in the classroom environment can affect motivation over time. One major motivation within the school environment was grades. Baker and Wigfield (1999) mentioned that students can be motivated by grades, and grades can be formulated into academic goals. Anderman and Maehr (1994) linked the school

environment as an influence to students' goal priorities, and Maehr (1990) stated that motivation was affected by factors within school culture. Furthermore, Self-Brown and Mathews (2003) found that students who were able to set goals on a frequent basis (i.e., weekly) were less likely to have low math performance. They were more likely to (a) gain confidence in their math abilities over time and (b) change their personal goal orientations in order to reflect changes in performance. Abramovich et al. (2013) concluded that students who received extrinsic motivators were more likely to exhibit concern over math performance, which created a performance avoidance orientation within personal goals. Future research studies about culture, reinforcement, goal orientation, and motivation can be conducted to specify more factors that can affect overall experiences with tokens and other sources of motivation. Studies may need to be conducted across subjects in order to determine if the same results are found in other areas besides math.

The third conclusion was based on findings from Research Question 2 and Research Question 3. It was determined that there were students who had high scores on tests when they did not receive what they wanted. There were also instances where students had low test scores when they frequently received the tokens they preferred. Teachers admitted that there were mixed results in terms of performance and behavior. Both teachers and students had mixed reactions about the token systems. The teachers felt that the token systems were not as integrated into the classroom practices as they could have been, and the work required was very tedious. The students felt that more choices were needed for the prize list, especially in the latter half of the 9-week quarter.

As stated in Chapter 1, there were a variety of studies that were able to outline benefits concerning tokens, particularly within the areas of student participation and academic achievement (Abramovich et al., 2013; Coyle, 2013; Miller, 1981; Truchlicka et al., 1998). For instance, Coyle (2013) outlined that middle school students enjoyed receiving tickets within PBIS. There were students who enjoyed the points and coins systems in this study, and this was especially prevalent within the seventh grade. This study, however, did have several limitations in terms of what could be implemented within the nine-week time period. Cameron and Pierce (1996) found that design restrictions can negatively impact studies about extrinsic and intrinsic motivators. The systems within this study did not fully address the needs and preferences of the participants, and the management of it was considered overwhelming. McLaughlin (1975) observed that inadequacies in terms of management and compatibility can negatively impact results. Furthermore, McClintic-Gilbert et al. (2013) found that academic achievement increased as the amount of extrinsic motivation decreased. In this study, math test averages were highest within the control group. Future studies could focus on prizes and how prizes can affect achievement in a token system. Results may not be the same if students were allowed more choice and flexibility in terms of prizes. Another area that requires attention is the need to see the extent to which instructional practices may play a role in math achievement when token systems are implemented during a particular study. Some instructional practices that require attention are the use of lectures, ability grouping, computer technology, praise, small groups, formative assessments, and differentiation in the classroom.

Recommendations

There were strengths and weaknesses found for this mixed methods study. The major strengths of this study were the (a) amount of standardization and integrity required, (b) training process, (c) ability to address different perspectives during the study, (d) amount of data accumulated concerning the research questions, and (e) ability to determine similarities and differences between theory and practice. The weaknesses included the (a) tedious management of the token systems, (b) lack of observation permitted during implementation, (c) prize list choices, (d) lack of flexibility in terms of scheduling, and (e) unanticipated ability grouping protocols that were within the school environment. Recommendations for future studies that pertain to tokens are based on all ten of these areas. Further recommendations for studies that involve classroom behavior, achievement goal orientation, and motivational preferences are stated within this section as well.

Recommendations for Teachers and Students

When implementing future research studies, there are aspects of the study procedures that students and teachers may need to assess moving forward. Teachers and students benefit from the improvements that are listed within this section. Recommendations for teachers and students are as follows:

- When implementing token systems of this magnitude, teachers and students need to use it in place of the major PBIS reinforcement system they have. During the study, the teachers were adding it onto the systems they currently had. This made their tasks more cumbersome overall. The control groups needed to use the reinforcement systems already in place, but the treatment groups could have had more of a prize selection if the old system was suspended for them during the time frame of the study.

- Have the reinforcement logs developed so that it is easy to integrate with attendance sheets and spreadsheets. A redesign of the reinforcement log form may be necessary for future studies. Students need to have a pencil pouch, binder, or personal container for their coins. This could be a part of the researcher's supply list, or teachers could make students aware of the fact they need to be able to store coins somewhere.

- Teachers and students can discuss the type of tokens they prefer to use within a selected list of options. The prizes were designed in this manner, and giving this option to the design increases the amount of flexibility teachers have when administering tokens. Students can rank their preferences, and the top two choices would be used by all teachers.

- The values in the prize list are based on the 4:1 feedback ratio, so the amount of tokens required for the prizes need to be reasonable for students using the tokens. Another prize list may need to be developed after the switch point where token values are adjusted to fit the behavior patterns and preferences of students who participate.

- The instructional activities, which include the incorporation of tokens, need to closely match the learning styles of students. Teachers could agree on a set of activities and behaviors where tokens would be applicable to students. This can make the administration of tokens more manageable in the long run.

- The extent of ability grouping within schooling must be reported to the researcher before training and random assignment in order to determine the amount of influence it may have on treatment results. If ability grouping is present, it may be necessary to focus strictly on a particular subgroup. It may not be possible to focus on particular subgroups if the number of participants are too small for statistical credibility.

Recommendations for Researchers

In light of the strengths and weaknesses within this study, suggestions have been created for the researcher. These suggestions outline possibilities for improvements to

this study. They do take into account the conceptual frameworks of operant conditioning and goal theory as well as the results found within this particular study.

- Consider extending the time of the study to include the whole school year. This study presented a snapshot of one quarter out of the school year. With additional time and funding, more data could be collected about students' behavior and performance. The results may be more meaningful to a school system if the study was yearlong. The use of tokens from onset would have quickly established them as a typical part of school culture.

- The use of remediation and follow-up strategies could be useful in tracking how students perform and behave after treatment. For instance, students and teachers could fill out a survey or check sheet about their overall experiences. This could help the researcher gain more insights about the project.

- Adding a performance measure during the switch point can help to glean knowledge about students' performance over time. Due to the time constraints and limitations within this study, it was decided that referral information would be collected at that time. Another idea is to utilize the motivational preferences check sheet during the pretest and switch point time frames. This would help document students' preferences with tokens at the beginning and the middle of the study. More trends could be established within the data.

- Researchers can have observation sessions in the classroom to see how teachers implement the token systems and prize lists. The amount of interaction that researchers have during observations would depend on their established research roles. The amount of recording that is acceptable during such events would be dependent on what students, teachers, and administrators consented to at the beginning of the study.

- Consider using a combination of interval and ratio schedules during the study. The fixed ratio schedule was not beneficial for all students. Future schedules may want to explore options with interval scheduling, or teachers may want to use a variable ratio schedule to meet students' instructional needs.

- Consider adding research questions about vicarious reinforcement, referral punishment systems, and response cost systems. These areas were not thoroughly explored during this study. The amount of questions developed depend on the systems that will be implemented during a particular study.

- Consider adding research questions pertaining to the cultural habits and instructional techniques within schools. This information could be obtained by interviewing teachers or administrators. This study did accumulate information about various habits and techniques, but not all of them were relevant to the research questions.

The recommendations provided for future studies on tokens help to inform teachers, students, and researchers on how they can incorporate a standardized behavioral support system into the overall learning environment. They are based on well-documented research procedures and teaching practices that are relevant to Georgia schools. The school environment is frequently changing in order to address the needs of all students. Token media use is an evidence-based practice that can be utilized by teachers and administrators when they are implementing one or more decision-making frameworks (e.g., RTI and PBIS).

Token reinforcement is not the only instructional solution that is available for educators, yet it has been overlooked to a certain degree within past research. By implementing these recommendations, educators can provide more motivation, flexibility, inclusivity, and collaboration within their respective classrooms. Researchers who follow the suggestions above would be able to connect theory with practice in ways that maintain the integrity and rigor required for credible studies. This study and the concepts contained within it are of great interest to educators, students, and researchers in the fields of education and psychology.

REFERENCES

Abramovich, S., Schunn, C., & Higashi, R. M. (2013). Are badges useful in education?: It depends upon the type of badge and expertise of learner. *Educational Technology Research & Development*, *61*(2), 217-232. doi:10.1007/s11423-013-9289-2

Akin-Little, K. A., Eckert, T. L., Lovett, B. J., & Little, S. G. (2004). Extrinsic reinforcement in the classroom: Bribery or best practice. *School Psychology Review, 33*(3), 344-362.

Alberto, P., Heflin, L. J., & Andrews, D. (2002). Use of the timeout ribbon procedure during community-based instruction. *Behavior Modification*, *26*(2), 297-311.

American Psychological Association. (2010). *Ethical principles of psychologists and code of conduct.* Retrieved from http://www.apa.org/ethics/code/principles.pdf

American Psychological Association, Coalition for Psychology in Schools and Education. (2015). *Top 20 principles from psychology for preK–12 teaching and learning.* Retrieved from http://www.apa.org/ed/schools/cpse/top-twenty-principles.pdf

Ames, C. (1984). Achievement attributions and self-instructions under competitive and individualistic goal structures. *Journal of Educational Psychology, 76*(3), 478-487.

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*(3), 260-267.

Anderman, E. M., & Maehr, M. L. (1994). Motivation and schooling in the middle grades. *Review of Educational Research*, *64*(2), 287-309.

Anderson, E., McLaughlin, T. F., Derby, K. M., & Williams, M. (2012). The effects of contingent rewards on the decrease of improper verbalizations of middle school students with learning disabilities. *Academic Research International, 2*(1), 366-371.

Baker, L., & Wigfield, A. (1999). Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. *Reading Research Quarterly*, *34*(4), 452-477. doi:10.1598/RRQ.34.4.4

Barringer, C., & Gholson, B. (1979). Effects of type and combination of feedback upon conceptual learning by children: Implications for research in academic learning. *Review of Educational Research, 49*(3), 459-478.

Batuev, A. S., & Sokolova L. V. (2004). From physiological theory to psychological facts (celebrating the 100th anniversary of I. P. Pavlov's Madrid speech). *Neuroscience and Behavioral Physiology, 34*(7), 711-720.

Baum, W. M. (2011). What is radical behaviorism? A review of Jay Moore's conceptual foundations of radical behaviorism. *Journal of the Experimental Analysis of Behavior, 95*(1), 119-126. doi: 10.1901/jeab.2011.95-119

Benabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies, 70*(244), 489-520.

Borrero, C., Vollmer, T. R., Borrero, J. C., Bourret, J. C., Sloman, K. N., Samaha, A. L., & Dallery, J. (2010). Concurrent reinforcement schedules for problem behavior and appropriate behavior: Experimental applications of the matching law.

Journal of the Experimental Analysis of Behavior, 93(3), 455-469. doi: 10.1901/jeab.2010.93-455

Bullock, C. E. (2006). *Stimulus functions in token-reinforcement schedules* (Doctoral dissertation). ProQuest Dissertations and Theses database. (Order No. 3224511)

Calkins, M. W. (1921). The truly psychological behaviorism. *Psychological Review, 28*(1), 1-18.

Cameron, J., & Pierce, W. D. (1996). The debate about rewards and intrinsic motivation: Protests and accusations do not alter the results. *Review of Educational Research*, *66*(1), 39-51.

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, *140*(4), 980-1008. doi:10.1037/a0035661

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology, 51*(1), 171-200. doi:10.1146/annurev.psych.51.1.171

Coyle, L. A. (2013). *Students' lived experiences with the positive behavioral interventions and supports (PBIS) program in middle school* (Doctoral dissertation). ProQuest Dissertations and Theses database. (Order No. 3596107)

Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (3rd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and Conducting Mixed Methods Research* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Cronk, B. C. (2012). *How to Use SPSS: A Step-by-step Guide to Analysis and Interpretation* (7th ed.). Glendale, CA: Pyrczak Publishing.

Cross, L. M. (1981). *Effects of a token economy program in a continuation school on student behavior and attitudes* (Doctoral dissertation). ProQuest Dissertations and Theses database. (Order No. 8124477)

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research, 71*(1), 1–27. doi:10.3102/00346543071001001

Devers, R., & Bradley-Johnson, S. (1994). The effect of token reinforcement on WISC-R performance for fifth-through ninth-grade American Indians. *Psychological Record*, *44*(3), 441-449.

Diener, C. I., & Dweck, C. S. (1978). An analysis of learned helplessness: Continuous changes in performance, strategy, and achievement cognitions following failure. *Journal of Personality and Social Psychology*, *36*(5), 451-462. doi:10.1037/0022-3514.36.5.451

Dowson, M., & McInerney, D. M. (2003). What do students say about their motivational goals?: Towards a more complex and dynamic perspective on student motivation. *Contemporary Educational Psychology*, *28*(1), 91-113.

Eccles, J. S., Wigfield, A., Midgley, C., Reuman, D., Iver, D. M., & Feldlaufer, H. (1993). Negative effects of traditional middle schools on students' motivation. *The Elementary School Journal*, *93*(5), 553-574.

Elliot, A. J., & Covington M. V. (2001). Approach and avoidance motivation. *Educational Psychology Review, 13*(2), 73-92.

Ferster, C. B., & Skinner, B. F. (1957). *Schedules of Reinforcement.* New York, NY: Appleton-Century-Crofts, Inc.

Foster, T. A., Hackenberg, T. D., & Vaidya, M. (2001). Second-order schedules of token reinforcement with pigeons: Effects of fixed- and variable-ratio exchange schedules. *Journal of the Experimental Analysis of Behavior, 76*(2), 159–178. doi:10.1901/jeab.2001.76-159

Gaughan, E. J. (1985). *The relationship between point earning behavior and academic achievement in a token economy for emotionally disturbed children* (Doctoral dissertation). ProQuest Dissertations and Theses database. (Order No. 8509364)

Georgia Department of Education. (2011, September 19). *Response to Intervention: Georgia's student achievement pyramid of interventions.* Retrieved from http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Curriculum-and-Instruction/Documents/RTI%20document%20Full%20Text.pdf

Georgia Department of Education. (2013, June). *Addressing climate, safety, and discipline in Georgia schools: School-wide positive behavioral interventions and supports.* Retrieved from http://www.gadoe.org/Curriculum-Instruction-and-Assessment/Special-Education-Services/Documents/PBIS/PBIS%20Final%20white%20paper_%20Sept%204.pdf

Goodman, G. (Ed.). (2010). *Educational Psychology Reader: The Art and Science of How People Learn*. New York, NY: Peter Lang.

Guba, E. G. (Ed.) (1990). *The Paradigm Dialogue.* Newbury Park, CA: Sage Publications, Inc.

Habaibeh-Sayegh, S. (2014). *The effectiveness of a token economy program in improving behavior and achievement* (Doctoral dissertation). ProQuest Dissertations and Theses database. (Order No. 3630049)

Hackenberg, T. D. (2009). Token reinforcement: A review and analysis. *Journal of the Experimental Analysis of Behavior*, *91*(2), 257–286. doi:10.1901/jeab.2009.91-257

Hagopian, L. P., Kuhn, D. E., & Strother, G. E. (2009). Targeting social skills deficits in an adolescent with pervasive developmental disorder. *Journal of Applied Behavior Analysis, 42*(4), 907-911. doi: 10.1901/jaba.2009.42-907

Hansen, S. D., & Lignugaris/Kraft, B. (2005). Effects of a dependent group contingency on the verbal interactions of middle school students with emotional disturbance. *Behavioral Disorders, 30*(2), 170-184.

Haselhuhn, C. W., Al-Mabuk, R., Gabriele, A., Groen, M., & Galloway, S. (2007). Promoting positive achievement in the middle school: A look at teachers' motivational knowledge, beliefs, and teaching practices. *Research in Middle Level Education Online*, *30*(9), 1-20.

Hayenga, A., & Corpus, J. (2010). Profiles of intrinsic and extrinsic motivations: A person-centered approach to motivation and achievement in middle school. *Motivation & Emotion*, *34*(4), 371-383. doi:10.1007/s11031-010-9181-x

Heidegger, M. (1962). *Being and Time* [Google Books version]. Retrieved from http://books.google.com/books

Hersen, M., & Gross, A. M. (2008). *Handbook of Clinical Psychology* (Vol. 2). Hoboken, NJ: John Wiley & Sons, Inc.

Hoeltzel, R. C. (1973). *Reading rates and comprehension as affected by single and multiple-ratio schedules of reinforcement within a token economy as measured by precision teaching techniques.* (Doctoral dissertation). ProQuest Dissertations and Theses database. (Order No. 74-9066)

Husserl, E. (1931). *Ideas: General Introduction to Pure Phenomenology* (W. R. Boyce Gibson, Trans.). London: George Allen & Unwin Ltd.

Iverson, M. K. (2010). *Evaluation of a token economy's effectiveness in a self-contained classroom in a psychiatric rehabilitation facility with secondary students diagnosed with emotional and behavioral disabilities* (Doctoral dissertation). ProQuest Dissertations and Theses database. (Order No. 3420563)

Joosten, A., Bundy, A., & Einfeld, S. (2009). Intrinsic and extrinsic motivation for stereotypic and repetitive behavior. *Journal of Autism & Developmental Disorders*, *39*(3), 521-531. doi:10.1007/s10803-008-0654-7

Kazdin, A. E. (1982). The token economy: A decade later. *Journal of Applied Behavior Analysis, 15*(3), 431–445. doi:10.1901/jaba.1982.15-431

Kazdin, A. E., & Bootzin, R. R. (1972). The token economy: An evaluative review. *Journal of Applied Behavior Analysis*, *5*(3), 343–372. doi:10.1901/jaba.1972.5-343

Kierkegaard, S. (1992). *Concluding unscientific postscript to the philosophical fragments* [Google Books version]. Retrieved from http://books.google.com/books

Knoster, T. P. (2014). *The Teacher's Pocket Guide for Effective Classroom Management* (2nd ed.). Baltimore, MD: Brookes Publishing.

Krueger, R. A., & Casey, M. A. (2015). *Focus Groups: A Practical Guide for Applied Research* (5th ed.). Thousand Oaks, CA: Sage Publications, Inc.

Lee, R., Sturmey, P., & Fields, L. (2007). Schedule-induced and operant mechanisms that influence response variability: A review and implications for future investigations. *Psychological Record, 57*(3), 429-455.

Lewis-Lancaster, A., & Reisener, C. (2013). Examining the results of a brief experimental analysis and reading fluency intervention with a middle school student. *Reading Improvement*, *50*(4), 166-174.

Lovitt, T. C., & Esveldt, K. A. (1970). The relative effects on math performance of single- versus multiple-ratio schedules: A case study. *Journal of Applied Behavior Analysis, 3*(4), 261-270. doi:10.1901/jaba.1970.3-261

Lynch, A., Theodore, L. A., Bray, M. A., & Kehle, T. J. (2009). A comparison of group-oriented contingencies and randomized reinforcers to improve homework completion and accuracy for students with disabilities. *School Psychology Review, 38*(3), 307-324.

Maag, J. W. (2001). Rewarded by punishment: Reflections on the disuse of positive reinforcement in schools. *Exceptional Children*, *67*(2), 173-186.

Mace, F. C., Pratt J. L., Zangrillo, A. N., & Steege, M. W. (2011). Schedules of reinforcement. In W. W. Fisher, C. C. Piazza, & H. S. Roane (Eds.), *Handbook of Applied Behavior Analysis.* New York, NY: The Guilford Press.

Machado, R. P. (2008). Nothingness and the work of art: A comparative approach to existential phenomenology and the ontological foundation of aesthetics. *Philosophy East & West*, *58*(2), 244–266.

Madsen, C. H., Madsen, C. K., Saudargas, R. A., Hammond, W. R., Smith, J. B., & Edgar, D. E. (1970). Classroom raid (rules, approval, ignore, disapproval): A cooperative approach for professionals and volunteers. *Journal of School Psychology, 8*(3), 180-185. doi:10.1016/0022-4405(70)90070-1

Maehr, M. L. (1990). *The psychological environment of the school: A focus for school leadership* (Contract No. RI17C80003). Urbana, IL: National Center for School Leadership.

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. doi:10.1016/j.jsp.2011.05.001

Marinak, B. A., & Gambrell, L. B. (2008). Intrinsic motivation and rewards: What sustains young children's engagement with text? *Literacy Research and Instruction, 47*(1), 9-26.

McClintic-Gilbert, M., Corpus, J. H., Wormington, S. V., & Haimovitz, K. (2013). The relationships among middle school students' motivational orientations, learning strategies, and academic achievement. *Middle Grades Research Journal, 8*(1), 1-12.

McDonald, M. E., Reeve, S. A., & Sparacio, E. J. (2014). Using a tactile prompt to increase instructor delivery of behavior-specific praise and token reinforcement and their collateral effects on stereotypic behavior in students with autism spectrum disorders. *Behavioral Development Bulletin*, *19*(1), 40-44.

McLaughlin, T. F. (1975). The applicability of token reinforcement systems in public school systems. *Psychology in the Schools, 12*(1)*,* 84–89.

Mertler, C. A., & Vannatta, R. A. (2013). *Advanced and Multivariate Statistical Methods: Practical Application and Interpretation* (5th ed.). Glendale, CA: Pyrczak Publishing.

Miller, J., & Eller, B. F. (1985). An examination of the effect of tangible and social reinforcers on intelligence test performance of middle school students. *Social Behavior & Personality: An International Journal*, *13*(2), 147-157.

Miller, J. B. (1981). *The effects of selected motivational rewards on intelligence test performance of middle school students* (Doctoral dissertation). ProQuest Dissertations and Theses database. (Order No. 8128357)

Mongillo, G. (2006). *Instructional games: Scientific language use, concept understanding, and attitudinal development of middle school learners* (Doctoral dissertation). ProQuest Dissertations and Theses database. (Order No. 3213879)

Moore, J. (2011). Behaviorism. *Psychological Record, 61*(3), 449-465.

Morgan, D. L. (2010). Schedules of reinforcement at 50: A retrospective appreciation. *Psychological Record, 60*(1), 151-172.

Mucherah, W., & Yoder, A. (2008). Motivation for reading and middle school students' performance on standardized testing in reading. *Reading Psychology*, *29*(3), 214-235. doi:10.1080/02702710801982159

National Center for Education Statistics. (2012). *Public elementary/secondary school universe survey data, 2011-2012 [Data files and code files]*. Retrieved from http://nces.ed.gov/ccd/pubschuniv.asp

National Center for Education Statistics. (2016). *The Condition of Education.* Retrieved from https://nces.ed.gov/programs/coe/pdf/coe_cgg.pdf

National Governors Association Center for Best Practices, Council of Chief State School Officers. (n.d.). *Frequently asked questions.* Retrieved from http://www.corestandards.org/wp-content/uploads/FAQs.pdf

Nelson, K. G. (2010). Exploration of classroom participation in the presence of a token economy. *Journal of Instructional Psychology*, *37*(1), 49-56.

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

Parish, J. G., & Parish, T. S. (1991). Rethinking conditioning strategies: Some tips on how educators can avoid "painting themselves into a corner." *Journal of Instructional Psychology*, *18*(3), 159-166.

Paschalis, A. P. (1987). Tokens and their economies: The Greeks had a use for them. *Journal of Applied Behavior Analysis, 20*(4), 427. doi:10.1901/jaba.1987.20-427

Patton, M. Q. (2002). *Qualitative Research and Evaluation Methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Pavlov, I. P. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex* [Google Books version]. Retrieved from http://books.google.com/books

Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology, 92*(3), 544-555. doi: 10.1037/0022-0663.92.3.544

Popkin, J., & Skinner, C. H. (2003). Enhancing academic performance in a classroom serving students with serious emotional disturbance: Interdependent group contingencies with randomly selected components. *School Psychology Review, 32*(2), 282-295.

Punch, K. F. (2014). *Introduction to Social Research: Qualitative and Quantitative Approaches.* London, England: Sage Publications.

Robustelli, S. L. (2006). *There's more than one yellow brick road: Examining revised achievement goal theory and its efficacy in explaining teachers' goals, students' goals, and achievement* (Doctoral dissertation). ProQuest Dissertations and Theses database. (Order No. 3240264)

Sansosti, F. J. (2012). Reducing the threatening and aggressive behavior of a middle school student with Asperger's Syndrome. *Preventing School Failure*, *56*(1), 8-18. doi:10.1080/1045988X.2010.548418

Sartre, J. P. (1943). *Being and Nothingness: The Principal Text of Modern Existentialism.* New York, NY: Gallimard.

Schlinger, H. D., Derenne, A., & Baron, A. (2008). What 50 years of research tell us about pausing under ratio schedules of reinforcement. *The Behavior Analyst, 31*(1), 39-60.

Schmandt-Besserat, D. (1992). *Before Writing: Vol. 1. From Counting to Cuneiform.* [Google Books version]. Retrieved from http://books.google.com/books.

Schneider, S. M., & Morris, E. K. (1987). A history of the term radical behaviorism: From Watson to Skinner. *The Behavior Analyst, 10*(1), 27-39.

Schunk, D. H. (2012). *Learning Theories: An Educational Perspective* (6th ed.). Boston, MA: Allyn & Bacon.

Seidman, I. (2006). *Interviewing as Qualitative Research: A Guide for Researchers in Education and the Social Sciences* (3rd ed.). New York, NY: Teachers College Press.

Self-Brown, S. R., & Mathews, I. (2003). Effects of classroom structure on student achievement goal orientation. *Journal of Educational Research*, *97*(2), 106-111.

Sheldon, K. M., & Kasser, T. (2008). Psychological threat and extrinsic goal striving. *Motivation and Emotion, 32*(1), 37-45. doi:10.1007/s11031-008-9081-5

Simon, S. J., Ayllon, T., & Milan, M. A. (1982). Behavioral compensation. *Behavior Modification*, *6*(3), 407-420.

Skinner, B. F. (1938). *The Behavior of Organisms: An Experimental Analysis*. New York, NY: Appleton-Century-Crofts, Inc.

Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Review, 52*(5), 270-277. doi:10.1037/h0062535

Skinner, B. F. (1953). *Science and Human Behavior* [Google Books version]. Retrieved from http://books.google.com/books

Skinner, B. F. (1958). Reinforcement today. *American Psychologist, 13*(3), 94-99. doi:10.1037/h0049039

Skinner, C. H., Williams, R. L., & Neddenriep, C. E. (2004). Using interdependent group-oriented reinforcement to enhance academic performance in general education classrooms. *School Psychology Review, 33*(3), 384-397.

Stage, S. A., & Quiroz, D. R. (1997). A meta-analysis of interventions to decrease disruptive classroom behavior in public education settings. *School Psychology Review, 26*(3), 333-368.

Strahan, D. B., & Layell, K. (2006). Connecting caring and action through responsive teaching: How one team accomplished success in a struggling middle school. *The Clearing House, 79*(3), 147-153.

Swain, J. C., & McLaughlin, T. F. (1998). The effects of bonus contingencies in a classwide token program on math accuracy with middle-school students with behavioral disorders. *Behavioral Interventions*, *13*(1), 11-19.

Taylor, D. L. (2000). *The effect of concurrent variable interval reinforcement schedules on children with attention deficit hyperactivity disorder and normal control children* (Doctoral dissertation). ProQuest Dissertations and Theses database. (Order No. 9974730)

Thomas, R. M. (2003). *Blending Qualitative and Quantitative Research Methods in Theses and Dissertations* [Google Books version]. Retrieved from http://books.google.com/books

Thompson, C. J., Locander, W. B., & Pollio, H. R. (1989). Putting consumer experience back into consumer research: The philosophy and method of existential-phenomenology. *Journal of Consumer Research*, *16*(2), 133–146. Retrieved from https://faculty.fuqua.duke.edu/~moorman/GeneralMills/Section1/Section1Documents/D.pdf

Thorndike, E. L. (1913). *Educational Psychology: The Psychology of Learning* [Google Books version]. Retrieved from http://books.google.com/books

Tolman, E. C. (1932). *Purposive Behavior in Animals and Men* [Google Books version]. Retrieved from http://books.google.com/books

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55*(4), 189-208.

Truchlicka, M., McLaughlin, T. F., & Swain, J. C. (1998). Effects of token reinforcement and response cost on the accuracy of spelling performance with middle-school special education students with behavior disorders. *Behavioral Interventions*, *13*(1), 1-10.

Tuckman, B. W., & Monetti, D. M. (2011). *Educational Psychology*. Belmont, CA: Wadsworth, Cengage Learning.

Unrau, N., & Schlackman, J. (2006). Motivation and its relationship with reading achievement in an urban middle school. *Journal of Educational Research*, *100*(2), 81-101.

Urdan, T., & Midgley, C. (2003). Changes in the perceived classroom goal structure and pattern of adaptive learning during early adolescence. *Contemporary Educational Psychology*, *28*(4), 524-551. doi:10.1016/S0361-476X(02)00060-7

U.S. Department of Education Office for Civil Rights. (2014, March). *Data snapshot: School discipline* (Issue Brief No. 1). Retrieved from Civil Rights Data Collection website: http://ocrdata.ed.gov/Downloads/CRDC-School-Discipline-Snapshot.pdf

Waddell, T. R., Leander, J. D., Webbe, F. M., & Malagodi, E. F. (1972). Schedule interactions in second-order fixed-interval (fixed-ratio) schedules of token reinforcement. *Learning and Motivation, 3*(1), 91–100.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review, 20*(2), 158–177.

Watson, J. B., & Raynor, R. (1920). Conditioned emotional reactions. *Journal of Experimental Psychology*, *3*(1), 1-14.

Webbe, F. M., & Malagodi, E. F. (1978). Second-order schedules of token reinforcement: Comparisons of performance under fixed-ratio and variable-ratio exchange schedules. *Journal of the Experimental Analysis of Behavior, 30*(2), 219–224. doi:10.1901/jeab.1978.30-219

Windholz, G. (1997). Ivan P. Pavlov: An overview of his life and psychological work. *American Psychologist, 52*(9), 941-946.

Wolfe, B. D., Dattilo, J., & Gast, D. L. (2003). Effects of a token economy system within the context of cooperative games on social behaviors of adolescents with emotional and behavioral disorders. *Therapeutic Recreation Journal, 37*(2), 124-141.

Wulfert, E., Block, J. A., Santa Ana, E., Rodriguez, M. L., & Colsman, M. (2002). Delay of gratification: Impulsive choices and problem behaviors in early and late adolescence. *Journal of Personality*, *70*(4), 533-552.

Yager, L. (2008). *The relationship between Mississippi school-based rewards programs and the behaviors of 6th grade students.* (Doctoral dissertation). ProQuest Dissertations and Theses database. (Order No. 3358526)

Young-Welch, C. (2008). *A mixed-method study utilizing a token economy to shape behavior and increase academic success in urban students* (Doctoral dissertation). ProQuest Dissertations and Theses database. (Order No. 3320692)

Zeiler, M. (1977). Schedules of reinforcement. In W. K. Wong & J. E. R. Staddon (Eds.), *Handbook of operant behavior.* Englewood Cliffs, NJ: Prentice Hall.

APPENDIX A:

Data Collection Timetable

Data Collection Timetable

Week	Quantitative Strand	Qualitative Strand
Week 1	1. Gathering Pretest, Goal Orientation, and Referral Data 2. Collect Consent/Assent Forms 3. Start of Quasi-experiment	1. Gathering Participant Info 2. Recruitment for Interviews
Week 2	1. Collect Reinforcement Records	1. Developing Instruments 2. Analysis of Data
Week 3	1. Developing Instruments 2. Analysis of Data 3. Collect Reinforcement Records	1. Interviews for Teachers and Students – Part 1
Week 4	1. Documenting Procedures 2. Analysis of Data 3. Collect Reinforcement Records	1. Documenting Transcription, Coding, and Categorizing 2. Analysis of Data
Week 5	1. Summary of Records and Data 2. Email Summary to Teachers 3. Switch Token Treatments 4. Collect Reinforcement Records	1. Transcriptions Complete for Part 1 2. Mail or Email Requested Copies 3. Developing Instruments
Week 6	1. Collect Reinforcement Records	1. Interviews for Teachers and Students – Part 2
Week 7	1. Documenting Procedures 2. Analysis of Data 3. Developing Instruments 4. Collect Reinforcement Records	1. Documenting Transcription, Coding, and Categorizing 2. Analysis of Data 3. Developing Instruments
Week 8	1. Scheduling Testing for Posttest 2. Collect Reinforcement Records	1. Transcriptions Complete for Part 2 2. Mail or Email Requested Copies
Week 9	1. Posttest Administered 2. Distribute Check Sheet 3. Collect Reinforcement Records	1. Interviews for Students and Teachers – Part 3
Week 10	1. Treatment Ends 2. Documenting Procedures 3. Analysis of Data	1. Documenting Transcription, Coding, and Categorizing 2. Analysis of Data
Week 11	1. Summary of Records and Data 2. Email Summary to Teachers	1. Transcriptions Complete for Part 3 2. Mail or Email Requested Copies 3. Email Summary to Teachers

Note. School visitation sessions are weekly.

APPENDIX B:

Goal Orientation Questionnaire

Goal Orientation Questionnaire

Name: _____ Date: _____

Gender: _____ Race: _____

Directions: Read over the items and answer in complete sentences.

1. What is your opinion about math class?

2. How well do you do on math tests?

3. How do you think you will do on the tests for this quarter?

4. List two personal goals that you have for math class.

   1.

   2.

5. How long do you think it will take you to complete the goals in question #4?

   Goal 1:

   Goal 2:

APPENDIX C:

Reinforcement Observation Log

Reinforcement Observation Log

Teacher's Name: _____ Subject: _____ Grade: _____

Student Name	Date	# of Tokens or Prizes	Reasons	# of Responses
John Smith	8/10 (Tues.)	1 Point	On task, raised hand, seated properly, answered questions correctly	4
Jane Doe	8/10 (Tues.)	1 Pencil	Exchanged four tokens	4

Total Tokens Awarded: _____ Total Prizes Awarded: ____

Key: Reasons = Appropriate Behaviors during class
Tokens = Points or Coins
4 Reasons = 1 Token
Prizes = Items Purchased with Tokens

APPENDIX D:

Motivational Preferences Check Sheet

Motivational Preferences Check Sheet

Name: _____ Date: _____

(This is to be passed out to ALL students after the quasi-experiment is complete to document personal preferences for each student participant.)

Directions – Out of the choices listed, put an X next to the one that you like the best.

_____ Points

_____ Coins

_____ Both

_____ Neither

APPENDIX E:

Teacher Consent Form

VALDOSTA STATE UNIVERSITY

Consent to Participate in Research

You are being asked to participate in a research project entitled "Quasi-Experimental Study of Middle School Tokens, Behavior, Goals, Preferences, and Academic Achievement." This research project is being conducted by Kelly Dreger, a graduate student in Curriculum and Instruction at Valdosta State University. The researcher has explained to you in detail the purpose of the project, the procedures to be used, and the potential benefits and possible risks of participation. You may ask the researcher any questions you have to help you understand this project and your possible participation in it. A basic explanation of the research is given below. Please read this carefully and discuss with the researcher any questions you may have. The University asks that you give your signed agreement if you wish to participate in this research project.

Purpose of the Research: This study involves research. The purpose of the study is to determine if token use produces any effects on end-of-quarter test scores and classroom behavior.

Procedures: You will implement a classroom token system (quasi-experiment) and participate in a focus group. These activities will be completed in approximately one quarter (nine weeks). The token system requires that points, coins, and prizes are given to students for appropriate behaviors. You will receive training in order to understand how to implement it. Not all blocks will receive the same system, and certain blocks will receive no tokens. You will be asked to keep a log of all the tokens and prizes you give to students over time. During the training session, refreshments will be served due to the nature and timing of the session.

Throughout the process, data about goals, suspensions, referrals, preferences, and end-of-quarter test scores will be collected. The time in which the data is collected would depend on its availability. Ideally, the researcher would like to observe your classroom and gather data periodically (i.e., weekly) in order to receive ongoing feedback about what is occurring. You will be interviewed about your experiences with tokens as well. The focus group interview is about an hour in length, and it will be scheduled at your convenience. Topics include your past experiences with tokens, your current practices in the classroom, and your opinions about tokens as an educational strategy. This discussion will be recorded with a digital voice recorder, and the researcher will write notes about what takes place.

Possible Risks or Discomfort: Because the amount of tokens awarded is not the same for all students, a minimal amount of uneasiness may occur in implementing the system at first. This would be lessened as the system becomes a part of day-to-day routines. A minimal amount of embarrassment could occur during the focus group if sensitive information is discussed. You

have the option to request that sensitive information is removed during transcription. If requested, copies of the transcripts will be made available for your review.

If you experience psychological distress as a result of your participation in this study, please contact the researcher at (229)-436-6425 or at kelly6754@yahoo.com. Neither the researcher nor Valdosta State University has made special provisions for services required to treat any psychological distress that results from participation in this research study.

By agreeing to participate in this research project, you are not waiving any rights that you may have against Valdosta State University for injury resulting from negligence of the University or its researchers.

Potential Benefits: You will be able to utilize and assess an educational strategy that will help to inform your teaching practices. The use of tokens would help to monitor behavior and reinforce performance goals. Your participation will help the researcher gain further insight about instructional experiences within today's schools.

Costs and Compensation: You will have to provide your own method of transportation in order to arrive at and depart from the school in which you work. All activities will take place in the school setting. No compensation (no money, gifts, or services) will be given for your participation in this research project.

Assurance of Confidentiality: Valdosta State University and the researcher will keep your information confidential to the extent allowed by law. Members of the Institutional Review Board (IRB), a university committee charged with reviewing research to ensure the rights and welfare of research participants, may be given access to your confidential information.

Hardcopies of confidential data will be kept in a locked security box. The voice recorder will be placed in this box as well. The researcher will be the only one who has the keys to these locked files. If you request hardcopies of your focus group transcript, then they will be mailed to your address. Electronic data and recorded audio files will be placed on a password-protected computer that the researcher has at home. The researcher will have full access to the information. You will be notified in advance if parents, administrators, or committee members request an electronic copy of data. If you request electronic copies of your transcripts, then they will be emailed to you at the e-mail address you provide.

All data will be kept for three years after the researcher graduates. This is a requirement of the IRB. Participants can receive copies of the results if requested at a future date. After three years have passed, all hardcopies will be shredded. Electronic information will be deleted from the computer. When reporting data from the study, the names of individual participants will not be used. Individual results will be reported in combination with data retrieved from other participants.

Voluntary Participation: Your decision to participate in this research project is entirely voluntary. If you agree now to participate and change your mind later, you are free to leave the study. Your decision not to participate at all or to stop participating at any time in the future will not have any effect on any rights you have or any services you are otherwise entitled to from Valdosta State University. During the focus group, you may skip any questions that you do

not want to answer. If you decide to withdraw during or after data collection stages, your information will be deleted and will not be included in the research results.

Information Contacts: Questions regarding the purpose or procedures of the research should be directed to Kelly Dreger at (229)-436-6425 or kelly6754@yahoo.com. Also, you may contact the supervising faculty, Dr. Steve Downey, at sedowney@valdosta.edu for additional information. This study has been approved by the Valdosta State University Institutional Review Board (IRB) for the Protection of Human Research Participants. The IRB, a university committee established by Federal law, is responsible for protecting the rights and welfare of research participants. If you have concerns or questions about your rights as a research participant, you may contact the IRB Administrator at (229)-333-7837 or irb@valdosta.edu.

Agreement to Participate: The research project and my role in it have been explained to me, and my questions have been answered to my satisfaction. I agree to participate in this study. By signing this form, I am indicating that I am 18 years of age or older. I have received a copy of this consent form.

I would like to receive a copy of my focus group transcript: _____ Yes _____ No

I would like to receive a copy of the results of this study: _____ Yes _____ No

Mailing Address: _____

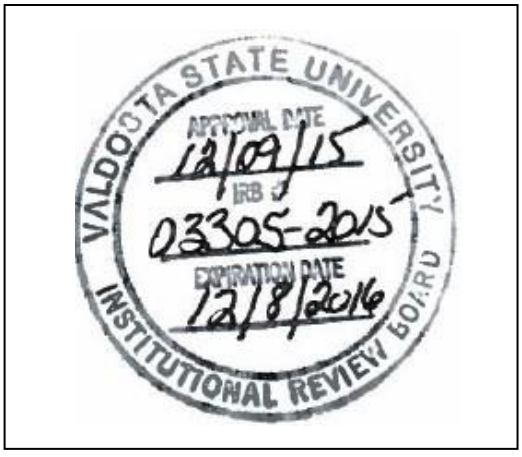
E-mail Address: _____

_____
Printed Name of Participant

_____
Signature of Participant Date

_____
Signature of Person Obtaining Consent Date

This research project has been approved by the Valdosta State University Institutional Review Board for the Protection of Human Research Participants through the date noted below:



APPENDIX F:

Parental Consent Form

VALDOSTA STATE UNIVERSITY

Parent/Guardian Permission for Child's/Ward's Participation in Research

You are being asked to allow your child (or ward) to participate in a research project entitled "Quasi-Experimental Study of Middle School Tokens, Behavior, Goals, Preferences, and Academic Achievement." This research project is being conducted by Kelly Dreger, a graduate student in Curriculum and Instruction at Valdosta State University. The researcher has explained to you in detail the purpose of the project, the procedures to be used, and the potential benefits and possible risks to your child (or ward). You may ask the researcher any questions you have to help you understand this study and your child's (or ward's) possible participation in it. A basic explanation of the research is given below. From this point on in this form, the term "child" is used for either a child or a ward. Please read the remainder of this form carefully and ask the researcher any questions you may have. The University asks that you give your signed permission if you will allow your child to participate in this research project.

Purpose of the Research: This study involves research. The purpose of the study is to determine if token use produces any effects on achievement tests and classroom behavior.

Procedures: Your child is eligible to receive tokens during a research study. If he/she receives tokens during the study, then he/she may participate in a focus group about the experience. These activities will be completed in approximately one quarter (nine weeks). During classroom tasks, he/she may receive points, coins, or prizes for appropriate behaviors. Data will be collected about suspensions, referrals, preferences, and end-of-semester test performance. Your child will be observed periodically in order to collect information for the study. The focus group interview, if applicable, would occur during school hours. It will last approximately one hour, and you will be notified in advance about when it will take place. The session will be recorded with a digital voice recorder, and the researcher will write notes about what takes place.

Possible Risks or Discomfort: Because the amount of tokens awarded is not the same for all students, a minimal amount of frustration may occur for those who receive little or no tokens. For those who do not receive any tokens, other instructional strategies will be used to make sure they receive positive feedback for their efforts. A minimal amount of embarrassment could occur during the focus group if sensitive information is discussed. You have the option to request that sensitive information is removed from transcripts. You can request copies of the transcripts to review the contents.

If your child experiences psychological distress as a result of his/her participation in this study, please contact the researcher at (229)-436-6425 or at kelly6754@yahoo.com. Neither the researcher nor Valdosta State University has made special provisions for services required to

treat any psychological distress that your child may suffer as a result of participating in this research study.

By granting permission for your child to participate in this research project, you are not waiving any rights that you or your child may have against Valdosta State University for injury resulting from negligence of the University or its researchers.

Potential Benefits: Your child will experience a strategy of learning that is based on their individual needs and behaviors. In the future, you may decide to use tokens as a way to link what occurs at school with what occurs at home. Your child's participation will help the researcher gain further insight about instructional experiences within today's schools.

Costs and Compensation: You will have to provide transportation for your child in order for him/her to arrive and depart from school. All activities will take place in the school setting. No compensation (no money, gifts, or services) will be given for your participation in this research project.

Assurance of Confidentiality: Valdosta State University and the researcher will keep your child's information confidential to the extent allowed by law. Members of the Institutional Review Board (IRB), a university committee charged with reviewing research to ensure the rights and welfare of research participants, may be given access to your child's confidential information.

With the teacher's approval, hardcopies of your child's confidential data will be kept in a locked security box. The voice recorder will be placed in this box as well. The researcher will be the only one who has the keys to these locked files. If you request hardcopies of your child's focus group transcript, then they will be mailed to your address. Electronic data and recorded audio files will be placed on a password-protected computer that the researcher has at home. The researcher will have full access to the information. You can notify the researcher or the teacher if you would like to have summaries of the research results. If you request electronic copies of your child's transcripts, then they will be emailed to you at the e-mail address you provide.

All data will be kept for three years after the researcher graduates. This is required by the IRB. You can receive copies of the results if requested at a future date. After three years have passed, all hardcopies will be shredded. Electronic information will be deleted from the computer. When reporting data from the study, the names of individual students will not be used. Individual results will be reported in combination with data retrieved from other participants.

Voluntary Participation: Your decision to allow your child to participate in this research project is entirely voluntary. If you agree now to allow your child to participate and you change your mind later, you are free to withdraw your child from the study at that time. Even if you give your permission and want your child to be part of the study, your child may decide not to participate at all, or he/she may leave the study at any time.

By not allowing your child to participate in this study or by withdrawing him/her from the study before the research is complete, you are not giving up any rights that you or your child have or any services to which you or your child are otherwise entitled to from Valdosta State University. Likewise, if your child decides on his/her own not to participate or to drop out of the study later

on, he/she is not giving up any rights, including rights to services from Valdosta State University to which he/she is otherwise entitled. During the focus group, your child may skip any questions that he/she does not want to answer. If you decide to withdraw your child during or after data collection stages, his/her information will be deleted and will not be included in the research results.

Information Contacts: Questions regarding the purpose or procedures of the research should be directed to Kelly Dreger at (229)-436-6425 or kelly6754@yahoo.com. Also, you may contact the supervising faculty, Dr. Steve Downey, at sedowney@valdosta.edu for additional information. This study has been approved by the Valdosta State University Institutional Review Board (IRB) for the Protection of Human Research Participants. The IRB, a university committee established by Federal law, is responsible for protecting the rights and welfare of research participants. If you have concerns or questions about your child's rights as a research participant, you may contact the IRB Administrator at (229)-333-7837 or irb@valdosta.edu.

Agreement to Participate: The research project and my child's (or ward's) role in it have been explained to me, and my questions have been answered to my satisfaction. I grant permission for my child to participate in this study. By signing this form, I am indicating that I am either the custodial parent or legal guardian of the child. I have received a copy of this permission form.

I would like to receive a copy of the focus group transcript: _____ Yes _____ No

I would like to receive a copy of the results of this study: _____ Yes _____ No

Mailing Address: _____

E-mail Address: _____

_____
Printed Name of Child/Ward

_____
Printed Name of Parent/Guardian

_____
Signature of Parent/Guardian Date

_____
Signature of Person Obtaining Consent Date

This research project has been approved by the Valdosta State University Institutional Review Board for the Protection of Human Research Participants through the date noted below:

APPENDIX G:

Child Assent Script

Child Assent Script

Hi. How are you? My name is Kelly Dreger. I'm a graduate student from Valdosta State University. Right now, I'm trying to learn about the use of points and coins in the classroom. I would like for you to help me with my research project, but let me explain what will happen if you decide to help me.

I will ask you to keep track of any points or coins you receive from your teacher during the next grading period. This is around nine weeks of school. You will be able to use them to get prizes for your behavior. Not everyone will receive the same things. Some of you may not receive any points or coins. If you have any questions, you can ask them to me or your teacher.

Later on, you might be interviewed so that you can share your feelings and opinions about what you receive. The ones who are picked will be interviewed in groups. I will record them and create written notes of the interviews. There are no right or wrong answers to the questions. There may be questions that you don't like answering. If so, you can choose not to answer the ones that make you uncomfortable. By being in the study, you will help me understand if the use of points and coins helps with your academic performance and your classroom behavior. You will learn about what interests you, and you will write down your preferences at the end of the study. Your opinions would help your teacher create activities that include what you like.

Your classmates would know what you receive in class, but they will not know what you have said during the interviews. Your teacher and your parents will know about what you receive, and they will keep track of them along with you. They will also know about what you say in the interviews. When I share the results of my study at the end of the project, I will not use your name. You will have an ID number on your documents instead, so no one will be able to tell who I'm talking about.

Your parents says it's okay for you to be in my study, but if you don't want to participate in the study, you don't have to do so. What you decide won't make any difference with your grades. I won't be upset, and no one will hold it against you if you don't want to be in the study. If you want to be in the study now but change your mind later, that's okay. You can stop at any time. If there is anything you don't understand, you should ask me about it so I can explain it to you.

You can ask me questions about the study at this time. If you have a question later that you don't think of now, you can ask your parents and teachers to call me or send me an email.

Do you have any questions for me now? Are there any comments you would like to make?

Would you like to be in my study to find out what you can receive for your behavior?

NOTES TO RESEARCHER: The child should answer "Yes" or "No." Only a definite "Yes" may be taken as assent to participate.

Name of Child: _____ **Parental Permission on File:** ☐ Yes ☐ No

(If "No," do not proceed with assent or research procedures.)

Child's Voluntary Response to Participation: ☐ Yes ☐ No

Signature of Researcher: _____ **Date:** -

_____

(Optional) Signature of Child: _____

APPENDIX H:

Interview Guides

Teacher Interview Guide

Note: This interview guide will be used with teachers for a one-hour individual interview. Each item has a question along with prompts. Prompts could be added or removed as the interview process occurs.

Part I - Introductory Experiences with Tokens

1. If applicable, describe one or more past experiences with tokens as a student.

   - Elementary
   - Middle
   - High School
   - College

2. If applicable, describe a few of your past experiences with tokens as a teacher.

   - Past Job Positions (if applicable)
   - Current Job Position (before study)
   - Students' Reactions
   - Personal Reactions

Part II – Procedures with Reinforcement System

1. Describe how you use points and coins for the study.

   - Awards for Appropriate Behaviors and Goals
   - Activities where tokens are used

2. How often do students receive tokens and prizes?

   - Per Day

- Per Week

3. How do students react to the use of points, coins, and prizes?

- General Impressions
- Examples
- What Works
- What Doesn't Work
- Changes over time based on Student Feedback

4. Explain what happens if students misbehave during class.

- Behaviors that lead to Referrals
- Strategies for dealing with Inappropriate Behaviors

Part III – Meaning Construction

1. Earlier we talked about how you use tokens in the classroom; how could tokens be used to support PBIS?

- In terms of support (performance and behavior)
- In terms of Tiers

2. What is your opinion about using tokens as an instructional strategy?

- Opinion on Points Systems
- Opinion about Coins Systems
- Reasons for Opinions
- Examples
- Suggestions

Student Focus Group Guide

Note: This interview guide will be used with students for a one-hour focus group interview. Each item has a question along with prompts. Prompts could be added or removed as the interview process occurs. I will assign an order for answering questions. Each student will have a chance to respond to the questions.

Part I - Introductory Experiences with Tokens

1. What are you learning in math class right now?

- Overview of Concepts
- Personal Goals concerning math

2. Describe a past school experience with points or coins if you have had one.

- For math
- For all subjects
- If not, ask if they witnessed the use of them in class.

3. How did you feel about what you experienced?

- Personal Reactions
- Recollection of Other's Reactions
- Reasons for Reactions

Part II – Procedures with Reinforcement System

1. Explain how you receive points, coins, and prizes this year for math.

- General Steps (in their own words)
- Reasons for tokens/prizes
- Activities where tokens are used

2. How often do you receive them?

- Per Day
- Per Week

3. What is your reaction when the teacher says you will be getting points, coins, and prizes during class?

- General Impressions
- Specific Feelings when receiving them
- Feelings about Other Students who receive them
- Reasons for Feelings (What Works/What Doesn't Work)

4. Explain what happens if you misbehave during class.

- Behaviors that Receive Warnings
- Behaviors that Lead to Referrals

Part III – Meaning Construction

1. What is your opinion about receiving points and coins in class?

- Opinion on Points Systems
- Opinion about Coins Systems

2. What is your opinion about the prizes you receive?

- Relevance of Prizes
- Suggested Changes

3. If you could change anything about what is done in math class, what would it be and why?

- Class Activities
- Learning Preferences

APPENDIX I:

Prize List Form

Prize List Form

(List Prizes You Can Afford. Have Students Vote for the Top Five.)

What You Can Supply (Examples: Free Time, Supplies, Candy)	**What Students Chose (From Most Votes to Least)**
1. _____	1. _____ = 20 Tokens
2. _____	2. _____ = 16 Tokens
3. _____	3. _____ = 12 Tokens
4. _____	4. _____ = 8 Tokens
5. _____	5. _____ = 4 Tokens
6. _____	
7. _____	
8. _____	
9. _____	
10._____	

Note: Coin Tokens are Fake Gold Coins

(Coins Supplied by Researcher)

Point Tokens are Marked Points on a Sheet of Paper

APPENDIX J:

Institutional Review Board Approval Form

Institutional Review Board (IRB)
for the Protection of Human Research Participants

NEW PROTOCOL APPROVAL

PROTOCOL NUMBER: IRB-03305-2015 **RESPONSIBLE RESEARCHER:** Kelly C. Dreger

PROJECT TITLE: Quasi-Experimental Study of Middle School Tokens, Behavior, Goals, Preferences, and Academic Achievement

APPROVAL DATE: 12/09/15 **EXPIRATION DATE:** 12/08/16

LEVEL OF RISK: X Minimal ☐ More than Minimal

TYPE OF REVIEW: X Expedited Under Category/ies 6&7 ☐ Convened (Full Board)

CONSENT REQUIREMENTS:

- X Adult Participants – Written informed consent with documentation (signature)
- ☐ Adult Participants – Written informed consent with waiver of documentation (signature)
- ☐ Adult Participants – Verbal informed consent
- ☐ Adult Participants – Waiver of informed consent
- X Minor Participants – Written parent/guardian permission with documentation (signature)
- ☐ Minor Participants – Written parent/guardian permission with waiver of documentation (signature)
- ☐ Minor Participants – Verbal parent/guardian permission
- ☐ Minor Participants – Waiver of parent/guardian permission
- X Minor Participants – Written assent with documentation (signature)
- ☐ Minor Participants – Written assent with waiver of documentation (signature)
- ☐ Minor Participants – Verbal assent
- ☐ Minor Participants – Waiver of assent
- ☐ Waiver of some elements of consent/permission/assent

APPROVAL: This research protocol is **approved** as presented. If applicable, your approved consent form(s), bearing the IRB approval stamp and protocol expiration date, will be mailed to you via campus mail or U.S. Postal Service unless you have made other arrangements with the IRB Administrator. Please use the stamped consent document(s) as your copy master(s). Once you duplicate the consent form(s), you may begin participant recruitment. **Please see Attachment 1 for additional important information for researchers.**

COMMENTS:

Lorraine Schmertzing 12/9/15

Thank you for submitting an IRB application.

Lorraine Schmertzing, Ed.D., IRB Chair Date

Please direct questions to irb@valdosta.edu or 229-259-5045.

Form Revised: 12.13.12