

EFFECT SIZE AND MODERATORS OF EFFECTS FOR
TOKEN ECONOMY INTERVENTIONS

A Dissertation

by

DENISE A. SOARES

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2011

Major Subject: Educational Psychology

Effect Size and Moderators of Effects for Token Economy Interventions

Copyright 2011 Denise A. Soares

EFFECT SIZE AND MODERATORS OF EFFECTS FOR TOKEN ECONOMY
INTERVENTIONS

A Dissertation

by

DENISE A. SOARES

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Kimberly J. Vannest
Committee Members,	Richard I. Parker
	Constance J. Fournier
	Steve E. Riechman
Head of Department,	Victor L. Willson

December 2011

Major Subject: Educational Psychology

ABSTRACT

Effect Size and Moderators of Effects for Token Economy Interventions.

(December 2011)

Denise A. Soares, B.S., Southwest Texas State University;

M.Ed., University of Houston

Chair of Advisory Committee: Dr. Kimberly J. Vannest

There is a clear call to use evidence-based practice (EBP) in schools, and a growing knowledge base of practices that have proven to be effective in helping students achieve in educational settings. In addition, the current trends of Positive Behavior Supports (PBS) and Response to Intervention (RtI) advocate for preventative and proactive strategies. Token economies (TE) are one intervention that is proactive and can be flexible to use with students across a wide range of behaviors and settings. According to Higgins, Williams, and McLaughlin, token economy (TE) is an effective way to improve classroom behavior. Unfortunately, limited recent research is available that evaluated the effects and moderators of token economies in classroom settings. The purpose of this investigation was to Meta-analyze the single case research on TE implemented in school and is the first to offer effect size analysis and identify moderators.

The use of TE's has been widely established as an evidence-based intervention for use in prisons, psychiatric hospitals, and school settings. However, very few articles

discuss size of effects to expect, the essential elements required, or the practical implementation issues within a classroom. Many myths surround the use of a TE, i.e., many assume a token system is effective only for individuals and this is not so, as TE is effective for groups as well as individuals. In an age of accountability and emphasis on preventative evidence based practice evidence for using a TE and how to implement a TE is needed in our literature. Empirical evidence for the use of a token economy in a classroom is presented along with suggested implementation ideas.

Twenty four studies were included in this Meta-analysis with an overall combined Tau-U ES of .78 of data showing improvement between phase A and B with CI_{90} [.72, .83]. Tau-U effect sizes ranged from .35 to 1.0. TE is effective with all ages evaluated (ages 3 – 15); however, statistically significant results indicated it was more effective with ages 6 - 15. Active ingredients (i.e. procedural steps) were evaluated, combined, and reported. Results indicate that TE is an evidence-based intervention to increase academic readiness behaviors and to decrease inappropriate behaviors.

DEDICATION

This is dedicated to my family. My husband, Barry, who made all this possible. This degree and dissertation could never have been completed without your inspiration, support, and encouragement. He knew exactly when to give me a push and when to give me a boost. I am so lucky to have you by my side, I love you.

My parents, Ron & Lynne, who always taught me I could do whatever I put my mind to. My sister, Barbara, just for keeping it all in perspective.

Jamie, Jennifer, Justin, Megan, & Sydney - may you also be motivated and encouraged to reach your dreams.

Judy, for her support, calming voice, and interpretation skills. I would not have seen the end without her. Lauren, for being my constant cheerleader. I also dedicate this dissertation to my many friends who have supported me throughout the process. I will always appreciate all they have done.

ACKNOWLEDGEMENTS

Many people have influenced my learning. I have been called a perpetual student, which is more than likely an accurate description of who I am. My educational journey is unending, so while I am appreciative of specific people with regard to my doctoral endeavors, it would be impossible to name all who have influenced me to travel the path to where I am today; needless to say, I would not be here without you.

A sincere appreciation and special thanks to my dissertation chair, Dr. Kimberly Vannest, whose continued support and encouragement were vital to the completion of this study. Dr. Vannest, thank you for being a step ahead of me and keeping me on track. I also thank you for your patience and tremendous guidance in shaping and carrying out my study, for always encouraging me to continue down this path. I have thoroughly enjoyed working with you. To my dissertation committee, Dr. Parker, Dr. Fournier, Dr. Riechman, I appreciate your insight into my study; your positive approach has encouraged me to be the best researcher I can be.

To my true friends for being so supportive, indulging me when I needed to vent and being excited with me when I learned something new, I thank you for being witnesses to this part of my life. A lifetime of thanks I give to my family members who have stood by me and encouraged me to move forward. Eternal thanks to my husband, Barry, who endured my many attitudes (positive and negative) during the process, I could not have succeeded without you!

NOMENCLATURE

APA	American Psychological Association
CI	Confidence Interval
ES	Effect Size
EBI	Evidence-based Intervention
EBP	Evidence-based Practice
IDEIA	Individual with Disabilities Education Improvement Act
MBD	Multiple Baseline Design
NCLB	No Child Left Behind
PABAK-OS	Prevalence and Bias Adjusted Kappa for Ordinal Scales
PBS	Positive Behavior Support
SED	Serious Emotional Disturbance
SCR	Single Case Research
TE	Token Economy

TABLE OF CONTENTS

	Page
ABSTRACT	iii
DEDICATION.....	v
ACKNOWLEDGEMENTS	vi
NOMENCLATURE	vii
TABLE OF CONTENTS.....	viii
LIST OF FIGURES	x
LIST OF TABLES.....	xi
CHAPTER	
I INTRODUCTION: THE IMPORTANCE OF RESEARCH.....	1
Introduction	1
Purpose Statement	5
Research Questions.....	6
Definitions and Key Terms	6
Organization of Study	8
II MANUSCRIPT I: THE EFFECTIVENESS OF TOKEN ECONOMY.....	10
Purpose and Research Questions.....	15
Method	15
Data Collection.....	16
Results.....	24
Discussion	29
Limitations	32
Implications.....	33
Conclusion.....	34

CHAPTER		Page
III	MANUSCRIPT II: MODERATED EFFECTS FOR TOKEN ECONOMY INTERVENTIONS FROM SINGLE CASE RESEARCH DESIGNS.....	36
	Potential Moderators.....	37
	Purpose and Research Questions.....	42
	Method.....	43
	Results.....	49
	Discussion.....	56
	Limitations.....	60
	Implications.....	61
IV	MANUSCRIPT III: TOKEN ECONOMY: MYTHS, APPLICATIONS, AND SUMMARY.....	63
	Case Study.....	64
	Introduction.....	65
	Evidence for Effectiveness.....	68
	Myths.....	69
	Pitfalls.....	74
	Summary.....	76
V	FINAL THOUGHTS.....	77
	REFERENCES.....	78
	APPENDIX A.....	105
	APPENDIX B.....	110
	APPENDIX C.....	127
	APPENDIX D.....	134
	VITA.....	137

LIST OF FIGURES

FIGURE	Page
A-1 Forrest Plot of Effect Size for 24 Included Studies and the Overall ES.....	105
A-2 Example of You're Buzz'n-Gram: A Token Earning Card.....	106
A-3 Visual Chart of Expected Behavior.....	107
A-4 Bella's Notebook Visual Chart	108
A-5 Simple Steps for a Token Economy	109

LIST OF TABLES

TABLE	Page
B-1 Summary of 29 Previously Reviewed Articles Included in Two Review Studies	110
B-2 Protocol for Design Strength	113
B-3 Summary of Participant and Setting Characteristics.....	114
B-4 Cross Tabulation Report from NCSS.....	115
B-5 Quality Review of TE Studies – Internal Validity Only	116
B-6 Included Studies and Effect Sizes in Alpha Order by 1 st Author.....	118
B-7 Summary of Included Studies and Effect Sizes	119
B-8 Inter-rater Reliability for Coded Categories	120
B-9 Summary of Participant Characteristics	120
B-10 Summary of Study Characteristics.....	121
B-11 Active Ingredients (Steps) of Token Economy Systems.....	122
B-12 Combination of Active Ingredients (Steps) Comparison for TE	123
B-13 Summary Statistics of Moderators for a TE	125
B-14 No-cost Backup Reinforcers	126

CHAPTER I

INTRODUCTION: THE IMPORTANCE OF RESEARCH

Introduction

Teachers need simple, efficient, effective interventions. Twenty percent of children demonstrate behavior problems in schools that do not respond to typical classroom management strategies; this challenges teachers to use more intense classroom management techniques (Sugai & Horner, 2002). However, school personnel struggle to find feasible (i.e. simple and easy to implement) classroom management techniques.

Behavior problems demonstrated by one or more student can interfere with teaching and learning of all students in the class (Carr, Taylor, & Robinson, 1991). For the student demonstrating the inappropriate behavior, instructional time can be lost; both in class through off task behavior and by being removed from class for disciplinary action, (Gest & Gest, 2005). This loss of academic instruction can be the beginning of an escalating cycle of failure (Scott, Nelson, & Liaupsin, 2001) with increased student frustration, the potential for further behavior problems, and more exclusion from instruction. Johns (2000) found that students with challenging behaviors and learning difficulties also receive less academic engaged time with their teachers than students without challenging behaviors. In addition, other students have difficulty focusing on academic instruction both when the behavior is

This dissertation follows the style of *Exceptional Children*.

demonstrated and while the teacher responds to the behavior. Teacher response also interrupts the pace of instruction (Johns, 2000). To prevent this negative cycle, teachers need knowledge of effective strategies to intervene proactively.

However, empirically supported strategies are insufficient. Strategies must be feasible for teachers to implement. One factor of feasibility is the amount of time required to plan and implement the intervention (Chafouleas, Riley-Tillman, & Christ, 2009). Teachers may be reluctant to implement interventions, with as low as 50% of the school day being spent on instructional activities (Good, 1983; Thurlow, Ysseldyke, Graden, & Algozzine, 1983). Therefore, complex interventions with multiple active ingredients (i.e. procedural steps) might not be implemented with fidelity and this potentially impact effectiveness. Token economy (TE) is one such intervention.

Token Economy

The use of a token economy is one behavioral intervention in which the individual can learn specific skills to obtain rewards and satisfaction contingent on the display of the appropriate behavior in a given setting (Cooper, Heron, & Heward, 2007). A token economy system is a type of intervention or behavior management system that uses a token (i.e., stickers, coupons, colored chips, etc.) to set the occasion and subsequently increase the probability of a behavior change (Kerr & Nelson, 1998). A token economy is any type of "structured treatment in which desirable behaviors are rewarded with tokens which are exchangeable for valuable goods or activities" (Lecomte, Liberman, & Wallace, 2000, p. 1312).

Theoretical Model

Token Economy is based on the principal of operant conditioning. Operant conditioning involves developing association between various consequences and behaviors in order to achieve the desired outcome (Zirpoli, 2005). Skinner used positive reinforcers to strengthen behaviors and used punishment to discourage behaviors that were not desired.

B.F. Skinner, a behaviorist, coined the term operant to refer to “active behavior that operates upon the environment to generate consequences” (Skinner, 1953). This theory explained how a range of learned behaviors are acquired and displayed. Operant conditioning has been linked to token economies when used in educational settings, research, and practice.

Token Economy Effective Settings

Token economies (TE) have been studied in a variety of settings with diverse subjects and behaviors (see Kazdin, & Bootzin, 1972; and Kazdin, 1982, for a review). TE exhibit success not only in schools (Akin-Little & Little, 2004; Cavalier, Ferretti, & Hodges, 1997; Christensen, Young, & Marchant, 2004; Filcheck, McNeil, Greco, & Bernard, 2004) but also in residential treatment centers (Barkley, Hastings, Tousel & Tousel, 1976), mental health hospitals (Ayllon, & Azrin, 1965; Berryman, O’Brien, & Cummins, 1983; Cotler, Applegate, King, & Kristal, 1972; Hundert & Batstone, 1978; Kaufman & O’Leary, 1972; Mayhew & Anderson, 1980), prison detention centers (Bassett, & Blanchard, 1977; Bippes, McLaughlin, & Williams, 1986; Holt, Hobbs, & Hankins, 1976), after school

programs (Turkewitz, O'Leary, & Ironsmith, 1975), colleges (Stilitz, 2009), and church schools (Swiezy, Matson, & Box, 1992).

Token Economy Effective Populations

Studies also indicate successful application of TE across population types. TE show positive effects for students with emotional and behavioral problems (Center & Wascom, 1984; Gaughan & Axelrod, 1989), intellectual disabilities (Baine, 1973; Carey, Mosk, & Kranchuck, 1981; Cotler et al., 1972; Forness & MacMillan, 1972; Kazdin & Geesey, 1980; Kazdin & Mascitelli, 1980), Attention Deficit Hyperactivity Disorder (Carlson et al., 1992; DuPaul & Weyandt, 2006; Johnson, Handen, Lubetsky, & Sacco, 1994), learning disabilities (Cavalier et al., 1997; Higgins, Williams & McLaughlin, 2001) and schizophrenia (Ulmer, 1976).

Token Economy Shows Promise

Research verifies the effectiveness of TE and the wide spread use, it continues to be one of the more effective forms of behavior modification (Matson & Boisjoli, 2009). The TE is robust and adaptable for treating a range of behaviors in various settings (Matson & Boisjoli, 2009). Despite a solid literature base demonstrating the effectiveness of this intervention strategy, the inherent complexity of the intervention needs to be simplified to improve fidelity and establish it as an evidence-base intervention for teachers.

Multiple individual research studies demonstrate effects of TE in multiple settings with differing populations; however, no meta-analysis has been found in the literature. This means no effect size evaluation of moderators in TE is available in

the current research, no effect size calculations are reported, no meta-analysis has been conducted, and no evaluation of moderators is available. Reviews have been done that attempt to summarize results across studies, but they struggle to collapse findings since the studies vary dramatically in procedures or components. Studies report a range “active ingredients” [four (Cooper, Heron, & Heward, 1987; Miltenberger, 2000; O’Leary & O’Leary, 1977) to seven (Miltenberger, 2000)]. For instance, differing combinations of the following active ingredients are included in the TE studies: (a) defining the target behavior, (b) type of token, (c) reinforcement schedule, (d) reinforcement menu, (e) reinforcement survey, (f) exchange rate, and (g) response cost. The essential or required active ingredients (or combination of active ingredients) for effects have yet to be identified empirically in the literature with one exception Vannest, Reynolds, & Kamphaus (2008) identify 5 basic elements and 12 procedural steps; however, this review is not an empirical test. Clearly more work is needed in this area.

Purpose Statement

The purpose of this study is four-fold: (1) to identify overall effects; (2) to determine range of effects (i.e., confidence intervals), (3) the determine moderators of effects, and (4) to determine the number of active ingredients (i.e. procedural steps) that are needed to maintain effectiveness. Moderators to be evaluated include within subject moderators (i.e. age, outcome variable of academic or behavioral), setting moderators (i.e., location of the intervention), and there are procedural moderators (i.e., inclusion of response cost, or verbal reminder).

Research Questions

This research sets out to answer the following questions:

- (1) What is TE's overall effectiveness and range of effectiveness as an intervention across studies (measured ES)?
- (2) With what confidence do we know this (confidence interval)?
- (3) Is TE different across age groups?
- (4) Is TE more effective in special education settings or general education settings?
- (5) Does the implementation of a response cost enhance the effectiveness of a TE?
- (6) Does a verbal reminder of token earning during the intervention phase enhance the effectiveness of TE?
- (7) What are the most influential active ingredients?
- (8) Is TE differentially effective for academic versus behavioral?

Definitions and Key Terms

The following definitions are provided to ensure consistency and understanding of the terms used throughout the study.

1. At-risk Students – Is defined as most frequently manifested by poor academic and social skills that promote a general disconnection with the school culture (McDonald, 2002).
2. Evidence Based Research - The No Child Left Behind Act (2001) requires the application of evidence-based research to educational practice and defines

it as "research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs."

3. Behavior Management – Behavior management is defined as all those actions, teachers and parents engage in to enhance the probability that children develop effective behaviors that are personally self-fulfilling, productive, and socially acceptable
4. Token Economy - A token economy is based on the premise of Operant Learning Theory (Skinner, 1931) in which the use of a behavior's antecedent and/or its consequence influences the occurrence and form of behavior. Kazdin (1977) defined token economy, as a behavioral technique in which the desired behavior changes achieved by delivering tokens for the performance of a desired behavior.
5. Token - Tokens are a secondary reinforcer and not worth anything themselves but are exchanged for something of value called a back-up reinforcer (Alberto & Troutman, 2003; Martin, & Pear, 2003).
6. Back-up reinforcers – Tokens themselves have no intrinsic value but can be exchanged for other reinforcers called back-up reinforcers. They can be material or privileges but are chosen for the appeal of the individual or group of students.
7. Effect Size - Effect sizes tell us about the strength of an intervention, the size of the effect.

8. Confidence Interval - The confidence interval can tell us about the degree of trustworthiness of our findings and the degree of error that might occur.
9. Single Case Research - SCR designs have distinctive abilities to (a) permit reliable changes in the dependent variable to be detected and (b) permit valid inferences to be concluded given observed changes (Kazdin, 1982; Kratochwill, 1992).
10. Active Ingredients – Active ingredients are the procedural steps to the token economy system and will be coded as moderators. A moderator variable is one that influences or alters the relationship of other variables (Holmbeck, 1997).

Organization of Study

This chapter introduces and discusses Token Economy as an evidence-based intervention. The rationale needed to identify active ingredients of a token economy system to improve fidelity of implementation in a classroom setting is discussed. In conclusion, this section examined the experimental research questions specific to this study and defined key terms.

Chapter II is the first manuscript of this two-manuscript dissertation was a meta-analysis to evaluate the overall ES and CI of TE. Twenty four studies were included with an overall combined Tau-U ES of .78 of data showing improvement between phase A and B with CI_{90} [.72, .83]. Tau-U effect sizes ranged from .35 to 1.0 with a standard error range of .09 to .31. Results indicate that TE is an evidence-

based intervention to increase academic readiness behaviors and to decrease inappropriate behaviors.

Chapter III is the second manuscript of this two-manuscript dissertation was a meta-analysis to evaluate the moderator effects of TE. Twenty-four studies were included. Age is the only moderator that produced statistically significant results. Active ingredients (i.e. procedural steps) were evaluated, combined, and reported. Results indicate as few as two steps are needed for effectiveness of a TE.

Chapter IV is written with teachers in mind. It is written in a prose for practitioner format, targeting the widely circulated teacher journal “Teaching Exceptional Children” this journal requires no more than one reference per paragraph and requests call outs and other stylistically different formats. As you read you may notice these differences, they are intentional. The author’s research focus is pre-service teacher education, teacher development, and research with practical implementation.

CHAPTER II

MANUSCRIPT I: THE EFFECTIVENESS OF TOKEN ECONOMY

A token economy (TE) is a commonly used, highly acceptable intervention appropriate for a range of educational settings (e.g. self-contained, content mastery, resource rooms, general education), fitting easily within daily routines of teachers and students. Derived from operant learning theory (Skinner, 1931), TE is a secondary reinforcement system (Alberto & Troutman, 2003) in which an individual earns tokens for targeted behaviors and exchanges accumulated tokens for a “larger” reinforcer (Kazdin, 1971).

Token economies (TE) have been studied in a variety of settings with diverse subjects and behaviors (see Kazdin, & Bootzin, 1972; and Kazdin, 1982, for a review). TEs exhibit success not only in schools (Akin-Little & Little, 2004; Cavalier, Ferretti, & Hodges, 1997; Christensen, Young, & Marchant, 2004; Filcheck, McNeil, Greco, & Bernard, 2004) but also in residential treatment centers (Barkley, Hastings, Tousel, & Tousel, 1976), mental health hospitals (Ayllon, & Azrin, 1965; Berryman, O’Brien, & Cummins, 1983; Cotler, Applegate, King, & Kristal, 1972; Hundert & Batstone, 1978; Kaufman & O’Leary, 1972; Mayhew & Anderson, 1980; Ulmer, 1976), prison detention centers (Bassett, & Blanchard, 1977; Bippes, McLaughlin, & Williams, 1986; Holt, Hobbs, & Hankins, 1976), after school programs (Turkewitz, O’Leary, & Ironsmith, 1975), colleges (Stilitz, 2009), and church schools (Swiezy, Matson, & Box, 1992).

Four prior reviews evaluated TE. Two are evaluative reviews that did not pose any research question (Kazdin, 1972; Kazdin & Bootzin, 1982). As stated in Weiss, evaluative reviews are an assessment of a process as a means of contributing to the improvement of the process (Weiss, 1998). Two are systematic reviews that answered specific research questions about the use of TE (Dickerson, Tenhual, & Green-Paden, 2005; Matson & Boisjoli, 2009). The Cochran Collaboration states a systematic review is focused on a specific research question and summarizes literature relevant to the research question (Higgins & Green, 2011). The

The two evaluative reviews (Kazdin & Bootzin, 1972; Kazdin, 1982) discuss three things: (a) both reviews discussed strengths and weakness of TE; (b) second, the advantages and obstacles that impede implementation of TE; and (c) third, concerns of generalization procedures. However, the evaluative reviews do not provide data or analysis from source articles. Kazdin (1982) found that TE are effective in restrictive environments but continues to raise concerns of obstacles that may impede generalization to an educational setting.

The other two reviews (Dickerson, Tenhual, & Green-Paden, 2005; Matson & Boisjoli, 2009) are systematic literature reviews that answered specific research questions (Higgins & Green, 2011). Dickerson et al. (2005) sought to update the schizophrenia treatment recommendations using TE to improve socially appropriate behaviors in a hospital setting. Dickerson et al, (2005) reviewed 13 controlled studies that focused on using TE specifically in a hospital setting; however, there are clear limitations in generalizing to school age children in a school setting. Matson

and Boisjoli (2009) was more closely aligned to school research by reviewing 16 studies, which 87% of the intervention took place in a school. This systematic review sought to answer the effectiveness of TE as a therapeutic intervention that would mimic a naturally occurring reinforcement systems (e.g. money) with the developmentally disabled and autism population. (See Table B-1 for an inclusive summary).

Although there is agreement of the effectiveness of TE in published research, it is crucial to evaluate previous research on the intervention related variables that could generalize to schools. Dickerson, Tenhula, and Green-Paden (2005) targeted TE implemented in hospital settings with one thousand seventy four participants ranging from 18 to 55 year old, 29% of whom were diagnosed with schizophrenia, 13% psychotic disorder, and 57% with other mental illness. Dickerson et al., (2005) found that TE was effective in increasing adaptive behaviors such as work performance, social interactions, and daily living care skills of patients in hospital settings based on increased scores rating scales for socially appropriate behaviors; however no effect size (ES) or confidence intervals (CI) were reported.

Matson and Boisjoli (2009) reviewed 16 studies targeted TE in multiple settings such as, schools, home, summer camps, group homes, state hospitals, and a developmental center with one hundred sixty four participants ranging in age from 4 to 18 years old, approximately 91% were children with intellectual disabilities and 8% with autism. Matson and Boisjoli (2009) found that TE was associated with an

increase in social, behavioral, and academic outcomes; however, no effect size (ES), confidence intervals (CI), or quality of research were reported.

While these four previous reviews of literature (Dickerson, Tenhual, & Green-Paden, 2005; Kazdin, 1982; Kazdin & Bootzen, 1972; Matson & Boisjoli, 2009) found token economies were effective in reducing inappropriate behavior and increasing academic achievement, the following errors of omission were present. A majority of articles reviewed do not report the effectiveness with school age participants in an educational setting, overall effect size (ES), confidence intervals (CI), or the quality of research of a TE.

In the two systematic literature reviews, the authors found TE to be an effective intervention through review of 29 single case research (SCR) studies; however, the authors provided no quantitative information regarding the level of effectiveness. Much of the intervention research in schools utilizes single case design (Horner, Carr, Halle, Odom, & Wolery, 2005), a scientific methodology that may provide some justification for effective interventions by suggesting a causal relationship between an intervention and its effects (Buysse et al., 1995; Horner et al., 2005). Because of varying needs of the individual or group with similar characteristics, SCR has proven relevant for defining educational practices at the level of the individual learner (Horner, et al., 2005) and provides a level of rigor by demonstrating experimental control much like a randomized control-group (Shavelson & Towne, 2002). Quality SCR is a reliable method of contributing to the field of education because: (a) the intervention is operationally defined, (b) the

setting of the intervention is defined, (c) the practice is implemented with fidelity, (d) a functional relationship is established between the intervention and the results, and (e) the results are replicated across studies, researchers, and participants (Horner et al., 2005).

ES and CI are necessary to empirically establish the token economy's effectiveness. The American Psychological Association (APA) requires that effect sizes (ES) and confidence intervals (CI) present data in a way that is interpretable (Wilkinson & APA Task Force on Statistical Inference, 1999) in order to give researchers enough information to assess the magnitude of the observed effect or relationship (APA, 2001 p. 26). ES provide this information by assessing the difference between groups or the strength of a relationship between variables. The CI indicates the degree of finding's trustworthiness and the range likely to contain the true effect size (Parker & Vannest, 2009). However, SCR has typically not been accessible to statistical measures like ES and CI (Parker & Vannest, 2009; Parker, Vannest & Brown, 2009) calculated through meta-analytical analysis.

A meta-analysis allows us to aggregate the results of multiple studies in order to provide a bigger picture of the effects of an intervention. It is an accepted method of summarizing the results of empirical studies within the behavioral, social, and health sciences (Kavale, 2001; Lipsey & Wilson, 2001). This is important because TE shows promise across settings and participants but there are clearly some unanswered questions. Technology has not been available previously for us to examine it given most TE are SCR studies; however new techniques provide options

to analyze through a meta-analysis. Meta-analysis is needed, as a single intervention study is not sufficient to identify a practice as effective (Thompson, 2006). This design has many benefits such as; detects small or moderate relationships, obtains a more precise estimate of a relationship, guides future research and finds patterns across studies (Gall, Gall, & Borg, , 2006). Although several single-case studies have demonstrated TE's effectiveness with various populations, it is critical to evaluate that data using a common metric -- i.e., an effect size measure -- via a meta-analysis (Kavale, 1984, 1998, 2001). The previous systematic reviews of TE provided answers to specific research questions; however, neither were meta-analytical designs.

Purpose and Research Questions

The purpose of this meta-analysis was to investigate the effectiveness of a token economy system implemented with children in public schools by calculating an overall ES with CI's. The research questions were: (a) what does existing evidence suggest regarding the effects of token economy interventions? (Measured ES), and (b) with what confidence do we know this (confidence interval)?

Method

This meta-analysis was conducted using the five-stage model of the integrative literature review suggested by Cooper (1982). The five stages are: (a) Problem formulation (e.g. research questions), (b) Data collection (e.g. identify studies through searching and acquiring and inclusion criteria, (c) Data evaluation (e.g. coding of articles) , (d) Analysis & interpretation, and (e) Public presentation.

This section describes the meta-analytic methodology for assessing the effects of a Token Economy Intervention in a school setting. The parts described in this section will be (b) data collection, (c) data evaluation, and (d) analysis and interpretation. First, the data collection phase, which includes the literature search methods and inclusion criteria are explained. Next, the data evaluation, which includes the inclusion reliability -- setting, location, and inclusion of a single case graph, -- is described. The last phase to be described is the analysis and interpretation, which includes the data extraction process and the effect size calculation, followed by a discussion of the assessment of the included studies' methodological quality.

Data Collection

Search Methods

Standard methods identified by Cooper and Hedges (1994) were used to search PsycINFO and Education Resource Information Center (ERIC) electronic databases. Key words, Boolean strings and truncated words used to conduct the search including but not be limited to: (a) token economy (b) intervention, (c) reinforcement, (d) contingency management, (e) systematic positive reinforcement, (f) tokens, (g) operant conditioning, (h) applied behavior analysis, (i) back-up reinforcers, (j) behavior therapy, (k) points, and/or (l) response cost.

Inclusion Criteria

Studies were included in the analysis only if they met five criteria. These criteria were established to ensure the data would answer the research questions. First, studies must measure effects of a TE intervention, defined as a program in

which students earn tokens for desirable behaviors and then exchange the tokens earned for back-up reinforcers (Alberto & Troutman, 2003; Martin & Pear, 2003). Second, study venues included only U.S. classroom settings with school age children (age 3-21). Third, studies must have been published from 1980 to 2011, in order to create a manageable set of articles representing typical public school classrooms, which compare to a modern day classroom settings that include general and special education. Studies published subsequent to prior reviews (i.e. Kazdin, 1980) were included. Fourth, studies utilized SCR methodology with a clearly readable graph of data and group studies were excluded to allow for comparison of effect sizes (Lipsey & Wilson, 2001). SCR designs allow for visual inspection of the data, especially at the phase change from baseline to intervention (Kazdin, 1982). Visual analysis is important to inspect the trend, slope, and intercept gap (Suen & Ary, 1989). Additionally, the visual graph allowed for extraction of original raw data for new ES analysis (see Data Extraction). When studies included multi-component interventions, the interventions assessed were required to be graphed individually so that the TE intervention could be separated and analyzed. Last, studies must have been published in peer-reviewed journals to ensure authors meet the accepted standards for their field and to prevent dissemination of irrelevant findings, unwarranted claims, unacceptable interpretations, and personal views (Ludwick, Dieckman, Herdtner, Dugan, & Roche, 1998). Publications that had not undergone peer review such as dissertations, descriptive articles, and unpublished desk copies were excluded.

For this meta-analysis, an effort was made to include as much of the population of empirical research as possible given the criteria for inclusion outlined above. The previously listed search terms were combined in the search engine so that any study in which the title or abstract contained a key word would be included in the results. The titles and abstracts of the studies acquired through this process were examined and were retained for screening if they were TE studies appearing to focus on school age children.

In addition to the search of databases, a hand search for titles related to token economies or secondary reinforcement was completed by reviewing the table of contents in the major journals in special education, school psychology, and behavioral psychology (as determined by ISI rating for the past 2 years): e.g., *Journal of Positive Behavior Intervention*, *Behavior Therapy*, *Behavioral Interventions*, *Journal of Special Education*. These journals were selected based on their prominence in the field of publishing intervention related articles. This process did not yield any additional studies that were not already evaluated in the abstract and full review stage.

Reference pages of all the resulting screened articles were inspected for additional eligible studies. Full documents for the resulting studies were acquired, at which time they were examined relative to the inclusion criteria. Any study that did not clearly fail the basic inclusion criteria was retained for review. This strategy captured many articles written by several of the more prominent and well-known

authors and experts in this field; however, it did not yield any additional studies that were not found in the original database search.

Data Evaluation

Inter-rater Reliability of Literature Search. Reliability of the literature search was achieved through inter-rater reliability checks between the two doctoral students completing the search. The two reviewers calibrated the search by (a) agreeing to the exact key words to use (e.g. token economy, classroom, intervention), (b) agreeing on two journals relevant to the search (e.g. Journal of Positive Behavior Intervention and Journal of Special Education), and (c) using a specific search engine (e.g., psychinfo). Trial checks were performed to calibrate the two reviewers by evaluating the same articles located in a specific journal. Reliability checks occurred during the article gathering stage when the reviewers' reliability was compared based on the articles they selected for inclusion using the inclusion criteria. Official reliability of the articles located was assessed through simple Percent of Agreement ($\text{sum of agreement} / \text{total number of agreement} + \text{disagreements} \times 100$). Initial agreement was 100%.

Coding. Social science research involves capturing and analyzing data. In order to make sense of the data for better analysis, the researcher must code or "label" the data. A protocol was adapted from Tolan, Bass, Henry, & Schoeny (2008) in order to code many relevant study characteristics. The complete coding protocol for this research synthesis may be found in Appendix A. In addition to bibliographic information, coded study characteristics included extensive

information about setting of study, participants, intervention features, research design, methodology, and measures used.

Inter-rater Reliability of Coding. Reliability of data coding was ensured through inter-rater reliability checks between the two doctoral students doing the coding. Before calculating reliability, coding sheet training and a trial coding were performed to calibrate the two observers. The two raters identified each coding variable using one example and one non-example. Official coding began when a minimum acceptable value (range from .80 to .90) of inter-rater agreement was met (Hartmann, Barrois, & Wood, 2004).

Cohen's Kappa reliability agreement using NCSS (Hintze, 2004) was calculated by entering the agreement/disagreement matrix for analysis. For cross tabulation, matrix data was entered into the NCSS statistics program, which provides the Cohen's Kappa (Kappa) reliability index that adjusts for the expected chance agreement. Kappa is a conservative measure of reliability and perhaps even underestimates agreement (Ary & Suen, 1989; Strijbos, Martens, Prins, & Jochems, 2006).

Additionally, reliability was calculated using Prevalence and Bias Adjusted Kappa for Ordinal Scales (PABAK-OS). Rater scores were transferred into an agreement matrix that allowed for the calculation of inter-rater reliability with PABAK-OS which considers special attributes of ordinal data by assigning differential weights. Just as PABAK (Byrt, Bishop, & Carlin, 1993) corrects

Kappa's undesirable sensitivities with nominal scales, PABAK-OS does the same for ordinal scales (Parker, Vannest, & Davis, in press).

Analysis and Interpretation

Data Extraction. Data extraction is a process that allows a researcher to digitize the original visual graph with the use of software. Digitizing data results in the exact reconstruction of the original graphic data to numeric raw data to enable proper comparisons (Glass, 1976). All single subject graphs of included articles were digitized in Getdata Graph Digitizer (Version 2.21) from getdata.com.ru. If graphs were not available in the article or were illegible, the researcher attempted to contact the first author for the visual graph. If the graph was not located, the article was excluded.

The researcher followed a four-step process to digitize the data. First, PDF versions of included articles were created. Second, each SCR graph was scanned into Get Data Software. Third, the X and Y-axes were defined by setting the values, and each data point was converted from published graph to raw data. Last, the values were imported into an Excel[®] spreadsheet (data ordered as each column = new phase).

Effect Size. Effect sizes (ES) were used to interpret the outcomes of individual studies in relation to each other. By calculating effect size, a comparison was made of the magnitude of change from one study to another (Thalheimer & Cook, 2002) and may be interpreted in relation to each other. ES may also be combined to produce an overall estimate of the relationships among those same

variables across a field of study. Effect sizes were calculated on the initial baseline versus intervention contrast (A1 vs. B1) for each design, and their data were collapsed across individual students or behaviors within a single study to produce an ES (Parker & Hagan-Burke, 2007). The effect Size used was Tau-U. Tau-U is a non-overlap with trend, a relatively new effect size in education research (Parker, Vannest, Davis & Sauber, 2011). Overlapping data are defined as the fewest data-points, which would be removed from either phase to eliminate all data overlap between phases. Tau-U extends Tau non-overlap by controlling for monotonic (positive) baseline trend (Parker et al., in press)

Parker et al. (2010) summarize Tau-U as “having statistical power that is flexible and can calculate trend only, non-overlap between phases only, or a combination of the two.” (p. 3, in press). Following the steps outlined for calculating Tau-U (Parker et al., 2011), the result is a conservative measure that offers important benefits. Benefits of Tau-U’s nonparametric “bottom-up” approach include: (a) consistency with visual analysis; (b) applicability to short data series and simple designs; (c) appropriateness with any design; (d) characterization by strong statistical power, which is of the strongest parametric tests (91% to 115%); (e) it control in phase A trend and (f) usefulness at three levels: non-aggregated data from a single client, aggregated data from a complex design, and meta-analyses (Parker et al., 2011).

Tau-U effects can be combined over multiple phase contrasts to provide an effect size for the overall design and then can be further combined across designs

and studies. The data was analyzed using the Tau-U calculator (singlecaseresearch.org) for individual ES. The ES and standard error were entered into the statistical program WinPEPI (Abramson, 2010) for analysis. The algorithm for WinPEPI to calculate the overall ES is the weighted average of all individual ES, with weights equaling the inverse of the Variance (not the standard error). The software also provided confidence intervals. Interpretation categories for ES effectiveness in Tau-U have not been established; however, Tau-U is in the strength of association family and therefore effect size interpretations are recommended minimum effect size (.2), moderate effects (.5), and strong effects (.8) (Ferguson, 2009).

Assessment of Methodological Quality. After ES was calculated across all studies, the quality of each study was assessed for methodological quality. Evaluating methodological quality is an important variable when conducting a meta-analysis to calculate ES across studies. Including studies with low quality in the calculation can increase the potential for untrustworthy results. A protocol was developed to examine quality of a design (see Table B-2) as a measurement score calculated using a 1-3 (weak, medium, and strong) rubric with criteria based on recently published guidelines for evaluating the methodology of single case designs (Horner et al., 2005; Kratochwill et al., 2010). The following features were evaluated in order to determine point value assigned for quality: (a) the type of design and (b) number of phases.

Results

The results of this meta-analysis are presented in the same sequence as in the methods section using the five-stage model of the integrative literature review suggested by Cooper (1982). The phases described in this section will be (b) data collection, (c) data evaluation, and (d) analysis and interpretation. First, data collection, which includes information about the results of the literature search and article selection outcomes, will be reported. The second phase, data evaluation, which includes participant and setting information as well as inter-rater reliability and coding quality are reported. The last phase, analysis and interpretation, reports (a) design features, (b) effect size calculation and confidence intervals, and (c) the methodology quality results are reported.

Data Collection

Search Results. Article citations were initially generated by searches of electronic databases. These initial searches yielded a high number of results (n=1,011), which was expected due to the number of key word searches and the word combinations. Article titles indicating non-related articles determined omissions, resulting in 322 remaining studies for further review. After omissions based on titles, 322 abstracts were reviewed for inclusion criteria. Abstracts without enough information to determine inclusion/exclusion remained for full review. Articles were excluded when they did not meet the criteria based on abstract review.

Full articles were then screened for inclusion. Two hundred and seventy eight studies were eliminated because the study was not conducted in a school, was

exclusively a descriptive study, or was without peer-review. Two doctoral students to determine if they met each inclusion criteria independently evaluated the remaining 44 articles.

Article Selection Results. The full texts of forty-four primary studies were examined for potential inclusion in this meta-analysis. Twenty studies were excluded: four addressed multi-components interventions for which the intervention data could not be separated, three did not include a visual graph of data from which raw data could be digitized, one was set in a classroom within a residential treatment center, one focused on the implementation of the token economy by paraprofessionals and did not include intervention data for the children in the study, and one included the intervention in the baseline data. After excluding these studies, the literature search resulted in 24 SCR studies in which token economies were the intervention in a classroom setting with school age children.

Data Evaluation

Participants and Setting. The sample consisted of 24 included studies that involved 84 students and produced 79 individual effect sizes. Table B-3 summarizes some of the features of these investigations. Studies were sorted into two groups based on age, preschool children, and school age children. Thirty-five percent of the studies took place with preschool age children ranging from three to five year olds, and sixty-five percent of the studies took place with school age children ranging in age from 6 – 15 year old. Outcome measures included behavior (79%) and academic readiness behaviors (21%). Eighteen (72%) of the studies included participants with

a special education eligibility; one (4%) included participants that were deaf, two (8%) studies included students with an intellectual disability, three (12%) studies included participants that were learning disabled, four (16%) included students that were emotional/behavioral disorders, and eight (30%) were identified special education but did not state an eligibility category. Interventions occurred equally in general and special education classrooms, 12 in special education, and 12 studies took place in general education.

Inter-reliability of Coding Quality of Studies. A detailed description of coded variables is presented in the methods section and the instrument used is located in Appendix A. Reliability was calculated using Cohen's Kappa (Kappa) and PABAK-OS. NSCC was used to produce a Cross Tabulation Report matrix and the Kappa for reliability. Kappa was .935. The cross tabulation report (Table B-4) was input into the PABAK-OS calculator (www.singlecaseresearch.org). PABAK-OS results indicate 23 hits and 1 near miss resulted in .94 CI₉₀ [.82 to 1.10].

Analysis and Interpretation

Designs of Included Studies. Seven of the studies included were Multiple Baseline (MB) Designs with 2 – 4 phases, six of the seven were MB across participants with the remaining one across behaviors. One study (Kilmas & McLaughlin, 2007) was a changing criterion design across three phases and three behaviors. The remaining studies (n=16) were Reversal Designs (see Table B-5).

Overall Effect Size Calculations of Included Studies. Twenty-four studies included 79 individual effect sizes (ES). ES were calculated for individual

participants in each study and then combined into an individual ES for each study. These studies were from 21 separate first authors. Kazdin (1980a; 1980b) and Salend (1985; 1986; 1988) are the first authors on multiple studies included in the analysis. When the same author is primary author on different studies it is possible that the design of the studies are identical which could potentially skew the results for this analysis either in a positive or negative direction depending on the quality of the design. In order to generate the independent effect sizes used in the analysis, the following procedure was used. The graph was digitized by scanning and defining the x and y-axis in order to obtain raw data to calculate the Tau-U effect size. These effect sizes provided a measure of the differences between phase A (baseline) and phase B (intervention): 92% of the studies contributed multiple (2 to 12 per study) effect sizes, and 8% contributed only a single effect size. Effect sizes within all studies were independent and treated as such until combined using WinPEPI (Abramson, 2010) software.

The overall Tau-U effect size was .78 of data showing improvement between phase A and B and within phase B with CI_{90} [.72, .83] (see Figure A-1). Tau-U effect sizes ranged from .35 to 1.0 with a standard error range of .09 to .31. The 24 included studies and their associated effect sizes are listed in Table B-6. Using Ferguson's (2009) ES interpretation suggestions for strength of association ES, sixteen of the studies had a strong effect size of .80 or above, five studies had a moderate ES (.50 - .79), two studies had a recommended minimum effect (.20- .49).

Despite the variability in ES, it does appear that TE is an effective intervention in classrooms.

Methodological Quality. Each of the 24 studies was reviewed for internal validity by the rules listed in the methods section (See Table B-5). Twelve studies (Filcheck et al., 2004; Himle, Woods & Bunaciu, 2008; Kazdin & Geesey, 1980; Kazdin & Mascitelli, 1980; Klimas & McLaughlin, 2007; Rosenberg, 1986; Salend & Allen, 1985; Simon, Ayllon, & Milan, 1982; Smith & Fowler, 1984; Sran & Borrero, 2010; Stevens et al., 201; Sullivan & O’Leary, 1990) were of *low* quality, eight studies were of *medium* quality (Center & Wascom, 1984; Conyers et al., 2004; De Martini-Scully et al., 2000; Higgins et al., 2001; Reitman, 2004; Salend & Lamb, 1986; Salend, Tintle & Balber, 1988; Truchlicka et al., 1998) four of *high* quality (Maglio & McLaughlin, 1981; McGoey & DuPaul, 2000; Mottram et al., 2002; Musser et al., 2001). Therefore, 50% of the studies were of weak quality, and 50% were of medium to strong quality (see Table B-7). Studies rated weak quality had a combined overall Tau-U of .76 CI₉₀ [.68, .84]. The eight studies with a medium quality rating had a combined overall Tau-U of .74 CI₉₀ [.65, .83]. The four studies categorized as strong quality had a combined overall Tau-U of .91 CI₉₀ [.77, 1.00]. The twelve combined medium and strong quality studies rated had a combined Tau-U of .79 CI₉₀ [.72, .86].

If the 12 low quality studies were excluded from the analysis the overall Tau U increased to .79 CI₉₀ [.72, .86] from a combined (24 studies) overall Tau-U of .75

CI₉₀ [.70, .80]. Given this, .79 is a better estimate of overall effect size as the low quality of the twelve studies, render study results questionable.

Discussion

This review provides additional information to the field regarding the overall effectiveness of TE when implemented in school setting from 1980 to 2011. While prior reviews addressed findings of TE implemented in a variety of settings, we contribute the analysis, which provides an overall ES and CI in order to draw conclusions about the effectiveness of TE on school age children. With 24 single-subject studies producing 79 effect sizes we found that, TE positively affects behavior and academic readiness behaviors in both general and special education settings. Consistent with earlier studies (Kimlas & McLaughlin, 1981; Salend, Tintle, & Balber, 1988; Sran & Borrero, 2010; Stevens et al., 2011; Truchlicka, McLaughlin & Swain, 1988), the present study confirmed that a token system was effective in increasing academic readiness behaviors and decreasing inappropriate behaviors (Center & Wascom, 1984; Conyers et al., 2004; DeMartini Scully et al., 2000; Filcheck et al., 2004; Higgins, Williams, & McLaughlin, 2001; Himle et al., 2008; Kazdin & Geesey, 1980; Kazdin & Mascitelli, 1980; Maglilio & McLaughlin, 1981; McGoey & DuPaul, 2000; Mottram et al., 2002; Musser et al., 2001; Reitman, 2004; Rosenberg, 1986; Salend & Allen, 1985; Salend & Lamb, 1986; Simon, Ayllon, & Milan, 1982; Smith & Fowler, 1984; Sullivan & O'Leary, 1990).

Moderate overall effects (.79) and measurement qualities of TE's were found. This current investigation's findings suggest that, when offered in naturalistic

settings, token economy intervention can have a significant positive effect on decreasing inappropriate behaviors and/or increasing academic readiness behaviors. Collectively, these results build on positive results reported by other research teams that have conducted related reviews examining TE and their effectiveness (Kazdin & Bootzin, 1972; Kazdin, 1982; Dickerson, Tenhual, & Green-Paden, 2005; Matson & Boisjoli, 2009).

Previous studies and reviews have focused on a wide range of settings (including hospital, prison detention centers, residential treatment centers, mental health facilities and after school programs), but those focusing exclusively on interventions implemented in schools are less common. A major issue in the literature concerns the extent to which generalization effects are maintained to other settings (Dickerson et al., 2004; Kazdin & Bootzin, 1972). Although conclusions can be drawn about the efficacy of specific procedures, generalization to the broader population of persons with behavioral problems and maintenance of beneficial effects over time is difficult to assess. These issues could be resolved based on design; however, many studies included here did not control for generalization or follow up. Some argue the restrictiveness of a hospital setting allows for more control; however, the basic premise of a token economy is to give control over the behaviors to the students to allow for their own choices. The appropriate choices are then rewarded with tokens. The TE promotes an increase in academic readiness behaviors and a reduction of inappropriate behaviors so it should be primarily evaluated on this basis. Previous studies in restrictive settings such as hospitals or

prisons to a school setting with less control makes the relevance of prior studies questionable for educators.

A noteworthy point is all included studies used token economies as an intervention; however, they inevitably varied in duration, consistency, and intensity. Duration and intensity were not evaluated independently. Duration of the intervention lasted anywhere from four (in a MBD) to 32 days. It is essential to ensure consistency among staff in carrying out a TE program and in administering reinforcements to students that positive, immediate, and specific. These standards are difficult to meet. There is no convincing evidence that children receiving the intervention in a more intensive setting (e.g. self-contained special education) benefit more than those in less intensive programs such as a general education setting. Since intensity may represent a variety of intervention differences, the construct of intensity should be better defined to avoid confusion by other intervention variables.

Token methods have proven to be flexible to the extent that they can be applied to include individual children or entire classrooms (Filcheck, McNeil, Grecos & Bernard, 2004). While we can evaluate the ES per study and the degree of confidence, we did not evaluate the intensity of the intervention. Additionally, the majority of the studies included worked with individual children or small (2-6) groups. Only two studies (Filcheck, et al., 2004; Salend & Lamb, 1986) worked with entire classrooms. Despite varying treatment effects, contexts, and populations, TE's have served to increase positive behavior skills in children at risk for negative outcomes.

Methodological quality of the studies have not been evaluated as a unique group. The quality is a measure of the internal validity or trustworthiness of the findings (Vannest, et al., 2010). Significance of strong quality is we can believe the results to be true. Twelve studies (Filcheck et al., 2004; Himle, Woods & Bunaciu, 2008; Kazdin & Geesey, 1980; Kazdin & Mascitelli, 1980; Klimas & McLaughlin, 2007; Rosenberg, 1986; Salend & Allen, 1985; Simon, Ayllon, & Milan, 1982; Smith & Fowler, 1984; Sran & Borrero, 2010; Stevens et al., 2011; Sullivan & O’Leary, 1990) were found to be weak quality; therefore, their results could be questioned. Given the results, methodological quality did not appear to explain differences in effectiveness rather quality only explained the research design.

Our data can only speak to the overall effect size of TE and to what degree we can believe the results. These data are important because we obtained further analytical evidence that TE is efficiently capable of reducing challenging behaviors. In addition, the TE was effective at increasing academic readiness behaviors among student’s school settings.

Limitations

The present study, while broad in scope, nevertheless suffers from limitations that must be taken into account when considering its findings. First, while broad search criteria and an exhaustive screening process were used, a relatively small number of studies were located. The possibility exists that identified gaps in the literature base are actually the result of gaps in the search strategy employed here, particularly where earlier studies are concerned. At the same time, however, the

focus here was on interventions in the types of educational settings where which children frequently spend time, and other researchers (e.g., Schneider & Goldstein, 2008) have identified research conducted in these natural environments as an area of need.

Second, is the use of only A-B phase data in the calculation of the overall ES. Most of the studies included additional phases beyond the intervention phase. The additional phases could elicit additional information if analyzed. Third, this meta-analysis was limited to SCR, excluding any group designs so the extent of the research is reduced in that it does not summarize all available evidence on the effects of a token economy. Additionally, published articles are generally positive results so with the understanding that a bias may exist in favor of publishing studies with positive results; thus, it is a limitation that this meta-analysis to only included peer-reviewed published works and excluded unpublished desk copies, dissertations, and theses.

Implications

It was evident that the token economy was a successful intervention at decreasing inappropriate behaviors and increasing academic readiness behaviors. Additionally, the results may have implications for teachers in settings, such as classrooms, wherein reinforcers for appropriate behavior are often provided after a long delay (e.g., during a morning activity children may earn going to the playground early for lunch) tokens can be delivered immediately and on multiple occasions throughout the day.

An awareness of the collaboration necessary for implementing the token economy and the success of a token economy can positively change the manner in which students function. In addition, as school faculties and leaders implement a shared set of interventions with uniform procedures for data collection and communication, a greater sense of togetherness and purpose will take root and the overall building climate will improve.

Teachers, students, parents, and community members will all reap the benefits from the overall academic and climate improvement. Reducing the inappropriate behavior displayed by students will have a ripple effect that extends beyond the brick and mortar of the school building and manifest itself deep within the community. Increased teacher job satisfaction and improving the climate of a school will likely foster the development of a school that is a professional learning community at work. Once this transformation occurs, the possibilities are limitless.

Conclusion

The current meta-analysis differs from previous research synthesis by including ESs and CIs. The aim of the study was to use single case research studies to determine the overall effects of the use of TE. Overall, given that instructional and academic interventions are the most widely researched, and that this intervention may be integrated into existing classroom activities across a wide variety of settings, the potential for their utility in classrooms is great. What remains unanswered is whether TE's produce effects across contexts that are long lasting and that make an

impact on meaningful behaviors. Studies examining these active ingredients are needed, as well as ones that include the generalization of a TE to multiple settings.

CHAPTER III

MANUSCRIPT II: MODERATED EFFECTS FOR TOKEN ECONOMY

INTERVENTIONS FROM SINGLE CASE RESEARCH DESIGNS

The earliest behavioral interventions used teacher behavior in the form of social approval, disapproval, and ignoring to shape students' classroom behavior (Becker, Madsen, Arnold, & Thomas, 1967; Madsen, Becker, & Thomas, 1968). However, some research suggested that teacher behavior alone was insufficient to reduce disruptive classroom behavior, but that the use of concrete rewards through a token economy for students who behaved appropriately was effective (O'Leary, Becker, Evans, & Saudargas, 1969). Initially, token economies were used to reward appropriate behavior while ignoring inappropriate behavior. Derived from operant learning theory (Skinner, 1931), TE is a secondary reinforcement system (Alberto & Troutman, 2003) in which an individual earns tokens for targeted behaviors and is able to exchange accumulated tokens for a "larger" reinforcer (Kazdin, 1971) usually referred to as a back-up reinforcer.

The predominant treatment approach for promoting the social, adaptive, and behavioral functioning of children has been based on behavioral theory (Bregman, Zager, & Gerdtz, 2005). However, the sophistication of intervention strategies has increased substantially, reflecting advancements in techniques and refinements in strategies. TE, a behavior management strategy, is one such complex intervention with sufficient research that support of effectiveness (Dickerson, Tenhual, & Green-Paden, 2005; Higgins, William, & McLaughlin, 2001; Kimlas & McLaughlin, 1981;

Matson & Boisjoli, 2009; Salend, Tintle, & Balber, 1988; Sran & Borrero, 2010; Stevens et al., 2011; Truchlicka, McLaughlin & Swain, 1988). However, increasing the complexity of intervention (i.e. adding multiple procedural steps) might have a negative impact on teacher acceptability and fidelity. Teachers do not have sufficient time (Ingersoll, 2003; Vannest, Hagan-Burke, & Parker, 2006; Vannest, Soares, Harrison, Brown, & Parker, 2010), are ill prepared to implement behavioral management strategies (Lannie & McCurdy, 2007) and have multiple diverse learners who require varied behavioral strategies (Grazano, 2005). Although effective behavior management does not guarantee effective instruction, it establishes the context of a structured learning environment (Emmer & Stough, 2001) whereby increasing the potential to decrease the high levels of stress and symptoms of burnout for teachers (Berliner, 1986; Browsers & Tomic, 2000; Espin & Yell, 1994).

Intervention complexity can be minimized by increasing implementer understanding of moderators. A moderator variable (M) is a variable that alters the strength of the causal relationship. A moderator analysis can conceivably identify which moderators have the most impact, whereby improving the ability of teachers to effectively manage interventions by using only the moderators, which produce the most effect.

Potential Moderators

In order to evaluate the necessity of potential moderators it is essential to compare, combine, and contrast findings of relevant studies. One way of achieving

this is through a meta-analysis, which combines quantitative results across a set of studies about the same topic to measure the impact in moderator variables thus explaining the differences in moderators (Cooper & Hedges, 1994; Glass, McGaw, & Smith, 1981; Hedges & Olkin, 1985; Hunter & Schmidt, 1990; Rosenthal, 1991). The moderator variable is a third variable (in addition to the primary independent and dependent variables) that may influence the differences in the strength or direction of observed relationships between the primary variable of interest (Steel & Kammeyer-Mueller, 2002). Differences are typically quantified using ES indices (Holmes, 1984; Snyder & Lawson, 1993). Several potential moderators of the included token economy (TE) studies can be explored as potential systematic sources of between study differences.

Potential moderators include the following: (a) student characteristics including age and setting (b) response cost, (c) verbal reminder during token earning phase, (d) active ingredients, and (e) academic versus behavior goals.

Student Characteristic Variables

Student characteristics variables are age/grade and instructional setting. TE has been found to be effective for elementary age students (Akin-Little & Little, 2004; Christensen, Young, & Marchant, 2004; Filcheck, McNeil, Greco, & Bernard, 2004), junior high students (Carlson, et al., 1992; Cavalier, Ferretti, & Hodges, 1997; Feindler, Marriott, & Iwata, 1984; Heaton & Safer, 1982; Safer, Heaton, & Parker, 1981), and high school (Schellenberg & Skok, 1991; Wheeler, Freagon, & Stern, 1985). Even though TE have found to be effective we do not know if they are more

effective for any one group. TE has been found to be effective with student with high incidence disabilities (e.g. EBD, LD, ID, and communication disorders), low incidence disabilities (e.g. blind, deaf, developmental delays, physical impairment, autism) and children with no disabilities (Truchlicka, Mclaughlin, & Swain, 1998) in numerous instructional settings such as inclusive, general education, special education and alternative settings (DeMartini-Scully, Bray, & Kehle, 2000; Rhode, Jenson, & Reavis, 1993). The type of setting in which the intervention is delivered is relevant. For example, an intervention delivered in a self-contained special education setting may yield stronger effects, since a greater number of natural opportunities due to the lower student teacher ratio are available for reinforcement. Research supports that self-contained settings might yield strong initial effects but perhaps may have poor generalizations to a less restrictive environment (Odom, et al., 2003).

Response Cost (RC)

RC is a procedure that attempts to decrease behavior by contingently withdrawing a specific amount of reinforcement following an inappropriate behavior or response (Kazdin, 1975). With differential results across studies, RC's impact on TE's effectiveness is an unanswered question. Previous research supports successful RC procedures (Broughton & Lahey, 1978; Gresham, 1979; Rapport, Murphy, & Bailey, 1980; Witt & Elliot, 1982) but notes RC can be time consuming. Other studies (Phillips, Phillips, Fixsen, & Wolf, 1971) found RC has harmful side effects when included in TE systems; such as, the opportunity for the implementer to over

penalize and the possibility of decreasing the incentive of demonstrating the target behavior. To overcome these side effects, Witt and Elliot (1982) included extra privileges that could be removed using RC. Some studies of TE found that when response cost was added to TE the effectiveness increase and remained even after the RC was removed (Azrin & Holz, 1966; Broughton & Lahey, 1978; Kazdin, 1972). In contradiction, Witt and Elliot (1982) found no conclusive evidence of the effect of RC after removal.

Verbal reminders during the token earning phase (cueing) by the implementer

A cue is any type of signal used to prompt another person to either engage in or disengage from a particular behavior. Research concerning the effectiveness of verbal reminders for earning tokens within a TE is inconsistent. The literature reveals proponents of cueing (Latham & Locke, 1991), as well as those who find negative effects of cueing (Balcazar, Hopkins, & Suarez, 1985, p. 65; Kluger & DeNisi, 1996) when using a TE.

Active Ingredients

Active ingredients are the component or procedural steps needed to implement the intervention effectively. Knowing what active ingredients are essential for TE to be effective across behaviors, populations and settings is important because no intervention can be effective if one or more of its essential active ingredients are missing (Yap, Aldersebaes, Railsback, Shaughnessy, & Speth, 2000). Vannest, Reynolds, and Kamphaus (2008) identified 5 basic elements and 12 procedural steps, but their review was not an empirical test. Other studies report a

number of steps or “active ingredients” ranging from four (Cooper, Heron, & Heward, 1987; O’Leary & O’Leary, 1977) to seven (Miltenberger, 2000). Differing combinations of these active ingredients have been included in the TE studies: (a) defining the target behavior, (b) type of token, (c) reinforcement schedule, (d) reinforcement menu, (e) reinforcement survey, (f) exchange rate, and (g) response cost. Cooper et al. (1987) proposed six active ingredients (i.e., select tokens, identify target behaviors, select backup reinforcers, establish ration of exchange, write procedures about token presentation/exchange, and field test). All of these components overlap with Vannest et al. (2008) except the field test. Different researchers may use some of the same active ingredients, additional active ingredients, or fewer active ingredients.

Academic versus Behavior Goals

TE is the most widely researched and validated behavioral intervention in the schools (McLaughlin & Williams, 1988; Swain & McLaughlin, 1998). Behavior modification research with children has demonstrated effectiveness of TE in producing improvement when successful performance on academic subjects is selected as the "target behavior" for modification (Kilmas & McLaughlin, 2007; Salend, Tintle, & Balber, 1988; Sran & Borrero, 2010) as well as behavioral outcomes (Center & Wascom, 1984; De-Martini-Scully, Bray & Kehle, 2000). However, no study evaluated if TE works equally well with academic goals and well as behavioral goals.

Purpose and Research Questions

The purpose of this study was to investigate the moderators of effects in a token economy system. Including only the necessary moderators in the TE design might reduce the possibility of teachers becoming over stressed by classroom management interventions and the lack of time to deal with the resulting complexity. In addition, identifying only necessary active ingredients might decrease teacher frustration (Epstein, Atkins, Cullinan, Kutash, & Weaver, 2008; Lewis, Hudson, Richter, & Johnson, 2004; Lhospital & Gregory, 2009; Maccini & Gagnon, 2006; Walker, 2004) and time. If teachers find an intervention more acceptable because it is less frustrating and time consuming, they are more likely to implement it (Mathews, McLaughlin, & Hunsaker, 1980; Vannest, Mahadevan, Mason, & Temple-Harvey, 2009; Witt & Elliot, 1982). However, while multiple individual studies have been conducted with each of the stated participant characteristics, no meta-analysis have evaluated the differing effects across studies or moderating variables. In addition, moderator effects on TE have not been evaluated.

The research questions seek to identify if difference of effects exist in potential moderators:

1. Is TE different across age groups?
2. Is TE more effective in special education settings or general education settings?
3. Does the implementation of a response cost enhance the effectiveness of a TE?

4. Does a verbal reminder of token earning during the intervention phase enhance the effectiveness of TE?
5. What are the most influential active ingredients?
6. Is TE differentially effective for academic versus behavioral?

Method

The following sections outline the process by which empirical research on token economies was acquired, screened, examined, and combined. First, a description of the literature review procedures will be described. Next, the system for coding studies is presented, along with method for calculating inter-rater reliability. Finally, the procedure for calculating effect sizes and combining studies with like moderators is discussed.

Comprehensive Literature Review

Three search strategies were used to secure a systematic representative sample of published studies. First, relevant studies were identified through computer searches of PsycINFO and Education Resource Information Center (ERIC) electronic databases using search terms and their variants: (a) token economy (b) intervention, (c) reinforcement, (d) contingency management, (e) systematic positive reinforcement, (f) tokens, (g) operant conditioning, (h) applied behavior analysis, (i) back-up reinforcers, (j) behavior therapy, (k) points, and/or (l) response cost. Second, the reference lists of each identified study were examined. Third, hand searches were conducted in relevant repeating journals.

Studies were included in the analysis only if they included the following five criteria. First, studies must measure effects of a TE intervention, defined as a program in which students earn tokens for desirable behaviors and then exchange the tokens earned for back-up reinforcers (Alberto & Troutman, 2003; Martin & Pear, 2003). Second, study venues included only U.S. classroom settings of school age children (age 3-21). Third, studies must have been published from 1980 to 2011, in order to create a manageable set of articles representing typical public school classrooms, which compare to a modern day classroom settings that include general and special education. In addition, studies were included that were not available prior to Kazdin's review in 1980. Fourth, studies utilized SCR methodology with a clearly readable graph of data, group studies were excluded to allow for continuity in comparison of effect sizes (Lipsey & Wilson, 2001) This allowed for visual inspection of the data, especially at the phase change from baseline to intervention – referred to as the intercept gap (Kazdin, 1982). Furthermore, visual analysis allowed for an inspection of the trend line and overlapping data between baseline and intervention. Additionally, the visual graph allowed for extraction of original raw data for new ES analysis (see Data Extraction). When studies included multi-component interventions, the interventions assessed were required to be graphed individually so that the TE intervention could be separated and analyzed. Last, studies must have been published in peer-reviewed journals to ensure authors meet the accepted standards for their field and to prevent dissemination of irrelevant findings, unwarranted claims, unacceptable interpretations, and personal views

(Ludwick, Dieckman, Herdtner, Dugan, & Roche, 1998). Publications that had not undergone peer review such as dissertations, descriptive articles, and unpublished desk copies were excluded.

For this meta-analysis, an effort was made to include as much of the population of empirical research as possible given the criteria for inclusion outlined above. The previously listed search terms were combined in the search engine so that any study in which the title or abstract contained a key word would be included in the results. The titles and abstracts of the studies acquired through this process were examined and were retained for screening if they were TE studies of school age children.

In addition to the search of databases, a hand search for titles related to token economies or secondary reinforcement was completed by reviewing the table of contents in the major journals in special education, school psychology, and behavioral psychology (as determined by ISI rating for the past 2 years): e.g., *Journal of Positive Behavior Intervention*, *Behavior Therapy*, *Behavioral Interventions*, *Journal of Special Education*. These journals were selected based on their prominence in the field of publishing intervention related articles. This process did not yield any additional studies that were not already evaluated in the abstract and full review stage.

Reference pages of all the resulting screened articles were inspected for additional eligible studies. Full documents for the resulting studies were acquired, at which time they were examined relative to the inclusion criteria. Any study that did

not clearly fail the basic inclusion criteria was retained for review. This strategy captured many articles written by several of the more prominent and well-known authors and experts in this field; however, it did not yield any additional studies that were not found in the original database search.

Coding

All included articles were coded by “labeling” each piece of data. Researchers complete this coding process in order to make sense of the data for capturing and analyzing potential moderating variables. Potential moderating variables in each of the included studies will be coded so differences between the studies can be examined for their potential effect. Possible variables are: (a) student characteristics, (b) response cost; (c) active ingredients; (d) verbal reminder during token earning phase, and (e) academic or behavior goals. A protocol adapted from Tolan, Bass, Henry, and Schoeny (2008) will be used to code relevant study characteristics; it is included in Appendix A. Operational definitions for the potential moderators can be found in Appendix A. In addition to general study information (e.g. author, publication type, location of study), coded study characteristics will include extensive information about each study’s research design, participants, context, and educational setting, as well as any information that would help identify or calculate study effect sizes.

Inter-rater Reliability of Coding

Retained studies were coded by the author. In order to establish inter-rater reliability for the coding protocol used here, a second doctoral student in Special

Education served as a second rater. Initially, the two raters coded several articles together and discussing in particular any studies containing unclear information. Official coding did not begin until a minimum acceptable value (range from .80 to .90) of inter-rater agreement was met (Hartmann, Barrois, & Wood, 2004). Next, the raters coded independently 6 studies (25% of the total number included). Inter-rater reliability was defined as the frequency of agreement on codes divided by the total number of coded categories, expressed as a percent.

Cohen's Kappa reliability agreement using NCSS (Hintze, 2004) was calculated by entering both agreements and disagreements to a 2 X 2 matrix. For cross tabulation, matrix data was entered into the NCSS statistics program, which provided the Cohen's Kappa (Kappa) index. Kappa adjusts for the expected chance agreement, thereby making it a conservative measure of reliability and perhaps even underestimating agreement (Ary & Suen, 1989; Strijbos, Martens, Prins, & Jochems, 2006).

Reliability was also reported in PABAK-OS. Rater scores were transferred into an agreement matrix that allows for the calculation of inter-rater reliability with PABAK-OS; "prevalence and bias adjusted Kappa for ordinal scales" considers special attributes of ordinal data by assigning differential weights. Just as PABAK (Byrt, Bishop, & Carlin, 1993) corrects Kappa's undesirable sensitivities with nominal scales, PABAK-OS does the same for ordinal scales (Parker, et al., 2011).

Agreements and disagreements in coding were resolved through discussion.

Data Extraction

Although the intervention (TE) is the same across the studies, digitizing was used to obtain raw data for recalculation of ES. All single subject graphs of included articles were digitized in Getdata Graph Digitizer (Version 2.21) from getdata.com.ru. Raw data from digitized primary studies was transformed into a numerical scale to enable proper comparisons (Glass, 1976).

First, PDF versions of included articles were obtained. Each SCR graph was scanned into Get Data Software. The X & Y-axis were defined by setting the values. Each data point was converted from published graphs to raw data using the software. The values were then imported to an Excel[®] spreadsheet with a column added for phases (see Parker et al., 2007). The numeric data allows analysis for calculation of effect size.

Effect Size and Moderator Calculations

The effect size statistic represents the size and direction of the relationships among variables in a study (Lipsey & Wilson, 2001). By calculating ES, the outcomes of individual studies may be interpreted in relation to each other. Critical phase contrasts (A vs. B; Baseline vs. Intervention) were identified for each design, and their ES will be aggregated for one or more ES representing the entire design. Effect Sizes will be “non-overlap with trend,” a new method: Tau-U

The data was analyzed using the Tau-U calculator (singlecaseresearch.org) for individual ES. The ES and standard error will then be entered into the statistical

program WinPEPPI (Abramson, 2010) for analysis. The software also provides confidence intervals.

Moderator effects have also been analyzed using the Tau-U ES and a T-Test statistic. Differences among effect sizes might be related to different study characteristics. To analyze the impact of study characteristics on the variability of the effect size a t-test is computed. The moderators were coded dichotomously using 1 for yes and 0 for no. For example, if a response cost was used in the intervention the moderator was coded 1. A t-test statistic was used to evaluate the differences in means of the two groups. This procedure was repeated for each moderator.

Results

The results of this study will be presented in the following sequence. First, general information about the results of the literature search will be reported. Second, the results of coding and inter-rater reliability of coding will be described. Finally, analysis of each moderator will be reported.

Literature Review Results

The researcher included studies in the analysis that met the following five criteria. First, studies addressed a TE intervention that met this definition: TEs are programs in which students earn tokens for desirable behaviors and then exchange the tokens earned for back-up reinforcers (Alberto & Troutman, 2003; Martin & Pear, 2003). For studies of multi-component interventions, the interventions assessed were required to be graphed individually. Second, study venues included only U.S. classroom settings in public schools of school age children (age 3-21).

Third, studies were published from 1980 to 2011 in order to compare to a modern day classroom settings. Fourth, studies were SCR designs and included a visual graph of data that allowed for extraction of original raw data for new ES analysis. Last, studies were published in peer-reviewed journals to ensure authors meet the accepted standards for their field and to prevent dissemination of irrelevant findings, unwarranted claims, unacceptable interpretations, and personal views. Publications that had not undergone peer review such as dissertations and unpublished desk copies were excluded.

The full text of forty-four primary studies were examined for potential inclusion in this meta-analysis. Twenty studies were excluded for the following reasons: three took place in a school that was located in a hospital setting, five did not have legible visual graphs from which digitizing could occur, five had a multi-component intervention in which the data could not be separated, and seven did not undergo the peer review.

Inter-rater Reliability of Coding

Data was captured and coded using an adapted protocol from Tolan, Bass, Henry, and Schoeny (2008). The coded categories are: setting of the study, participants, intervention features, methodology, and measures. Reliability was calculated on the five coding categories listed in Table B-8 and was 96% or higher for all of the areas coded, with an overall reliability for all of the codes of 99%.

Participants and Characteristics of Included Studies

Table B-9 summarizes the number of studies, participants, and calculated ES as well as age, disability category and the intervention setting of the included studies. The final group of 24 included studies involved 84 participants and produced 79 individual effect sizes. Sixty-five percent of the studies took place with children ranging from 6 – 15 year old, 35% of studies were implemented with children ranging from 3 to 5 year old. Eighteen (72%) of the studies included participants with a special education eligibility: one (4%) included participants that were deaf, two (8%) studies included students with an intellectual disability, three (12%) studies included participants that were learning disabled, four (16%) included students that were emotional/behavioral disorders, and eight (30%) were identified special education but did not state an eligibility category. The setting of the interventions occurred equally in general and special education classrooms, 12 studies took place in general education and 12 in special education.

Table B-10 includes a summary of study characteristics including dependent variable, educational status of the participants, and inclusion status of response cost (RC). TE were implemented to decrease behavior in 63% of the included studies and 37% used TE to increase academic readiness behaviors (e.g. attentive class behavior, assignments complete, assignments per minute). Seventy-one percent of studies were conducted in special education settings and 29% were conducted in general education settings. RC was used, as part of the intervention in 67% of the studies and 33% did not include RC.

Analysis of Moderators

To explain study differences, several moderator variables were examined. The results of an analysis on student characteristics (e.g. age and setting), as well as response cost and intervention related variables will be discussed.

Student Characteristics Moderator – Age. In answer to research question one is TE different across age groups? Two groups were established, preschool (ages 3-5) and grade school (ages 6-15) because research established that particular problem behaviors such as noncompliance is more common in preschool settings (Taplin & Reid, 1973). This is of particular importance as these behaviors are often associated with later academic and social readiness. Noncompliance with instructions is common in preschool settings (Crowther, Bond, & Rolf, 1981) and may be particularly common when children are asked to terminate a preferred activity (e.g., free play) or initiate a nonpreferred activity (e.g., clean-up). Results indicated a statistically significant difference in ES, $\text{Tau-U} = .90$ and $.71$, with significance better than $.05$, at $.03$. A two-sample t-test was conducted to compare age $>$ or $= 5$ and age < 5 for participants in TE intervention. There was a significant difference in the scores for 3-5 year olds ($M=.71$, $SD=.17$) having a moderate effect size and 6-15 year olds having a strong effect size ($M=.90$, $SD=0.17$); $t(23) = 2.272$, $p = 0.03$. Specifically, the results showed the mean effect size for students' ages 6-15 was significantly higher than for 3-5 year olds. This suggests that the TE interventions were equally effective in reducing inappropriate classroom behavior

and or increasing academic readiness behaviors with all ages; however, the effectiveness was greater older children.

Classroom Setting. The second research question was to determine if classroom instructional setting (general education or special education) was a moderator of TE effects. TE is equally effective in both general education (ES = .85) and special education (ES = .86) classrooms. A two-sample t-test was conducted to compare the differences across classroom settings. The intervention occurred equally in special education (12) and general education (12). The differences in the mean ES was not statistically significant at $p = .86$. Mean ES for regular classroom setting ($M = .85$, $SD = .19$) compared to a special education setting ($M = .86$, $SD = .21$); $t(23) = .18$, $p = .86$.

Response Cost. The third research question asked if implementation of a response cost is a moderator of the effectiveness of a TE? Results indicate the effect sizes (.84 and .89) were not statistically significant. Dichotomous coding of 0 for no response cost and 1 for response cost allowed for a comparison of this moderator. A two-sample t-test was conducted; however, yielded no significant differences. Eight studies reported no use of a response cost ($M = .89$, $SD = .21$); Sixteen studies reported use of a response cost ($M = .84$, $SD = .19$); $t(23) = .72$, $p = .48$.

Verbal Reminder during Intervention. The fourth research question inquired about verbal reminder during intervention. The results indicated strong effects for using a verbal reminder but this moderator did not necessarily indicate statistically significant differences. Twenty-four studies were included in this

analysis with 19 studies reporting the use of a verbal reminder of token earning during the intervention. Nineteen studies reported using a verbal reminder ($M = .88$, $SD = .13$); five studies did not report inclusion or lack of inclusion on using verbal reminders during intervention ($M = .75$, $SD = .24$); $t(23) = .85$, $p = .40$.

Active Ingredients. The research question sought difference in the steps or combination of steps for effectiveness of TE. The researcher considered five steps. The five steps were: (1) Included was a visual chart of Goals, (2) study outlined specific token earning times, (3) the token was identified, (4) included was a visual chart or display of reinforcers and their cost, and (5) access to back up reinforcer. By definition of a TE (Alberto & Troutman, 2003) step 5 was dropped from the analysis as all 24 studies included access to a backup reinforce leaving four steps for analysis. Table B-11 summarizes the following results.

Step 1 – Visual Chart. Twenty-four studies were included in this analysis with 9 studies reporting they used a visual chart of student goals during token earning times. The overall mean effect sizes for using a visual chart are strong ($M = .94$) compared to not using a visual chart ($M = .70$) this comparison yielded statistical significance at a $p = .05$.

Step 2 – Specific Times for Token Earning. All 24 studies were coded dichotomously using a 1 for yes and a 0 for no. Three studies did not list specific times for token earning ($M = .89$; $SE = .11$) and 21 studies listed they did have specific token earning times ($M = .85$; $SE = .04$) $t(23) = .32$, $p = .75$. Effect sizes

produced virtually no differences and there was no statistical significance in the combined means.

Step 3 – Tokens Were Identified. Tokens were identified in 22 of the included studies ($M = .84$; $SE = .04$) leaving 2 studies that did not identify a token ($M = .97$; $SE = .02$); $t(23) = 2.59$, $p = .01$. This comparison did yield statistical significance; however, with the ratio of comparison more than 2 data points are needed for a true comparison.

Step 4 – Visual Chart of Reinforcer and Cost. All 24 studies reported on this variable with only 3 reporting they did include a visual chart and cost of back up reinforce ($M = 1$; $SE = 0$) and 21 did not report the child was aware of the cost of back up reinforcers ($M = .84$; $SE = .04$); $t(23) = 3.7$, $p = .00$. This did yield statistical significance but only 3 data points reported using a chart so the results are not reliable.

The steps were combined in all possible combinations to determine if there is a differential effect using one or more of the steps (see Table B-12). There is no difference in having two steps versus four steps. Studies that included steps 2 and 4 were the same as studies that included all 4 steps so it is hard to determine the differential effect of step 3.

Effectiveness for Academic versus Behavior Goals. The fourth research question asked does TE have differing effects for academic versus behavior goals? The results indicated a strong effect for both academic and behavior goals so TE can be used as an intervention for both outcome measures. Five studies reported the use

of TE to increase academic readiness behaviors ($M = .75$, $SD = .34$); Nineteen studies reported the use of TE to decrease inappropriate behaviors in the classroom ($M = .88$, $SD = .13$); $t(23) = .85$, $p = .40$. A two-sample t-test was conducted to test the means of the use of TE for academic goals ($n=5$) versus behavior goals ($n=19$); however, yielded no statistically significant differences.

Discussion

The current study was conducted to examine the moderators and their effects on TE across all available peer reviewed studies focusing on interventions implemented in school settings from 1980 to 2011. Specifically, the current study evaluated age, classroom setting, response cost, verbal reminder, active ingredients, and differences of using a TE for academic or behavior goals. This was unique, as previous reviews (Kazdin & Bootzin, 1972; Kazdin, 1982; Dickerson, Tenhual, & Green-Paden, 2005; Matson & Boisjoli, 2009) did not complete analysis or discuss moderators and their effects.

The research question in the present study that examined age differences in participants requires some discussion. This moderator compared age range of 3 – 5 year olds ($M=.71$) to 6-15 year olds ($M=.90$). The researcher found a strong positive and statistically significant correlation for preschoolers versus older children, $t(23) = 2.31$, $p = .03$. TE was effective with both groups of children; however, it was most effective with children over the age of six possibly because older children are determined to be more compliant and have more control over their individual choices of backup reinforcers. The finding that older children complied with more

instructions than younger children is consistent with previous research (Brumfield & Roberts, 1998; Shriver & Allen, 1997) and suggests that the age of the child may be predictive of compliance with an intervention.

The second moderator to be examined was the classroom setting. The researcher set out to determine if a setting had an impact on the effectiveness of a TE. An analysis of instructional settings showed that no instructional category was associated with statistically significant results. While special education settings produced an overall ES of .85 smaller than those in a general education setting (.86), this difference was not significant. This moderator is fascinating in the context that special education services can be offered in a variety of settings so if the TE is an intervention that is needed by a special education student the intervention can be offered in general education settings and the effect will be similar. However, this is not consistent with DuPaul and Eckert (1997) where they found interventions had a greater impact on behavior when they were implemented in special education classrooms as opposed to implementation in general education.

Response cost can be applied within token economies and is a procedure that consists of immediate withdrawal of tokens for inappropriate behavior. Research concludes response-cost procedures have been effective in reducing the frequency of undesirable behavior when the magnitude of the cost significantly taxed the availability of backup reinforcers (Burchard and Barrera, 1972; Kazdin, 1972). In this study response cost as a moderator did not have a significant impact on the effectiveness of a TE. Eight studies did not use a response cost procedure ($M = .89$)

and sixteen studies did use a response cost ($M = .84$). The TE did not appear to be weakened if the RC procedure was not used. No statistically significant differences were obtained between groups [$t(23) = .73, p = .48$]. The mean effects of both groups were strong. As Kazdin (1973) pointed out, any given negative effect of a RC should not be cause for dissolution of TE since the RC is a component that can be removed from the intervention.

A verbal reminder (cueing) is means to induce an individual with added prompts to perform a desired behavior. Research concerning the effectiveness of verbal reminders is inconsistent. The mean effects of 19 studies which included a verbal reminder ($M = .88$) indicated strong effect; however, the five studies which did not use a reminder produced a moderate mean ES ($M = .75$). This result is indicative of the previous research and the inconsistent results produced; however, it is inconsistent with previous research that found negative effects of verbal prompts (Balcazar, Hopkins, & Suarez, 1985, p. 65; Kluger & DeNisi, 1996). Teacher prompting during intervention is time consuming and may not be needed in order for an intervention to be effective. Furthermore, this study did not produce statistically significant differences in the means for using verbal prompts.

Teachers are in need of effective interventions that can be implemented quickly and easily (Eber, Sugai, Smith, & Scott, 2002). Analysis in this study indicate step 1 (visual chart of goals/behaviors) and step 2 (specific token earning times) are the minimal essential elements for effectiveness of TE with the understanding step 5 is access to back up reinforcer. While differential effects of

step 3 cannot be determined results indicate that step 4 is not deemed necessary by other researchers as it is often omitted. However, insufficient data was unavailable for analysis in this study. Studies that included steps 1 and 2 combined were the only combination to produce statistical significance. Nine studies had both steps 1 and 2 producing a mean ES of .94. The remaining 15 studies that did not include both steps 1 and 2 produced a mean ES of .80 at $p = .05$.

The analysis regarding student behavior and increasing academic readiness behaviors through the utilization of the TE indicate that the program was effective for decreasing inappropriate behaviors, the academic readiness behaviors were also improved using a TE. The five studies that used TE for academic skills produced a mean ES of .75; the nineteen studies that used TE for behavior produced a mean ES of .88. TE is not statistically significant on either category. This research is consistent with and elaborates upon conclusions from Matson & Bisojoli (2009) which states that TE is still one of the more effective forms of behavior modification. With an orientation toward a positive environment, students can be rewarded systematically to progress toward individual or group goals. TE can be used on a continuum as a school-wide intervention or as an individual intervention reducing the inappropriate behavior displayed by students which can have a ripple effect that extends to the teacher's positive classroom management strategies. The TE appears to be one of the primary intervention models with evidence-based research to support it at an individual, small group, or school wide setting.

TE is one evidence-based intervention that can be implemented efficiently with minimal teacher effort as is evidenced in this study, if consideration is given to the essential active ingredients. While inconsistencies are found in the literature regarding the significance of the characteristics of students and the active ingredients necessary for effective implementation results here, indicate that effects can be found with many combinations of moderators. Teachers can quit attempting to implement complex, complicated TE systems as all that is needed is some tokens and a chart.

Limitations

There are limitations to the present study. First, I examined only SCR studies that were implemented in public schools employing the intervention TE. If the scope had been broadened to include all studies using TE, it might yield different conclusion. At the same time, this study, though small in scope was completed to justify the effects of moderators within a TE and many of the studies that included various settings (e.g. hospital, prisons, and residential treatment centers) implemented the TE in the same manor. Therefore, the outcome may in fact be more similar than different.

Next, the active ingredients (i.e. procedural steps) are difficult to access because many studies do not report the design of the intervention. Due to the potential complexity of the intervention specific measures may fail to capture similarities or differences due to lack of reporting of active ingredients.

Implications

Several implications are apparent from the current study. The token economy intervention is an effective behavior management tool for increasing academic readiness behaviors and decreasing inappropriate behavior. First, there are various ways to integrate a token economy into the special and general education classroom settings. This multifaceted intervention is successful with an individual or group implementation. It can be as simple as marbles in a jar to reward an entire classroom or individualized for a specific student for problem behaviors.

Second, tokens provided more opportunity for instructional practices because reinforcement could happen immediately and consistently during instruction by passing out tokens. This allows for a teacher to multi-task and not interrupt instruction to reinforce behavior.

Third, user-friendly interventions need to be identified for practitioners in search of reinforcement based protocols for behavior problems (OSEP Center on Positive Behavioral Interventions and Supports, Sugai, et al., 2000). A systems change needs to evolve throughout school facilities, beginning with informing educators about feasible, evidence-based practices. The intent of the TE intervention meta-analysis was to provide further evidence for reinforcement based procedures when working among children in school settings. The intervention was supported by literature suggesting that token economies encompass the notion of being included in a positive behavior support plan while providing an acceptable treatment alternative to punitive procedures.

According to the article, Center for Effective Collaboration and Practice (2005), it is important to modify by changing the rewards periodically to the program to avoid boredom. However, it is important that the students understand the programs principals before changing it.

CHAPTER IV

MANUSCRIPT III: TOKEN ECONOMY: MYTHS, APPLICATIONS, AND SUMMARY

Even experienced teachers sometimes struggle with classroom management in today's challenging and high-pressure environments. New students mid-year, differences in school and home expectations, language and cultural differences, exceptionalities and gifts can all present a need for a top-notch classroom management system. Classroom educators experience a loss of 4 hours of academic instruction each week due to behavior management (Walker & Gresham, 2003; Vannest & Parker, 2009). Add the time demands for academic performance, adequate yearly progress, and response to intervention programming and all teachers need an efficient, effective method for maximizing instructional time and motivating students to perform academically and behaviorally.

Studies tell us that while new teachers are likely to trust practices they learned in their credential programs or read about in journals, experienced teachers are less likely to do so (Landrum, Cook, Tankersley, & Fitzgerald, 2007). While the reasons for this may be many it seems clear that practices that are "supposed" to work sometimes fall short of teacher expectations. Nothing is more frustrating than leaving a workshop enthusiastic, only to abandon the technique a few days later when it falls apart or does not work "as advertised". Nor is it encouraging attending a workshop on best practices only to hear about chasing kids around the room with M&M's. However, many effective practices are not new; some may already be in

your repertoire. These practices consistently demonstrate effectiveness in the field and it may be time to revisit a “classic” that we once abandoned. Consider the following case and see if you have ever had similar experiences with an evidence-based practice.

Case Study

First year teacher Ms. Williams struggles to teach content to her 4th graders for more than 10 minutes without addressing inappropriate behavior. She remembers learning a “guaranteed technique” and sets out to implement a token economy. After a run to a teachers supply store (too expensive) and a local discount-mart (nothing right) she purchases some supplies spends most of the weekend designing her own system. Since her classroom theme is “bees” (who can fly even though physics says they can’t) she creates token earning cards with her theme (see Figure A-2) and laminates them, 50 bucks and a trip to a “less-than-a-dollar” store yields some tangible back up reinforcers such as pencils, stickers, small bottles of bubbles, and cheap headphones. She identifies her 3-5 classroom “rules” and writes out her expectations.

School starts Monday morning and Ms. Williams explains her system to the students, but by Wednesday, she is no longer consistently passing out tokens and she is out of “rewards”. One student is

stealing from the treasure chest and others are talking so much she begins to take away the tokens. A few other students are constantly asking her when they can go to the treasure chest. She realizes she needs a visual to remind herself and her students but feels like she created a monster system to bribe kids into acting how they are supposed to act in the first place.

For those of us who have ever been in Ms. Williams shoes, there are three common challenges in implementing any evidence-based intervention. First the teacher needs to use all the required components or steps of an intervention with accuracy (called fidelity), second the teacher has to believe there is benefit to student, third, the teacher must find acceptable the time, costs, and see few barriers to implementation (Vannest, Soares, Harrison, Brown, & Parker, 2010; Witt & Elliott, 1982). In our case study, Ms. Williams experienced each of these common challenges. The purpose of this paper is to provide information about practical methods and procedures for teacher application. This paper describes the evidence for a TE and addresses how each of the common barriers to implementation of an evidence-based practice might be overcome. We provide examples of what a token economy might look like and, address some of the myths in using a token economy as a tier one or tier two interventions.

Introduction

A Token Economy is a system of behavior modification derived from principals of operant conditioning. Behavior is changed through reinforcing the

occurrence of new behavior or reinforcing an increase in behavior. Token economies typically involve four to seven steps (although as few as two and as many as twenty) (see Figure A-5).

The primary goal of a TE is to increase desirable behavior and decrease inappropriate behavior. Something needed in every classroom, and every age group or type of student. TE's offer a system of extrinsic reinforcement or tokens for the accomplishment of work, the achievement of a goal, or the demonstration of certain behavior such as, turning in homework, listening to instruction, helping a neighbor, being a good sport, or arriving on time to class. Such systems of reinforcement can be as simple as adding a marble to a jar on the teachers desk or as complex as individual goals and individual records.

A token economy is an evidence-based practice with a long history of empirical support (see Table B-1). Despite this history and clear evidence of effects, a survey of teacher's classroom management practices show that token programs to increasing appropriate behavior were used by only 30 percent of teachers; and the average frequency of use for those 30 percent who used them at all was between "not at all" and "just a little" (Rosen, Taylor, O'Leary, & Sanderson, 1990).

A TE can be used with a group or individual. For example, Bella often threw her pencil across the room when she could not decide the answer to a math problem on her worksheet and was not receiving teacher attention. An individual TE could be used - Bella's notebook had a visual chart stating a behavioral expectation goal for using her pencil for writing only and raising a hand when help was needed (Figure

A-4). At the end of Math, if Bella had used her pencil for writing only, the teacher placed a star beside that goal. Earning enough stars led to a chance to choose the song played during dismissal time.

The TE could also change behavior as part of a class wide application. A classroom chart indicating 3-5 behaviors might include to raise a hand when you need help and to be safe and respectful (Figure A-3). Tokens could be “passed out” physically during class using raffle tickets, small printed “dollars” or tokens could be “passed out” metaphorically by checking off a record sheet or using a white board to keep “score”. Tokens are a secondary reinforcer and not worth anything themselves but are exchanged for something of value referred to as a back-up reinforcer (Alberto & Troutman, 2003; Martin, & Pear, 2003). The tokens acquire power when they are paired with back-up reinforcers.

TE is powerful enough to work in challenging environments like prisons, mental health hospitals, and psychiatric facilities (Comaty, Stasio, & Advokat, 2001; Paul & Lentz, 1977) yet easily adapted for use in classroom settings (Soares, 2011). Most adults and students alike have had some experience with token economies. For example, most of us have seen the home “chore chart” in which children receive stars after completing chores; and the stars earned something, maybe a trip to a movie. A TE is simply a contingency management system and the target behaviors can be the same for all students in a class or individualized for students with specific or specialized needs. So what is the evidence for using a TE in schools? How much behavior change might I expect to see?

Evidence for Effectiveness

Combining the results of 79 experiments reported in 24 prior studies TE can essentially change behavior to a large degree, 78% of the data comparisons between phases A (no token economy) and phase B (token economy) show improvement during phase B. TE is equally effective in general and special education settings and slightly more effective with older children (.90 for ages 6-15) but still effective with younger students (.71 for ages 3-5) (Soares, 2011).

As we revisit Ms. Williams, we see that she experienced all the common problems with implementing an evidence-based practice. First, she struggled to incorporate all the components “by Wednesday she was no longer passing out tokens and ran out of reinforcers”. Second, her belief in the likelihood that the intervention will benefit a student quickly faded Ms. Williams felt she created a monster “to get kids to do what they should be doing anyway”. Third, the cost and the time to set up her TE may be a barrier that causes her to quickly abandon the practice before she gets it figured out.

However, a token economy is one of the most proactive and effective behavioral interventions for improving school behavior (Higgins, Williams, & McLaughlin, 2001). TE does not have to be difficult to implement all the steps and the time; costs can be minimized to some upfront investment, and some creative thinking about backup reinforcers (see Table B-14) back up reinforcers that do not cost a dime. The most difficult challenge to address in teacher adoption of an evidence-based practice is their belief system about the practice. In figure A-5 we

provide the steps for accurate implementation, in table 14 we provide some creative suggestions for a low-cost, no-cost system. In this next section, we will walk through some common misunderstandings.

A history of use in schools is a breeding ground for myths about a practice. Things that have been around, may have been tried and abandon and there may be resistance to implementing something folks have “heard of” or “used to do”. Myths are dangerous because they sound credible, and may be based in partial truth (i.e. “I spent 50 dollars on prizes and I can’t afford to keep doing that”). The following myths are derived the literature and from my 20 years of experiences in school settings and informal conversations with teachers, administrators, and college professors and the responses to these myths are based on two empirical studies (Soares, 2011a; Soares 2011b).

Myths

Myths surrounding TE include beliefs that: (a) rewards decrease appropriate classroom behavior (Kohn, 1999; Lepper & Greene, 1978), (b) TE systems are a form of bribery (Kohn, 1999), (c) TE systems are complex (Miltenberg, 2007), (d) TE systems are for special education students only, (e) TE systems are good for young children only, and (f) TE systems are only used for rewarding appropriate behavior not for increasing academic readiness behaviors (i.e. on task to complete assignment, work completion). Myths generally come from partial truths. Any strategy can be implemented in such a way as to be less effective than intended. Additionally, pitfalls may have led to these myths. Pitfalls surrounding TE include:

(a) satiation (Green, Sternberg, & Lepper, 1976) (b) inconsistency and (c) overreliance on punishment (Doty, McInnis, & Paul, 1974). Examples on how to avoid the pitfalls are given.

Rewards Decrease Appropriate Classroom Behavior

One myth is that TE and positive reinforcement decrease internal motivation (Kohn, 1999) thus decreasing appropriate classroom behavior. This theory is derived from studies where University students who are performing well learn to expect reinforcement for academic performance (Deci, 1971) and no longer perform without it. One of the problems with this argument is published research focused on the motivation effect rather than performance. Simply stated, this theory believes the tokens (or reinforcement) diminish the effect of intrinsic motivation. This line of research has evolved over the years to include the improvement of incentives to retain the intrinsic motivation of behavior. In contrast, other lines of research supports that behavior is maintained or increased by reinforcement (Cooper, et al., 2007; Kaplan and Carter, 1995; Skinner, 1957). Reinforcement is given to bring about desirable change and to teach students to take responsibility for behavior. When a behavior is reinforced, the likelihood of the behavior increasing occurs. Soares (2011) found an overall effect size of .78 when the TE was implemented within 79 individual behaviors or children. The 79 experiments reported in 29 studies took place in a variety of settings, with both general and special education students, with a wide range of differences among the TE intervention. The strong effect size in a classroom token system indicates the delivery of a token and the

backup reinforcers are viable options for improving performance (Soares, 2011a; 2011b).

TE Is Bribery

Some feel the delivery of a token to be a bribe. Individuals that believe this feel that giving the child a token is simply persuading the child to do what the adult wants them to do and not teaching the child appropriate behaviors. However, according to Merriam-Webster, a bribe is something used to encourage unacceptable, inappropriate, or possibly illegal behavior. In contrast, however, planned positive reinforcement is very effective in promoting desirable change in student behavior. In a TE system, teachers provide tokens that reinforce the child after the child has demonstrated the predetermined expected behavior. Let me expand –your paycheck is reinforcement for doing your job and bonuses are reinforcements for going above and beyond expectations. Without these reinforcements, how likely is it that you would exhibit the appropriate behavior of showing up at work each day?

TE Is too Complex

Some teachers may create elaborate TE systems that require many complicated procedures. In fact, this concept can also be found in the professional literature. Kaplan and Carter (1995) list eight planning steps that need to occur before implementation of a TE for example: (1) what behavior needs to be displayed to earn tokens, (2) what kind of tokens are you going to use, (3) ratio of student behavior and tokens dispensed, (4) what are the backup reinforcers, (5) who dispenses the tokens, (6) when are tokens given, (7) how are they given, and (8)

when are tokens redeemed for backup reinforcers. That a high number of complex steps is required is a myth; however, TE does not have to be too complex for implementation by teachers in classrooms. Soares (2011) analyzed 5 steps generally seen in TE: (1) visual chart of behaviors, (2) specific token earning times, (3) tokens were identified, (4) visual chart of reinforcers and costs and (5) access to back up reinforcers. Soares (2011) found that TE can be implemented effectively as long as two steps are included; the students earn tokens and cash them in for backup reinforcers. Thus, TE's need not be complex. The necessary components are the earning of tokens and access to back-up reinforcers; thus, a jar filled with marbles is exchanged for a class pizza party.

TE Is for Special Education Settings Only

Another myth or misunderstanding is that TE is only used in special education settings. Some teachers believe that TE systems can only be implemented with small number of students in a very specialized environment. However, TE's are not a behavior strategy for just special education settings; in fact, TE can be flexible and easily adapted to any teaching setting or style. Soares (2011 b) found that TE produced strong effects in both general education and special education settings ($M=.86$ and $.84$, respectively). Teachers implementing TE in a general education classroom with large groups of students can reinforce the group each time there is one person that demonstrates the desired behavior. To adjust for the setting and the needs of the class TE can be as simple and flexible in design as needed.

TE Is only for Young Children Only

TE only works for young children is yet another myth. TE are often thought of as cute charts such as Figure A-2 You're Buzz'in. Some find it hard to conceptualize the same concept in such a way that would be motivating for older children. However, research indicates that the flexibility of TE extends to varying age groups (Soares, 2011). Soares (2011) found TE produced moderate to strong effects ($M=.71$) with ages ranging from 3 to 5 and strong effects ($M=.90$) with children ages 6 to 15. Younger students (age 5 or less) produced moderate to strong effect which is linked to the immediate reinforcement of the token (Filcheck, McNeil, Greco, & Bernard, 2004). Teachers can implement a TE with older children by using a money system with age appropriate reinforcers. Tokens could be represented by pretend money and exchanged for backup reinforcers such as computer time, free homework pass, free answer on a test, etc.

TE Is Only Good for Behavior Issues

The sixth myth surrounding TE is only used to increase appropriate behavior. Some believe that only appropriate behaviors such as raising a hand to get teachers attention can be addressed with a TE. However, both behavior goals and academic readiness goals can be reinforced and changed through the use of a TE. Soares (2011) found that TE produced moderate to strong effects with both behavior and academic goals. TE was slightly more effective with behavior goals ($M=.88$) than with academic goals ($M=.75$). The TE effectively increased behaviors such as compliance to instructions, improved hand raising to get the teachers attention, and

on-task behavior (Kazdin & Mascitelli, 1980; Musser, Bray, Kehle, & Jenson, 2001). In addition, the use of TE increased academic readiness behaviors such as, increasing duration of work time, and completing assignments (Kilmas & Mclaughlin, 2007; Salend, Tintle, & Balber, 1988). Teachers can increase academic readiness behaviors by providing a token each time the student demonstrates an academic readiness skill such as working for completing assignment.

These myths may be a part of the reason that only 30 percent of teachers report using a TE and of those whom do, they use it sparingly. In addition to the myths, which need debunking, there are pitfalls, which need avoiding. These pitfalls can take an effective intervention and render it ineffective.

Pitfalls

TE Creates Reward Satiation

Satiation happens when a reinforcer is no longer effective. Satiation can also occur if too much reinforcement is being delivered. This occurs if a student has unlimited access to the reinforcers (i.e. stickers – if a child is covered with stickers the stickers will eventually lose impact). Soda may not be a strong reinforcer to a student who has unlimited access to it. To avoid satiation teachers can do a few things. First, provide a menu of items for back-up reinforcers so that the student can select items that he or she desires. Second, frequently change the backup menu keeps the child more motivated in earning tokens and decreases the likelihood of satiation. Third, teachers can use reinforcement in the form of activities, social

opportunities, and learning activities, which tend to be more immune to satiation (Zirpoli and Melloy, 1993).

Inconsistency

Inconsistency occurs when teachers forget to follow their schedule, or miss opportunities to provide tokens. Teachers might forget to give a token to a student when the desired behavior is demonstrated or teachers give the students who are physically closest to them in proximity give more tokens than those in the back of the classroom. In actuality, reinforcement should be systematically planned in the development of the TE systems creating high likelihood of consistency across time and students. While consistency is important, teachers should fade the schedule of reinforcement through the delivery of tokens. Teachers will want to issue more tokens in the beginning to get buy in from the children. The changing of requirements promotes continual improvement in behavior or performance by the students while fading concrete reinforcers.

Overreliance on Punishment

Response cost generally refers to a “fine” or the removal of reinforcers (e.g., points, tokens, money, etc.) from the child, and is issued upon the display of an inappropriate behavior (Burchard, 1967; Siegel, Lenske, and Boren, 1969; Weiner, 1962). Response cost is a procedure within a TE where the token is removed for inappropriate behavior. This can be easy to overdo leaving you with a punitive system instead of a positive one. One way to avoid the use of punishment is to adopt a “time out from reinforcement” where a student can no longer earn a token for a

short period of time, but none are taken away. If you use response cost, avoid taking away more tokens than are given. Be clear on procedures for removing and reinstating tokens (or you might escalate the behavior). Soares (2011) found when evaluating 24 studies, 16-used RC and 8-did not. Results indicate both categories had strong effects ($M = .83$ and $.89$ respectively) with the studies who did not use RC having slightly stronger effects. Therefore, RC is not a feature that has to be implemented within the system for the TE system to be effective (Soares, 2011); however, one rule that might be beneficial is to use response cost sparingly and fairly.

Summary

Research supports many positive attributes about token economies. Token economies are one of the most flexible and effective behavior management strategies that can be used to motivate and reinforce human behavior (Kazdin, 1984). TE systems provide a systematic way for students to access desirable items and/or activities. For example, teachers can give the same reward every time or vary the reward; teachers can provide the reward every time the student demonstrates the desired behavior or at a random rate. These rewards can be issued through the use of a token economy.

CHAPTER V

FINAL THOUGHTS

The purpose of this paper was to present evidence for a TE, address some of the many myths surrounding TE, and provide practical application examples for procedures in easing teacher application. Soares (2011) evaluated twenty-four studies including 79 single effect sizes that 78% of the data comparisons between phases A and B show improvement during phase B. Results indicate that TE is an evidence-based intervention to increase academic readiness behaviors and to decrease inappropriate behaviors. Token economy systems can take on a wide variety of forms. They can range from very simple, short-lived systems to much more complex systems that require the child to work for days or even weeks before earning his reward. TE can be continually adjusted and updated to maximize their effectiveness in all settings with individuals, small groups, class wide or even school wide. While teachers struggle with implementing evidence based intervention that have not previously worked due to them making it more complexity, TE is one of the most efficient and effective interventions to change child behavior.

REFERENCES

- Abramson, J. H. (2010). Programs for epidemiologists—Windows version (WinPepi) [Computer software]. Retrieved from <http://www.brixtonhealth.com/pepi4windows.html>.
- Akin-Little, K. A., & Little, S. G. (2004). Re-examining the overjustification effect: A case study. *Journal of Behavioral Education, 13*, 179-192.
doi:10.1023/B:JOB.0000037628.81867.69
- Alberto, A. A., & Troutman, A. C. (2003). *Applied behavior analysis for teacher* (6th edition). New Jersey: Merrill Prentice Hall.
- Ary, D., & Suen H. K. (1989). *Analyzing qualitative behavioral observational data*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ayllon, T. & Azrin, N. H. (1965). The measurement and reinforcement of behavior of psychotics. *Journal of the Experimental Analysis of Behavior, 8*, 357-383.
- Azrin, N. H., & Holz, W. C. (1966). Punishment. In W. K. Honig (Ed.), *Operant behavior: Areas of research and application* (pp. 380-447). New York: Appleton-Century-Crofts.
- Baine, D. A. (1973). Beyond the token economy. *Mental Retardation Bulletin, 1*(3), 60-64.
- Balcazar, F., Hopkins, F., & Suarez, W. (1985). A critical objective view of performance feedback. *Journal of Organizational Behavior Management, 7*, 65-89. doi:10.1300/J075v07n03_05

- Barkley, R., Hastings, J., Tousel, R. & Tousel, S. (1976). Evaluation of a token system for juvenile delinquents in a residential setting. *Journal of Behavior Therapy and Experimental Psychiatry*, 7, 227-230.
- Bassett, J. E., & Blanchard, E. B. (1977). The effect of the absence of dose supervision on the use of response cost in a prison token economy. *Journal of Applied Behavior Analysis*, 10, 375-379.
- Becker, W. C. Madsen, C. H., Jr., Arnold, C. R., & Thomas, D. R. (1967). The contingent use of teacher attention and praise in reducing classroom behavior problems. *Journal of Special Education*, 1, 287-307.
- Berliner, D.C. (1986). In pursuit of the expert pedagogue. *Educational Researcher*, 15(7), 5-13.
- Berryman, D., O'Brien, P., & Cummins, R. (1983). The development of fluency, stimulus generalization and response maintenance of student academic engaged time in a special school: An educational mainstream preparation procedure validated against normative peer data. *Educational Psychology*, 3, 43-61.
- Bippes, R., McLaughlin, T. F., & Williams, R. L. (1986). A classroom token system in a detention center: Effects for academic and social behavior. *Journal for Remedial Education and Counseling*, 2, 126-132.
- Bregman, J.D., Zager, D. & Gerdtz, J. (2005). Behavioral interventions. In D.J. Cohen and F.R. Volkmar (eds.), *Handbook of autism and pervasive*

developmental disorders, 3rd Ed., (pp. 897-924). New York: John Wiley & Sons, Inc.

- Broughton, S. F., & Lahey, B. B. (1978). Direct and collateral effects of positive reinforcement, response-cost, and mixed contingencies for academic behavior. *Journal of School Psychology, 16*, 126-136. doi:10.1016/0022-4405(78)90051-1
- Browsers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education, 16*(2), 239-253.
- Brumfield, B. D., & Roberts, M. W. (1998). A comparison of two measurements of child compliance with normal preschool children. *Journal of Clinical Child Psychology, 27*, 109-116.
- Burchard, J. D. (1967). Systematic socialization: A programmed environment for the habilitation of anti-social retardates. *Psychological Record, 17*, 461-476.
- Burchard, J. D. & Barrera, F. (1972). An analysis of timeout and response cost in a programmed environment. *Journal of Applied Behavior Analysis, 5*, 271-282.
- Buysse, R.T., Kratochwill, T.R., & Elliott, S.N. (1995). Meta-analysis for single-case consultation outcomes: Applications to research and practice. *Journal of School Psychology, 33*, 269–85.
- Byrt, T., Bishop, J., & Carlin, J.B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology, 46*, 423-429. doi:10.1016/0895-4356(93)90018-V

- Carey, R.G., Mosk, M.D. & Kranchuck, K.B. (1981). Employing overcorrection as a cost contingency in a token economy. *Behavior Research of Severe Developmental Disabilities*, 2, 1016-1025.
- Carlson, C. L., Pelham, W. E., Milich, R., & Dixon, J. (1992). Single and combined effects of methylphenidate and behavior therapy on the classroom performance of children with attention-deficit hyperactivity disorder. *Journal of Abnormal Child Psychology*, 20, 213-232.
doi:10.1007/BF00916549
- Carr, E. G., Taylor, J. C., & Robinson, S. (1991). The effects of severe behavior problems in children on the teaching behavior of adults. *Journal of Applied Behavior Analysis*, 24, 523–535. doi:10.1901/jaba.1991.24-523
- Cavalier, A. R., Ferretti, R. P., & Hodges, A. E. (1997). Self-management within a classroom token economy for students with learning disabilities. *Research in Developmental Disabilities*, 18, 167-178. doi:10.1016/S0891-4222(96)00045-5
- Center, D. B., & Wascom, A. (1984). Transfer of reinforcers: A procedure for enhancing response cost. *Educational and Psychological Research*, 4 (1), 19-27.
- Chafouleas, S.M., Riley-Tillman, T.C., & Christ, T.J. (2009). Direct behavior rating (DBR): An emerging method for assessing social behavior within a tiered intervention system. *Assessment for Effective Intervention*, 34, 195-200.
doi:10.1177/1534508409340391

- Christensen, L., Young, K. R., & Marchant, M. (2004). The effects of a peer-mediated positive behavior support program on socially appropriate classroom behavior. *Education & Treatment of Children, 27*, 199-234.
- Center for Effective Collaboration and Practice (2005). *Classroom interventions for children with attention deficit disorder*. Retrieved from Center for Effective Collaboration and Practice website: <http://cecp.air.org/familybriefs>
- Comaty, J.E., M. Stasio & Advokat, C. (2001). Analysis of outcome variables of a token economy (TE) treatment program in a state psychiatric hospital. *Research in Developmental Disabilities, 22*: 233-253, 2001.
- Conyers, C., Miltenberger, R.G., Gubin, A., Barenz, R., Jurgens, M., Sailer, A., Haugen, M., & Kopp, B. (2004). A comparison of response cost and differential reinforcement of other behavior to reduce disruptive behavior in a preschool classroom. *Journal of Applied Behavior Analysis, 37*(3), 411-415.
- Cooper, H.M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of Educational Research, 52*, 291-302.
- Cooper, H. M., & Hedges, L. V. (Eds.) (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (1987). *Applied behavior analysis*. Columbus, OH: Merrill.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson.

- Cotler, S. B., Applegate, G., King, L. W., & Kristal, S. (1972). Establishing a token economy program in a state hospital classroom: A lesson in training student and teacher. *Behavior Therapy, 209*, 214-17.
- Crowther, J. H., Bond, L. A., & Rolf, J. E. (1981). The incidence, prevalence, and severity of behavior disorders among pre-school aged children in day care. *Journal of Abnormal Child Psychology, 9*, 23-42.
- Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18*, 105–115.
- De Martini-Scully, D., Bray, M. A., & Kehle, T. J. (2000). A packaged intervention to reduce disruptive behaviors in general education students. *Psychology in the Schools, 37*, 149-156. doi:10.1002/(SICI)1520-6807(200003)37:2<149::AID-PITS6>3.0.CO;2-K
- Dickerson, F. B., Tenhual, W. N., & Green-Paden, L. D. (2005). The token economy for schizophrenia: Review of the literature and recommendations for future research. *Schizophrenia Research, 75*, 405-416.
doi:10.1016/j.schres.2004.08.026
- Doty, D.W., McInnis, T., & Paul, G.L. (1974). Remediation of negative side effects of an on-going response cost system with chronic mental patients. *Journal of Applied Behavior Analysis, 7*, 191-198.
- DuPaul, G.J. & Eckert, T.L. (1997). The effects of school-based interventions for attention deficit/hyperactivity disorder: A meta-analysis. *School Psychology Review, 26*, 5-27.

- DuPaul, G. J. & Weyandt, L. L. (2006). School-based interventions for children and adolescents with attention deficit/hyperactivity disorder: Enhancing academic and behavioral outcomes. *Education and Treatment of Children, 29*(2), 341-358.
- Eber, L. Sugai, G. Smith, C. & Scott, T.M. (2002). Wraparound and positive behavioral interventions and supports in the schools. *Journal of Emotional and Behavioral Disorders, 10*, 136-173.
- Emmer , E.T. & Stough, L.M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist, 36*(2), 103-112.
- Epstein, M., Atkins, M., Cullinan, D., Kutash, K., & Weaver, R. (2008). *Reducing behavior problems in the elementary school classroom*. Retrieved from What Works Clearinghouse
http://ies.ed.gov/ncee/wwc/pdf/practiceguides/behavior_pg_092308.pdf
- Espin, C.A., & Yell, M.L. (1994). Critical indicators of effective teaching for preservice teachers: Relationships between teaching behaviors and ratings of effectiveness. *Teacher Education and Special Education, 17*, 154-169.
- Feindler, E.L., Marriott, S., & Iwata, M. (1984). Group anger control training for junior high school delinquents. *Cognitive Therapy and Research, 8*, 299-311.
doi:10.1007/BF01173000

- Ferguson, C. J. (2009). Is psychological research really as good as medical research? Effect size comparisons between psychology and medicine. *Review of General Psychology, 13*(2), 130-136.
- Filcheck, H. A., McNeil, C. B., Greco, L. A., & Bernard, R. S. (2004). Using a whole-class token economy and coaching of teacher skills in a preschool classroom to manage disruptive behavior. *Psychology in the Schools, 41*, 351-361. doi:10.1002/pits.10168
- Forness, S.R. & MacMillan, D.L. (1972). Reinforcement overkill: Implications for education of the retarded. *Special Education, 6*, 220-230.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2006). *Educational research: An introduction* (8th ed.). Boston: Pearson Allyn & Bacon. ISBN: 0205488498
- Gaughan, E. & Axelrod, S. (1989). Behavior and achievement relations with emotionally disturbed children: an applied study. *Psychology in the Schools, 26*(1), 89-99.
- Gest, S., & Gest, J. (2005). Reading tutoring for students at academic and behavioral risk: Effects on time-on-task in the classroom. *Education & Treatment of Children, 28*, 25-48.
- Glass, G. V (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3-8.
- Glass, G.V., McGaw, B. & Smith, M.L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

- Good, T. (1983). Classroom research: A decade of progress. *Educational Psychologist, 18*, 127-144.
- Graziano, C. (2005, *February/March*). School's out. *Edutopia*,, 40-44.
- Greene, D., Sternberg, B., & Lepper, M.R. (1976). Overjustification effect of token economy. *Journal of Personality and Social Psychology, 34*(6), 1219-1234.
doi: 10.1037/0022-3514.34.6.1219
- Gresham, F.M. (1979). Comparison of response cost and timeout in a special education setting. *Journal of Special Education, 13*, 199-208.
doi:10.1177/002246697901300211
- Hartmann, D. P., Barrios, B. A., & Wood, D. D. (2004). Principles of behavioral observation. In S. N. Haynes & E. M. Heiby (Eds.), *Comprehensive handbook of psychological assessment, Vol. 3: Behavioral assessment* (pp. 108-127). Hoboken, NJ: John Wiley and Sons.
- Heaton, R. C., & Safer, D. J. (1982). Secondary school outcome following a junior high school behavioral program. *Behavior Therapy, 13*, 226-231.
doi:10.1016/S0005-7894(82)80066-X
- Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Higgins, J.P.T., & Green, S. (eds.)(2011) *Cochrane handbook for systematic reviews of interventions*, Version 5.1.0. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

- Higgins, J. W., Williams, R. L., & McLaughlin, T. F. (2001). The effects of a token economy employing instructional consequences for a third-grade student with learning disabilities: A data-based case study. *Education & Treatment of Children, 24*, 99-106.
- Himle, M. B., Woods, D. W., & Bunaciu, L. (2008). Evaluating the role of contingency in differentially reinforced tic suppression. *Journal of Applied Behavior Analysis, 41*, 285-289.
- Hintze, J. (2004). NCSS 2004 [computer software]. Retrieved from <http://www.ncss.com>
- Holmbeck, G. N. (1997). Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology, 4*, 599–610. doi:10.1037/0022-006X.65.4.599
- Holmes, CT (1984). Effect size estimation in meta-analysis. *Journal of Experimental Education, 52*, 106-109.
- Holt, M. M., Hobbs, T. R. & Hankins, R. (1976). The effects of token reinforcement on delinquents' classroom behavior. *Psychology in the Schools, 13* (3), 341-347.
- Horner, R. H., Carr, E. G., Halle, J., McGee, J., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165-179.

- Hundert, J., & Batstone, D. (1978). A practical procedure to maintain pupils' accurate self-rating in a classroom token program. *Behavior Modification*, 2(1), 93-112. doi: 10.1177/014544557821006
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd edition). Thousand Oaks, CA: Sage
- Individuals With Disabilities Education Improvement Act of 2004, Pub. L. No. 108–446, 118 Stat. 2647 [Amending 20 U.S.C. § 1462 et seq.].
- Ingersoll, R. (2003). The teacher shortage: Myth or reality? *Educational Horizons*, 81, 146-152.
- Johns, B. (2000). Reaching them through teaching them: Curriculum and instruction for students with emotional/behavioral disorders. *Behavioral Disorders*, 10, 3-6.
- Johnson, C.R., Handen, B.L., Lubetsky, M.J., & Sacco, K.A. (1994). Efficacy of methylphenidate and behavioral intervention on classroom behavior in children with ADHD and mental retardation. *Behavior Modification*, 18 (4), 470-87.
- Kaplan, J. S., & Carter, J. (1995). *Beyond behavior modification: A cognitive-behavioral approach to behavior management in the school* (3rd ed.). Austin, TX: Pro-Ed.
- Kaufman, K.F. & O'Leary, K.D. (1972). Reward, cost, and self-evaluation procedures for disruptive adolescents in a psychiatric hospital school. *Journal of Applied Behavior Analysis*, 5, 299-309.

- Kavale, K.A. (1984). Potential advantages of the meta-analysis technique for research in special education. *Journal of Special Education, 18*, 61-72.
doi:10.1177/002246698401800107
- Kavale, K.A. (1998). Some methods are more effective than others. *Intervention in School and Clinic, 33*, 195-200. doi:10.1177/105345129803300401
- Kavale, K.A. (2001). Meta-analysis: A primer. *Exceptionality, 9*, 177-183.
doi:10.1207/S15327035EX0904_2
- Kavale, K.A. (2007). Quantitative research synthesis: Meta-analysis of research on meeting special educational needs. In L. Florian (Ed.), *The Sage handbook of special education* (pp. 207–221). Thousand Oaks, CA: Sage Publications.
- Kazdin, A.E. (1971). The effect of response cost in suppressing behavior in a pre-psychotic retardate. *Journal of Behavior Therapy and Experimental Psychiatry, 2*, 137-140. doi:10.1016/0005-7916(71)90029-2
- Kazdin, A.E. (1972). Response cost: The removal of conditioned reinforcers for therapeutic change. *Behavior Therapy, 3*, 533-546.
- Kazdin, A.E. (1973). The effect of vicarious reinforcement on attentive behavior in the classroom. *Journal of Applied Behavior Analysis, 6*, 71-78.
- Kazdin, A.E. (1975). *Behavior modification in applied settings*. Homewood, IL: Dorsey.
- Kazdin, A.E. (1977). *The token economy*. New York: Plenum Press.
- Kazdin, A.E. (1980). *Behavior modification in applied settings* (rev. ed.). Homewood, IL: Dorsey.

- Kazdin, A.E. (1982). *Single-case research designs: Methods for clinical and applied settings*. New York, NY: Oxford University Press.
- Kazdin, A.E. (1984). Statistical analysis for single-case experimental designs. In D.H. Barlow & M. Herses (eds.), *Single case experimental designs: Strategies for studying behavior change* (2nd ed.)(pp. 285-324). New York: Pergamon Press.
- Kazdin, A. E., & Bootzin, R. R. (1972). The token economy: An evaluative review. *Journal of Applied Behavior Analysis, 5*, 343-372. doi:10.1901/jaba.1972.5-343
- Kazdin, A.E., & Geesey, S. (1980). Enhancing classroom attentiveness by preselection of back reinforcers in a token economy. *Behavior Modification, 4*, 98-114.
- Kazdin, A. E., & Mascitelli, S. (1980). The opportunity to earn oneself off a token system as a reinforcer for attentive behavior. *Behavior Therapy, 11*, 68-78.
- Kerr, M. M., & Nelson, C. M. (1998). *Strategies for managing behavior problems in the classroom* (3rd ed.). Upper Saddle River, NJ: Merrill.
- Klimas, A. & McLaughlin, T. F. (2007). The effects of a token economy system to improve social and academic behavior with a rural primary aged child with disabilities. *International Journal of Special Education, 22* (3).
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary

feedback intervention theory. *Psychological Bulletin*, *119*, 254–284.

doi:10.1037/0033-2909.119.2.254

Kohn, A. (1999). *Punished by rewards: The trouble with gold stars, incentive plans, A's, praise, and other bribes*. New York: Houghton Mifflin.

Kratochwill, T. R. (1992). Single case research design and analysis: An overview.

In T. R. Kratochwill and J. R. Levin (Eds.), *Single-case research design and analysis* (pp. 1-14). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L.,

Rindskopf, D. M & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from What Works Clearinghouse

http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Kratochwill, T. R., & Shernoff, E. S. (2004). Evidence-based practice: Promoting evidence-based interventions in school psychology. *School Psychology Quarterly*, *19*, 342-364. doi:10.1521/scpq.19.4.342.53499

Kratochwill, T. R., & Stoiber, K. C. (2002). Evidence-based interventions in school psychology: Conceptual foundations of the procedural and coding manual of Division 16 and the Society for the Study of School Psychology Task Force. *School Psychology Quarterly*, *17*, 341-389.

doi:10.1521/scpq.17.4.341.20872

Landrum, T.J. Cook, B.G., Tankersley, M., & Fitzgeralds, S. (2007). Teacher perceptions of the useability of intervention information from personal versus data-based sources. *Education and Treatment of Children*, *30*(4), 27-42.

- Landrum, T. J., Tankersley, M., & Kauffman, J. M. (2003). What is special about special education for students with emotional or behavioral disorders? *The Journal of Special Education, 37*, 148–156.
doi:10.1177/00224669030370030401
- Lannie, A. L. & McCurdy, B. L. (2007). Preventing disruptive behavior in the urban classroom: Effects of the Good Behavior Game on student and teacher behavior. *Education and Treatment of Children, 30*, 85-98.
- Latham, G. P. & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes, 50*, 212-247.
doi:10.1016/0749-5978(91)90021-K
- Lecomte, T., Liberman, R.P. & Wallace, C. (2000). Identifying and using reinforcers to enhance treatment of persons with serious mental illness. *Psychiatric Services, 51*,1312-1314.
- Lepper, M. R. & D. Greene (Eds.) (1978). *The hidden costs of reward: New perspectives on psychology of human motivation*. Hillsdale, NJ: Erlbaum.
- Lewis, T. J., Hudson, S., Richter, M., & Johnson, N. (2004). Scientifically supported practices in EBD: A proposed approach and brief review of current practices. *Behavioral Disorders, 29*, 247-259.
- Lhospital, A. S., & Gregory, A. (2009). Changes in teacher stress through participation in pre-referral intervention teams. *Psychology in the Schools, 46*, 1098–1112. doi: 10.1002/pits.20455

- Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Applied Social Research Methods Series (Vol. 49). Thousand Oaks, CA: Sage Publications.
- Ludwick, R., Dieckman, B.C. Herdtner, S., Dugan, M., & Roche, M. (1998). Documenting the scholarship of clinical teaching through peer review. *Nurse Education, 23* (6): 17–20. doi:10.1097/00006223-199811000-00008
- Maccini, P., & Gagnon, J. C. (2006). Instructional adaptations and assessment accommodations by secondary special and general educators. *Exceptional Children, 72* (2), 217-234.
- Madsen, C.H., Becker, W.C., & Thomas, D.R. (1968). Rules, praise, and ignoring: Elements of elementary classroom control. *Journal of Applied Behavior Analysis, 1*, 139-150.
- Maglio, C. & McLaughlin, T.F. (1981). Effects of a token reinforcement system and teacher attention in reducing inappropriate verbalization with a junior high school student. *Corrective and Social Psychiatry, 27*, 140-145.
- Martin, G., & Pear, J. (2003). *Behavior modification: What it is and how to do it?* (7th ed.). Upper Saddle River, NJ: Simon & Schuster.
- Mathews, B., McLaughlin, T. F., & Hunsaker, D. (1980). Effects of teacher attention and activity reinforcers on on-task behavior. *Education and Treatment of Children, 3*, 13-19.
- Matson, J. L., & Boisjoli, J. A. (2009). The token economy for children with intellectual disability and or autism: A review. *Research on Developmental Disabilities, 30*, 240-248. doi:10.1016/j.ridd.2008.04.001

- Mayhew, G. L., & Anderson, J. (1980). Delayed and immediate reinforcement: Retarded adolescents in an educational setting. *Behavior Modification, 4*, 527-545.
- McDonald, A. (2002). Best practices for at risk children. Retrieved from: <http://www.sanmarcos.net/ana/bestpractices.html>
- McGoey, K.E. & DuPaul, G.J. (2000). Token reinforcement and response cost procedures: Reducing the disruptive behavior of preschool children. *School Psychology Quarterly, 15*, 330-343.
- McLaughlin, T.F., & Williams, R.L. (1988). The token economy in the classroom. In J.C. Witt, S.N. Elliott, & F.M. Gresham (Eds.). *Handbook of behavior therapy in education* (pp. 469-487). New York: Plenum.
- Miltenberger, R. (2000). Behavioral skills training to remediate deviant social behavior of an adolescent in residential treatment. *Cognitive and Behavioral Practice, 7*, 236-238. doi:10.1016/S1077-7229(00)80039-5
- Mottram, L. M., Bray, M. A., Kehle, T. J., Broudy, M., & Jenson, W. R. (2002). A classroom-based intervention to reduce disruptive behaviors. *Journal of Applied School Psychology, 19*(1), 65-74.
- Musser, E. H., Bray, M. A., & Kehle, T. J. (2001). Reducing disruptive behaviors in students with serious emotional disturbance. *School Psychology Review, 30*(2), 294-304.
- No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

- Odom, S.L., Brown, W. H., Frey, T., Karasu, N., Smith-Carter, L., & Strain, P. S. (2003). Evidence-based practices for young children with autism: Evidence from single subject design research. *Focus on Autism, 18*, 176-181.
- Odom, S.L., McConnell, S.R., McEvoy, M.A., Peterson, C., Ostrosky, M., Chandler, L.K., Spicuzza, R.J., Skellenger, A., Creighton, M., & Favazza, P.C. (1999). Relative effects of interventions supporting the social competence of young children with disabilities. *Topics in Early Childhood Special Education, 19*, 75-91.
- O'Leary, K.D., Becker, W.C., Evans, M.B., & Saudargas, R.A. (1969). A token reinforcement program in a public school: A replication. *Journal of Applied Behavior Analysis, 2*, 3-13.
- O'Leary, K. D. & Drabman, R. (1971). Token reinforcement programs in the classroom: A review. *Psychological Bulletin, 75*, 379-398.
- O'Leary, K. D., & O'Leary, S. G. (1977). *Classroom management: The successful use of behavior modification* (2nd edition.). New York: Pergamon Press.
- Parker, R.I. & Hagan-Burke, S. (2007). Single case research results as clinical outcomes. *The Journal of School Psychology, 45*, 637-653.
doi:10.1016/j.jsp.2007.07.004
- Parker, R.I., Hagan-Burke, S., & Vannest, K.J. (2007). PAND: an alternative to PND. *Journal of Special Education, 40*, 194-204.
doi:10.1177/0022466907040004010

- Parker, R.I., & Vannest, K.J. (2009). An improved effect size for single case research: Non-overlap of all pairs (NAP). *Behavior Therapy*, 40, 357-367.
- Parker, R.I, Vannest, K.J., & Brown, L.M. (2009). The improvement rate difference for single case research. *Exceptional Children*, 75, 135-150.
- Parker, R. I., Vannest, K.J., & Davis, J.L. (2010). Nine non-overlap techniques for single case research. *Behavior Modification*, 35(4), 303-322.
- Parker, R.I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, 35, 303-322. doi:10.1177/0145445511399147
- Parker, R.I., Vannest, K.J., Davis, J.L., & Sauber, S.B. (2011). Combining non-overlap and trend for single case research: Tau-U. *Behavior Therapy*, 42, 284–299. doi: 10.1016/j.beth.2010.08.006.
- Paul, G. L., & Lentz, R. J. (1977). *Psychosocial treatment of chronic mental patients: Milieu versus social-learning programs*. Cambridge: Harvard University Press.
- Phillips, E. L., Phillips, E. A., Fixsen, D. L., & Wolf, M. M. (1971). Achievement place: Modification of the behaviors of pre-delinquent boys within a token economy. *Journal of Applied Behavior Analysis*, 4, 45-50.
doi:10.1901/jaba.1971.4-45
- Rapport, M. D., Murphy, A., & Bailey, J. S. (1980). The effects of a response cost treatment tactic on hyperactive children. *Journal of School Psychology*, 18, 98-111. doi:10.1016/0022-4405(80)90025-4

- Reitman, D., Murphy, M. A. , Hupp, S. D. A. & O'Callaghan, P. M.(2004). Behavior change and perceptions of change: Evaluating the effectiveness of a token economy, *Child & Family Behavior Therapy*, 26(2), 17 - 36. doi: 10.1300/J019v26n02_02
- Rhode, G., Jenson, W.R., & Reavis, H.K. (1993). *The tough kid book: Practical classroom management strategies*. Longmont, CO: Sopris West, Inc.
- Rosen, L. A., Taylor, S. A., O'Leary, S.G., & Sanderson, W. (1990). A survey of classroom management practices. *The Journal of School Psychology*, 28(3), 257-269.
- Rosenberg, M. S. (1986). Maximizing the effectiveness of structured classroom management programs: Implementing rule-review procedures with disruptive and distractible students. *Behavioral Disorders*, 11, 239-248.
- Rosenthal, R. (1991). Parametric measures of effect size. In H. Cooper & L.V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231 - 244). New York: Sage.
- Safer, D. J., Heaton, R. C., & Parker, F. C. (1981). A behavioral program for disruptive junior high school students: Results and follow-up. *Journal of Abnormal Child Psychology*, 9, 483-494. doi:10.1007/BF00917797
- Salend, S.J. & Allen, E.M. (1985). Comparative effects of externally-managed response cost systems on inappropriate classroom behavior. *Journal of School Psychology*, 23, 59-67.

- Salend, S. J., & Lamb, E. A. (1986). Effectiveness of a group-managed interdependent contingency system. *Learning Disability Quarterly, 9*, 268–273.
- Salend, S.J., Tintle, L., and Balber, H. (1988). Effects of a student-managed response cost system on the behavior of two mainstreamed students. *The Elementary School Journal, 89*(1), 89-97.
- Schellenberg, T., Skok, R., & McLaughlin, T. F. (1991). The effects of contingent free time on homework completion in English with high school English students. *Child & Family Behavior Therapy, 13*(3), 1-12.
doi:10.1300/J019v13n03_01
- Schneider, N., & Goldstein, H. (2008). Peer-related social competence interventions for young children with communication and language disorders. In W. H. Brown, S. L. Odom, & S. R. McConnell (Eds.), *Social competence of young children: Risk, disability, and intervention* (pp. 233-252). Baltimore: Paul H. Brookes.
- Scott, T. M., Nelson, C. M., & Liaupsin C. J. (2001). Effective instruction: The forgotten component in preventing school violence. *Education and Treatment of Children, 24*, 309-322.
- Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Research Council, National Academy Press.

- Shriver, M. D., & Allen, K. D. (1997). Defining child noncompliance: An examination of temporal parameters. *Journal of Applied Behavior Analysis, 30*, 173–176.
- Siegel, G. M., Lenske, J., & Broen, P. (1969). Suppression of normal speech disfluencies through response cost. *Journal of Applied Behavior Analysis, 2*, 265-276.
- Simon, S. J., Ayllon, T., & Milan, M. A. (1982). Behavioral compensation: Contrastlike effects in the classroom. *Behavior Modification, 6*, 407-420.
- Skinner, B. F. (1931). The concept of the reflex in the description of behavior. *Journal of General Psychology, 5*, 427-458.
doi:10.1080/00221309.1931.9918416
- Skinner, B.F. (1953). *Science and human behavior*. New York: Macmillan.
- Skinner, B. F. (1957). *Verbal behavior*. New York, NY: Appleton.
- Smith, L.K. & Fowler, S.A. (1984). Positive peer pressure: The effects of peer monitoring on children's disruptive behavior. *Journal of Applied Behavior Analysis, 17* (2), 213-227.
- Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education, 61*, 334-349.
- Sran, S.K. & Borrero, J.C. (2010). Assessing the value of choice in a token system. *Journal of Applied Behavior Analysis, 43*, 553-557.

- Steel, P.D. & Kammeyer-Mueller, J.D. (2002). Comparing meta-analytic moderator estimation techniques under realistic conditions. *Journal of Applied Psychology, 87*, 96-111.
- Stevens, C., Sidener, T. M., Reeve, S. A., & Sidener, D. W. (2011). Effects of behavior-specific and general praise on acquisition of tacts in children with pervasive developmental disorders. *Research in Autism Spectrum Disorders, 5*, 666-669. doi:10.1016/j.rasd.2010.08.003
- Stilitz, I. (2009). A token economy of the early 19th century. *Journal of Applied Behavior Analysis, 42* (4), 925-926.
- Strijbos, J.; Martens, R.; Prins, F., & Jochems, W. (2006). Content analysis: What are they talking about? *Computers & Education, 46*, 29–48.
doi:10.1016/j.compedu.2005.04.002
- Suen, H. K., & Ary, D. (1989). Analyzing quantitative behavioral observation data. Hillsdale, NJ: Erlbaum.
- Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., Scott, T., Liaupsin, C., Sailor, W., Turnbull, A. P., Turnbull H. R., III, Wickham, D., Wilcox, B., & Ruef, M. (2000). Applying positive behavior support and functional behavioral assessment in schools. *Journal of Positive Behavior Interventions, 2*(3), 131–143.
- Sugai, G., & Horner, R.H. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Child and Family Behavior Therapy, 24*, 23-50.

- Sullivan, M. A., & O'Leary, S. G. (1990). Maintenance following reward and cost token programs. *Behavior Therapy, 21*, 139-149.
- Swain, J. C., & McLaughlin, T. F. (1998). The effects of bonus contingencies in a classwide token program on math performance with middle school students with behavior disorders. *Behavioral Interventions, 13*, 11-20.
- Swanson, H. L., & Sachse-Lee, C. (2000). A meta-analysis of single-subject design intervention research for students with learning disabilities. *Journal of Learning Disabilities, 33*, 114-166. doi:10.1177/002221940003300201
- Swiezy, N. B., Matson, J. L., & Box, P. (1992). The good behavior game: A token reinforcement system for preschoolers. *Child and Family Behavior Therapy, 14*, 21-32.
- Tankersley, M., Cook, B. G., & Cook, L. (2008). A preliminary examination to identify the presence of quality indicators in single-subject research. *Education and Treatment of Children, 31*, 523-548. doi:10.1353/etc.0.0027
- Taplin, P. S., & Reid, J. B. (1973). Effects of instructional set and experimenter influence on observer reliability. *Child Development, 44*, 547-554.
- Thalheimer, W., & Cook, S. (2002). *How to calculate effect sizes from published research articles: A simplified methodology*. Retrieved June 15, 2011 from http://work-learning.com/effect_sizes.htm.
- Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. New York: Guilford.

- Thurlow, M. L., Ysseldyke, J. E., Graden, J., & Algozzine, B. (1983). Instructional ecology for students in resource and regular education. *Teacher Education and Special Education, 6*, 248-254.
- Tolan, P., Henry, D., Schoeny, M., & Bass, A. (2008). *Mentoring interventions to affect juvenile delinquency and associated problems*. Oslo, Norway: The Campbell Collaboration.
- 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*, 1-10. doi:10.1002/(SICI)1099-078X(199802)13:1<1::AID-BIN1>3.0.CO;2-Z
- Turkewitz, H., O'Leary, K. D., & Ironsmith, M. (1975). Generalization and maintenance of appropriate behavior through self-control. *Journal of Consulting and Clinical Psychology, 43*, 577-583. doi:10.1037/h0076905
- Ulmer, R. A. (1976). *On the development of a token economy mental hospital treatment program*. Washington, D.C.: Hemisphere Publishing Corp.
- Vannest, K. J., Davis, J. L., Davis, C. R., Mason, B. A., & Burke, M. D. (2010). Effective intervention and measurement with a daily behavior report card: A meta-analysis. *School Psychology Review, 39*, 654-672.
- Vannest, K.J., Hagan-Burke, S. & Parker, R.I. (2006). Teacher time use across programs. *Journal of Educational Review, 29(4)*, 299-306.

- Vannest, K. J., Mahadevan, L., Harvey, K. T., & Mason, B. A. (2009). Educator perceptions of the impact of No Child Left Behind on special populations. *Remedial and Special Education, 30*, 148-159.
doi:10.1177/0741932508315378
- Vannest, K. J., Reynolds, C. R., & Kamphaus, R. (2008). *BASC-2 intervention guide for learning and behavior problems*. Minneapolis, MN: Pearson.
- Vannest, K.J., Soares, D.A., Harrison, J.R., Brown, L., & Parker, R.I. (2010). Changing teacher time. *Preventing School Failure, 54* (2) 86 – 98. DOI: 10.1080/10459880903217739
- Walker, H. M. (2004). Use of evidence-based interventions in schools: Where we've been, where we are, and where we need to go [Commentary]. *School Psychology Review, 33*, 398–407.
- Walker, H. M., Colvin, G., & Ramsey, E. (1995). *Antisocial behavior in school: Strategies and best practices*. Pacific Grove, CA: Brooks/Cole Publishing Company.
- Walker, H. M., & Gresham, F. M. (2003). School-related behavior disorders. In W. Reynolds & G. Miller (Eds.), *Handbook of psychology: Educational psychology* (Vol. 7, pp. 511–530). New York: Wiley.
- Weiner, H. (1962). Some effects of response cost on human operant behavior. *Journal of Experimental Analysis of Behavior, 5*, 201-208.
- Weiss, C.H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.) Prentice Hall.

- Wheeler, J., Freagon, S., & Stern, L. (1985). Teaching a severely handicapped woman self-management skills to decrease ambulation time. *Remedial & Special Education, 6*, 48-50. doi:10.1177/074193258500600110
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54* (8) 594-604.
- Williams, B.F., Williams, R.L., & McLaughlin, T.F. (1991). Classroom procedures for remediating behavior disorders. *Journal of Developmental and Physical Disabilities, 3*, 349-384.
- Witt, J. C, & Elliot, S. N. (1982). The response cost lottery: A time efficient and effective classroom intervention. *Journal of School Psychology, 20*,155-161. doi:10.1016/0022-4405(82)90009-7
- Yap, K., Aldersebaes, I., Railsback, J., Shaughnessy, J., & Speth, T. (2000). *Evaluating whole school reform efforts: A guide for districts and school staff* (2nd edition). Portland, OR: Northwest Regional Educational Laboratory.
- Zirpoli, T. J. (2005). *Behavior Management: Applications for Teachers*. Upper Saddle River, New Jersey: Pearson Prentice Hall.
- Zirpoli, T.J., & Melloy, K.J. (1993). Behavior management: Applications for teachers and parents. New York: MacMillan Publishing Company.
- Zlomke, L. (2003). Token Economies. *The Behavior Analyst Today, 4* (2), 177-193.

APPENDIX A

FIGURES

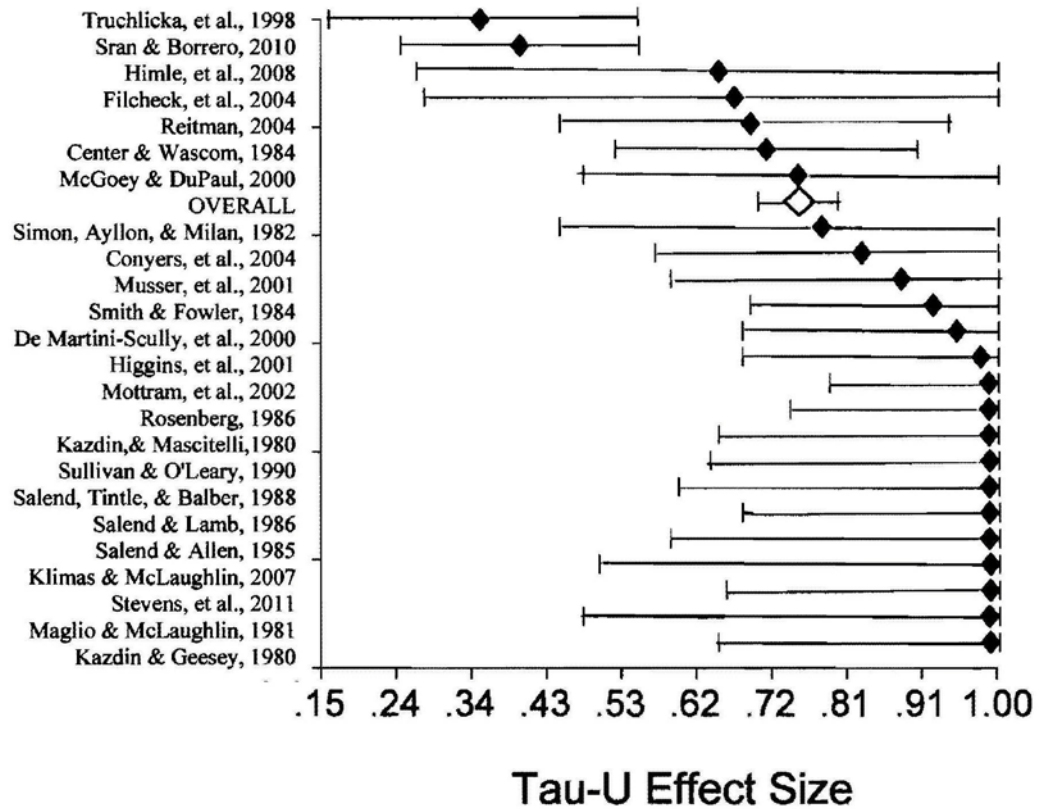
Figure A-1*Forrest Plot of Effect Sizes for 24 Included Studies and the Overall ES*

Figure A-2

Example of You're Buzz'n-Gram: A Token Earning Card

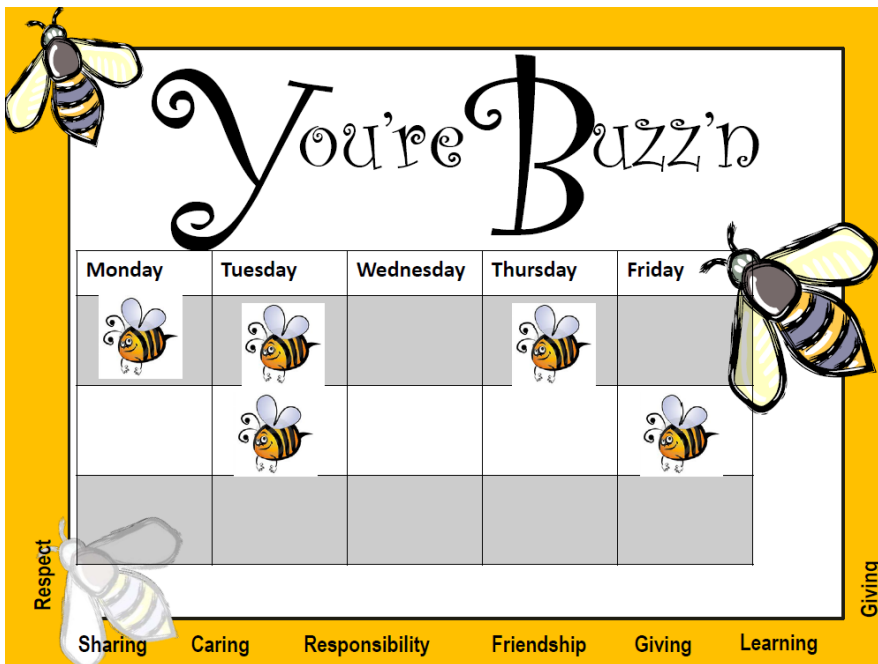
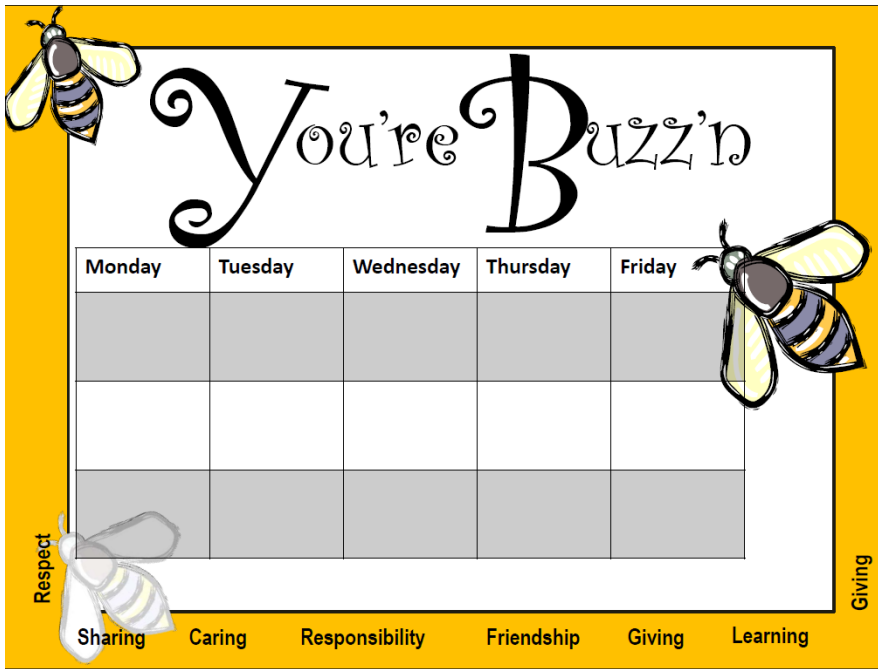


Figure A-3

Visual Chart of Expected Behavior



Figure A-4

Bella's Notebook Visual Chart

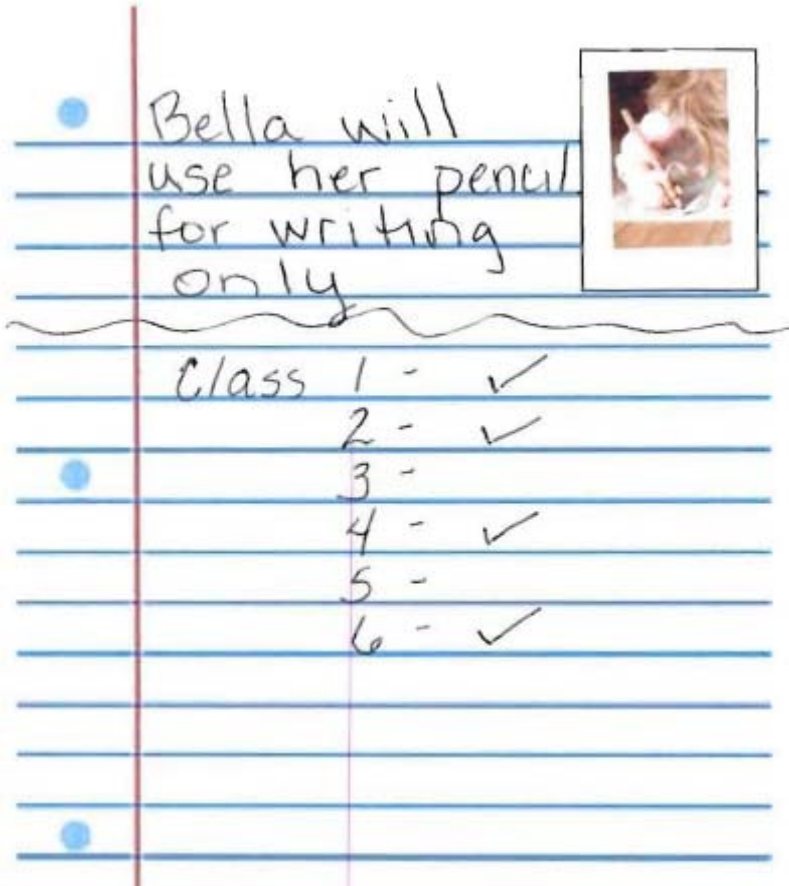
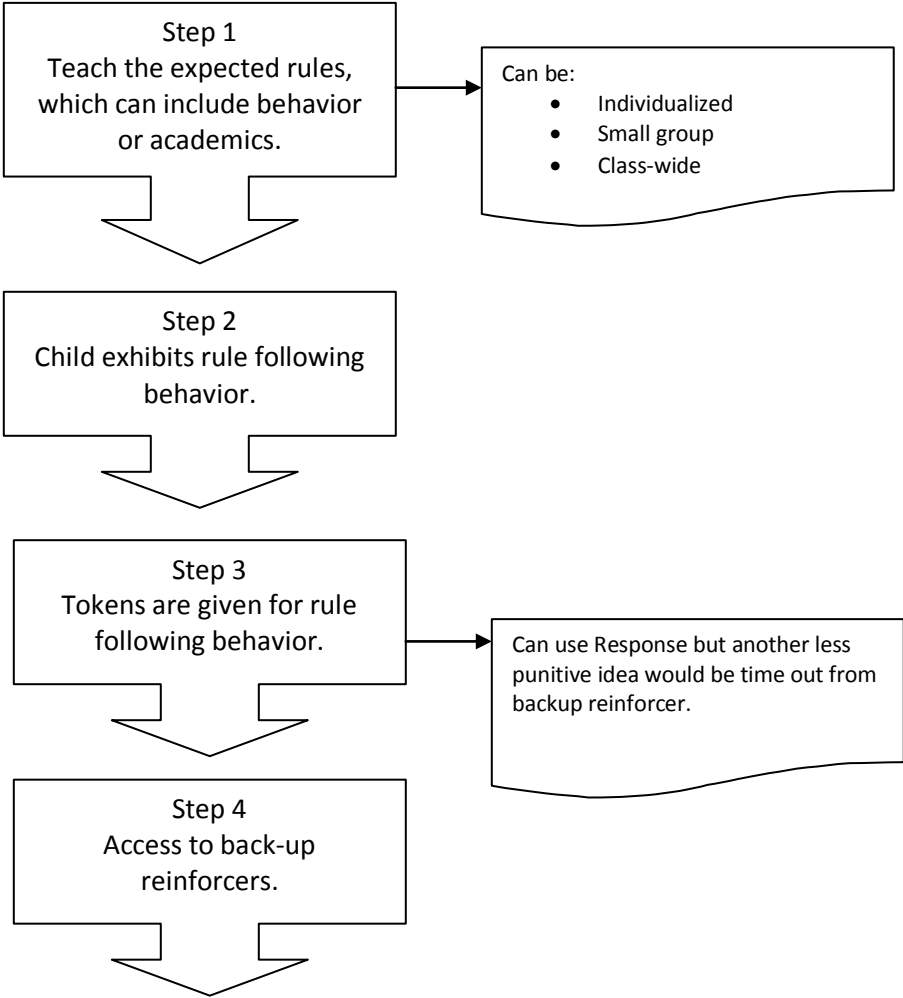


Figure A-5

Simple Steps for a Token Economy



APPENDIX B

TABLES

Table B-1*Summary of 29 Previously Reviewed Articles Included in Two Review Studies*

Review	Studies	# of Participants	Setting	Diagnoses	# of M/F	Target Behaviors
Dickerson, Tenhula, & Green-Paden, 2005	Ayllon & Azrin, 1965	44	Inpatient Illinois State Hospital	Schizophrenia	44 Female	Adaptive Behaviors such as job performance
Dickerson, Tenhula, & Green-Paden, 2005	Schaefer & Martin, 1966	40	Inpatient California State Hospital	Schizophrenia	40 Female	Daily living skills, social interaction, work performance
Dickerson, Tenhula, & Green-Paden, 2005	Marks et al., 1968	22	Inpatient Washington State Hospital	Schizophrenia	22 Male	Social behaviors
Dickerson, Tenhula, & Green-Paden, 2005	Shean & Zeidberg, 1971	42	Inpatient Virginia State Hospital	Psychotic	42 Male	Daily living skills, social interaction, work performance
Dickerson, Tenhula, & Green-Paden, 2005	Gripp & Magaro, 1971	45	Inpatient	Schizophrenia	45 Female	Work performance
Dickerson, Tenhula, & Green-Paden, 2005	Maley et al., 1973	40	Inpatient Virginia State Hospital	Schizophrenia	40 Female	Grooming, Cooperation with others, Appropriate Behaviors
Dickerson, Tenhula, & Green-Paden, 2005	Paul & Lentz, 1977	84	Inpatient Illinois State Hospital	Psychotic Disorder	Did not specify	Social Skills and Educational Activities
Dickerson, Tenhula, & Green-Paden, 2005	Baker et al., 1977	18	Inpatient Hospital in the UK	Schizophrenia	Did not specify	Adaptive Behaviors

Table B-1 continued

Review	Studies	# of Participants	Setting	Diagnoses	# of M/F	Target Behaviors
Dickerson, Tenhula, & Green-Paden, 2005	Elliott et al., 1979	18	Inpatient Hospital in the UK	Schizophrenia	18 Male	Hygiene, work habits, social interaction
Dickerson, Tenhula, & Green-Paden, 2005	Nelson & Cone, 1979	16	Inpatient West Virginia Hospital	Psychotic Disorder	16 Male	Personal hygiene, work performance and social skills
Dickerson, Tenhula, & Green-Paden, 2005	Rimmerman et al., 1991	617	Outpatient	Mental Illness residing in Adult Homes	Did not specify	did not specify
Dickerson, Tenhula, & Green-Paden, 2005	Lippman & Motta, 1993	36	Outpatient	Schizophrenia	Did not specify	Daily living skills, Work performance and Social Skills
Dickerson, Tenhula, & Green-Paden, 2005	Li & Wang, 1994	52	Inpatient Hospital in China	Schizophrenia	Did not specify	Life Skills
Matson & Boisjoli, 2009	Birnbrauer et al. 1965	17	School	Mild Intellectual Disability	Did not specify	Rates of Studying
Matson & Boisjoli, 2009	Wolf et al., 1968	16	School	Borderline to Average Intelligence	Did not specify	Academic Achievement and Report Card Grades
Matson & Boisjoli, 2009	Staats et al., 1970	24	School	Mild Intellectual Disability	Did not specify	Improved attention and work behavior
Matson & Boisjoli, 2009	Cotler et al., 1972	14	State Hospital Classroom	Borderline to Average Intelligence	14 male	Class participation and quiet working
Matson & Boisjoli, 2009	Jones & Kazdin, 1975	4	School	Mild to Borderline ID	Did not specify	Attentive, in seat
Matson & Boisjoli, 2009	Nay & Legum, 1976	11	School	Mild Intellectual Disability	Did not specify	Out of Seat, inappropriate verbalizations

Table B-1 continued

Review	Studies	# of Participants	Setting	Diagnoses	# of M/F	Target Behaviors
Matson & Boisjoli, 2009	Fox & Roseen, 1977	1	Home	Normal IQ, PKU	1 Male	Daily living skills
Matson & Boisjoli, 2009	Johnson et al., 1984	42	School	Mild Intellectual Disability	Did not specify	On task during testing
Matson & Boisjoli, 2009	Pruneti et al., 1989	20	School	Children with Head Injuries	Did not specify	Maladaptive Behaviors
Matson & Boisjoli, 2009	Steeves et al., 1970	2	Developmental Center	Autism	2 Males	Verbal and printing tasks
Matson & Boisjoli, 2009	Hung, 1977	4	Summer Camp	Autism	Did not specify	Spontaneous Questions
Matson & Boisjoli, 2009	Handen et al., 1984	1	Group Home	Autism	1 Male	Low levels of repetitive speech
Matson & Boisjoli, 2009	Odom et al., 1985	3	School	Autism	Did not specify	Social Initiation
Matson & Boisjoli, 2009	McDonald & Hemmes, 2003	1	School	Autism	Did not specify	Interactions with adults
Matson & Boisjoli, 2009	Boscoe & Byrne, 2003	1	School	PDD	Did not specify	Food refusal
Matson & Boisjoli, 2009	Tarbox et al., 2006	1	School	Autism	Did not specify	Attending to task

*Note: Dickerson, Tenhula, & Green-Paden, 2005 included 13 studies; Maston & Boisjoli, 2009 included 16 studies.

Table B-2*Protocol for Design Strength*

Design	# of Baselines/Phases	Score
Multiple Baseline Design	2 baselines	1 (weak)
	3 baselines	2 (medium)
	4 baselines or 3 baselines and a control	3 (strong)
Multiple Baseline Non-concurrent Design	Any number of baselines	1 (weak)
Single Baseline Design Reversal	ABA	1 (Weak)
	ABAB	2 (medium)
	ABAB+ any additional phases	3 (strong)
Changing Criterion Design		1 (weak)
AB Design		1 (weak)

**Note:* Design strength is the internal validity of the design. The number of subjects is important for measurement precision but not for design strength or internal validity. Additionally, if a C phase is added it will not change the design strength but will answer additional questions. If an AB design adds a maintenance phase it does not increase strength.

Table B-3*Summary of Participant and Setting Characteristics*

Study Characteristic		Number	Percent (%)
	Included studies	24	
	Total number of students	84	
	Number of ES	79	
Age	3-5	29	35
	6-15	55	65
Educational Status	Special Education	17	71
	General Education	7	29
Setting of Intervention	Special Education	12	50
	General Education	12	50
Goal	Academic	5	21
	Behavior	19	79

Table B-4*Cross Tabulation Report from NCSS*

		Rater 1				
Rater 2	1	2	3	4	Total	
1	2	0	0	0	2	
2	0	12	1	0	13	
3	0	0	5	0	5	
4	0	0	0	4	4	
Total	2	12	6	4	24	

*The number of rows with at least one missing value is 0

Table B-5*Quality Review of TE Studies - Internal Validity Only*

Author	Design	Quality (1-3) Rater1/Rater 2
Center & Wascom, 1984	ABAB	2/2
Conyers, Miltenberger, Gubin, Barenz, Jurgens, Sailer, Haugen, M., & Kopp, B. (2004).	ABAB	2/2
De Martini-Scully, Bray,& Kehle, 2000	MBD (2 baselines & 1 control)	2/2
Filcheck, McNeil, Greco, & Bernard, 2004	ABACC'	1/1
Higgins, Williams, & McLaughlin, 2001	MBD (3 baselines)	2/2
Himle, Woods, & Bunaciu, 2008	AB design replicated across 4 students	1/1
Kazdin & Geesey, 1980	ABC replicated with 2 students	1/1
Kazdin, & Mascitelli, 1980	ABC replicated with 2 students	1/1
Klimas, & McLaughlin, 2007	Changing Criterion	1/1
Maglio & McLaughlin, 1981	ABAB with 3 week fading period and 2 week F-U	3/2
McGoey & DuPaul, 2000	MBD (4 baselines)	3/3
Mottram, Bray,, Kehle,, Broudy, & Jenson, 2002	MBD (3 baselines & 1 control)	3/3
Musser, Bray, Kehle, & Jenson, 2001	MBD (3 baselines & 2 controls)	3/3
Reitman, 2004	ABAB	2/2
Rosenberg, 1986	AB	1/1

Table B-5 continued

Author	Design	Quality (1-3) Rater1/Rater 2
Salend & Allen, 1985	ABCBC (alternating treatments design across settings added reversal phases)	1/1
Salend & Lamb, 1986	ABAB	2/2
Salend, Tintle, & Balber, 1988	ABAB	2/2
Simon, Ayllon, & Milan, 1982	ABCB	1/1
Smith & Fowler, 1984	ABAC	1/1
Sran & Borrero, 2010	ABA	1/1
Stevens, Sidener, Reeve, & Sidener, 2011	MBD (2 baselines)	1/1
Sullivan & O'Leary, 1990	ABCBC	1/1
Truchlicka, McLaughlin, & Swain, 1998	MBD (3 baselines)	2/2

Table B-6*Included Studies and Effect Sizes in Alpha Order by 1st Author*

ID #	Author	Tau-U ES	Standard Error	90% CI
1	Center & Wascom, 1984	0.71	0.12	0.52, 0.90
2	Conyers, Miltenberger, Gubin, Barenz, Jurgens, Sailer, Haugen, M., & Kopp, B. (2004).	0.83	0.16	0.57, 1.00
3	De Martini-Scully, Bray,& Kehle, 2000	0.95	0.16	0.68, .22
4	Filcheck, McNeil,Greco,& Bernard, 2004	0.67	0.23	0.28, 1.00
5	Higgins, Williams, & McLaughlin,2001	0.98	0.18	0.68, 1.00
6	Himle, Woods, & Bunaciu, 2008	0.65	0.23	0.27, 1.00
7	Kazdin & Geesey, 1980	1	0.21	0.65, 1.00
8	Kazdin,& Mascitelli,1980	0.99	0.21	0.65, 1.00
9	Klimas, & McLaughlin, 2007	1	0.30	0.50, 1.00
10	Maglio & McLaughlin, 1981	1	0.31	0.48, 1.00
11	McGoey & DuPaul, 2000	0.75	0.17	0.48, 1.00
12	Mottram, Bray,, Kehle,, Broudy, & Jenson, 2002	0.99	0.12	0.79, 1.00
13	Musser, Bray, Kehle,& Jenson, 2001	0.88	0.18	0.59, 1.00
14	Reitman, 2004	0.69	0.15	0.45, 0.94
15	Rosenberg, 1986	0.99	0.15	0.74, 1.00
16	Salend & Allen, 1985	1	0.25	0.59, 1.00
17	Salend & Lamb, 1986	1	0.19	0.68, 1.00
18	Salend, Tintle, & Balber, 1988	1	0.24	0.60, 1.00
19	Simon, Ayllon, & Milan, 1982	0.78	0.20	0.4, 1.00

Table B-6 continued

ID #	Author	Tau-U ES	Standard Error	90% CI
20	Smith & Fowler, 1984	0.92	0.13	0.69, 1.00
21	Sran & Borrero, 2010	0.4	0.09	0.25, 0.55
22	Stevens, Sidener, Reeve, & Sidener, 2011	1	0.21	0.66, 1.00
23	Sullivan & O'Leary, 1990	1	0.22	0.64, 1.00
24	Truchlicka, McLaughlin, & Swain, 1998	0.35	0.12	0.16, 0.55
OVERALL		0.78	0.03	0.72, 0.83

*Note: If the higher end of the CI exceeded 1.00 the number was rounded down to 1.00.

Table B-7

Summary of Included Studies and Effect Sizes

Rating	Number of Studies	Tau-U	CI ₉₀	SE
Weak	12	.76	[.68, .84]	.05
Medium	8	.74	[.65, .83]	.05
Strong	4	.91	[.77, 1.00]	.08

*Note: the upper CI was rounded down to 1.0.

Table B-8*Inter-rater Reliability for Coded Categories*

Section	Interrater agreement
Setting of the study	100%
Study participants (e.g. gender, N, age, disability category, intervention setting)	100%
Intervention features (e.g. delivery, training, active ingredients)	97%
Methodology (e.g. design, graphs)	100%
Measures (e.g. DV, IV, validity)	98%
Overall	99%

Table B-9*Summary of Participant Characteristics*

Study Characteristic	Number	Percent
Included studies	24	
Total number of students	84	
Number of ES	79	
Age	3-5	35
	6-15	65
Educational Status	Special Education	71
	General Education	29
Setting of Intervention	Special Education	50
	General Education	50

Table B-10*Summary of Study Characteristics*

		Number	Percent
		(#)	(%)
Dependent Variable	Behavior	50	63
	Academic	29	37
Educational Status	Special Education	17	71
	General Education	7	29
Response Cost	Yes	16	67
	No	8	33

Table B-11*Active Ingredients (Steps) of Token Economy Systems*

Moderator		#	Tau-U ES	T value	Two tailed P value
Step 1- Visual Chart	Not Included	15	0.703		
	Included	9	0.944		
			0.141	2.093	0.047
Step 2- Specific Times	Not Included	3	0.891		
	Included	21	0.851		
			0.038	0.326	0.747
Step 3- Token Identified	Not Included	2	0.971		
	Included	22	0.845		
			0.124	2.599	0.016
Step 4- Visual of reinforcers and cost	Not Included	21	0.835		
	Included	3	1		
			0.164	3.700	0.001

Table B-12*Combination of Active Ingredients (Steps) Comparison for TE*

Moderator*		# of Studies	Tau-U ES	SE	T value	Two tailed P value
Step 1 & 2	Not Included	15	0.803	0.0578		
	Included	9	0.944	0.0344		
			0.141		2.098	0.0475
Step 1 & 3	Not Included	17	0.822	0.052		
	Included	7	0.937	0.044		
			0.114		1.660	0.110
Step 1 & 4	Not Included	21	0.835	0.044		
	Included	3	1	0		
			0.164		3.698	0.001
Step 1, 2, & 3	Not Included	17	0.822	0.052		
	Included	7	0.937	0.044		
			0.114		1.660	0.110
Step 1, 2, & 4	Not Included	21	0.835	0.203		
	Included	3	1	0		
			0.164		0.778	0.444
Step 1, 3, & 4	Not Included	21	0.835	0.203		
	Included	3	1	0		
			0.164		0.778	0.444

Table B-12 continued

Moderator*		# of Studies	Tau-U ES	SE	T value	Two tailed P value
Step 2 & 3	Not Included	5	0.922	0.063		
	Included	19	0.838	0.048		
			0.083		0.384	0.704
Step 2 & 4	Not Included	21	0.835	0.203		
	Included	3	1	0		
			0.164		0.778	0.444
Step 2, 3, & 4	Not Included	21	0.835	0.203		
	Included	3	1	0		
			0.164		0.778	0.444

*Note: Step 1- Visual Chart, Step 2- Specific Times, Step 3- Token Identified, and Step 4- Visual of reinforcers and cost. The above represents analysis of the combined steps. For example, Step 1 & 2 nine studies included both of the steps; fifteen studies included only one of the two steps.

Table B-13*Summary Statistics of Moderators for a TE*

Moderator	N of Studies	Tau-U ES	SE/SD	Lower CI	Upper CI	T value	two tailed P value
Age 6 - 15	18	0.904	0.176	0.816	0.991		
Age 3 - 5	6	0.710	0.177	0.523	0.897	2.317	0.033
SPED Setting	12	0.848	0.188	0.729	0.967		
General Ed. Setting	12	0.863	0.205	0.731	0.993	0.176	0.862
RC Not Included	8	0.893	0.212	0.714	1.000		
RC Included	16	0.838	0.195	0.734	0.941	0.726	0.476
Verbal reminder Included	19	0.883	0.342	0.324	1.000		
Verbal reminder Not Included	5	0.750	0.132	0.819	0.947	0.852	0.403
Academic Readiness Behavior Goal	5	0.750	0.342	0.324	1.000		
Behavior Goal	19	0.883	0.132	0.819	0.947	0.851	0.404

Table B-14*No-Cost Backup Reinforcers*

Reinforcer	Example
Special jobs in the class	A student wants to be an office assistant or assist with the custodians.
Computer time	Student wants to play an academic game for 15 minutes.
Academic extra assistance	Free answer on 1 test item, homework pass, late assignment pass (limited to a few days within the due date)
Special chair or work location	Student want to sit in the teacher's chair that rolls for the day or student can sit in a beanbag to do work.
Positive note home or phone call	Additional positive note home so that child can receive or do something special.
Buddy Time	Work with a buddy for an assignment
Choose your desk	Get to pick in the classroom where you want to sit. I generally gave 3 choices so that I could keep students away from negative situations.

APPENDIX C

TOKEN ECONOMY INTERVENTION
DATA CODING INSTRUMENT

This data-coding instrument was adapted from Tolan, Bass, Henry, & Schoeny (2008).

Article INFORMATION AND SCREENING

- A1. Study ID# ___ ___ ___ [ID]
- A2. Coding Date ___ - ___ - ___ [CODDATE]
- A3. Coder initials ___ [CODER]
- A4. Primary author (LN, FI) [AUTHOR]
-
- A5. Year of publication ___ ___ ___ [PUBYR]
- A6. Does study measure token economy as an outcome? [TE]
- 1. yes
 - 2. no (STOP)
- A7. Is the focus of this publication a Token Economy intervention intended to increase or decrease BEHAVIOR OR ACADEMICS of school age children? (DV might not necessarily be identified as token economy – contingency management, level system, token system are all appropriate) [TEINT]
- 1. yes – TE is stated as a primary goal
 - 2. yes – TE is a primary construct used to operationalize or measure the stated primary goal (e.g., participation, achievement)
 - 3. yes – as a secondary outcome
 - 4. no (STOP)
 - 99. cannot tell
- A8. Was this study conducted in North America? [USA]
- 1. Yes
 - 2. no (STOP)
 - 99. cannot tell (set aside)
- A9. Where was this study conducted? [SITE]
- 1. Preschool
 - 2. Head Start
 - 3. Elementary school
 - 4. Secondary (Junior High or High School)
 - 5. other: _____ (STOP)
(e.g., residential facility, prison, home, day care, or laboratory setting)
 - 99. cannot tell

A10. Indicate the type of paper/study below: [PAPER]

- 1. outcome/program/intervention evaluation (CONTINUE)
- 2. review of token economy studies (STOP)
- 3. position paper, editorial, book review (STOP)
- 4. guidelines for treatment or intervention (STOP)
- 5. qualitative research (STOP)
- 6. other: _____(STOP)
- 99. cannot tell (STOP)

A11. Indicate the source of the paper below: [SOURCE]

- 1. peer-reviewed journal
- 2. Dissertation (STOP)
- 3. technical report (STOP)
- 4. other: _____
- 99. cannot tell

SECTION A

Setting OF STUDY

B1. Primary author's discipline: [AUTHDISC]

- 1. education
- 2. psychology
- 3. child development
- 4. speech/language pathology
- 5. social work
- 6. other: _____
- 99. cannot tell

B2. Research setting [SETTING]

- 1. inclusive setting
- 2. general education
- 3. special education
- 4. alternative
- 6. other: _____
- 99. cannot tell

SECTION B

STUDY PARTICIPANTS

C1. Total N at beginning of study _____ [INITIALN]

C2. Total N at end of study _____ [FINALN]

C3. Race/ethnicity of participants – indicate predominant ethnicity [RACE]

- 1. Caucasian Specify _____

- 2. African American Specify _____
- 3. Hispanic/Latino Specify _____
- 4. Other Specify _____
- 99. cannot determine

C4. Total Female N _____ [FEMALE]

C5. Total males N _____ [MALES]

C6. Indicated socioeconomic status of majority of participants [SESCAT]

- 1. Low (at or below poverty line)
- 2. Working or lower middle class
- 3. Middle class or above
- 4. Combination
- 99. cannot tell

C7. Ages of Participants: _____

Participants for the study

C8. SPED Disability categories represented [TXDISABIL]

- 1. Intellectual/Developmental disabilities
- 2. Visual impairment
- 3. Hearing impairment
- 4. Multiple/severe
- 5. at-risk, incl. socially isolated or other risk factors as identified by authors)
- 6. Speech/language
- 7. Diverse group
- 8. typical
- 9. abused/maltreated
- 99. cannot tell

SECTION D

INTERVENTION FEATURES

D1. What do the authors call the intervention? [TXNAME]

D2. Who delivered the intervention? [INTVNIST]

- 1. teacher-mediated
- 2. Paraprofessional mediated
- 3. peer-mediated
- 4. teacher-mediated with caregiver component
- 5. multifaceted program w multiple contexts incl. home
- 6. caregiver mediated
- 7. experimenter
- 99. cannot tell

D3. Was the interventionist trained? [INTVNISTTR]

- 1. Yes
- 2. No
- 99 cannot tell

D3. Program model (for children) [MODEL]

- 1. environmental
- 2. academic
- 3. behavioral
- 4. other: _____

D4. Treatment fidelity: measure reported, or comments included [FIDELITY]

- 1. yes
- 2. no
- 99. cannot tell

D5. Duration of intervention [DURATION]

- 1. up to 2 weeks
- 2. 2 weeks to 1 month
- 3. 1-3 months
- 4. 4-6 months
- 5. 7-9 months
- 6. 10 months to 1 year
- 7. more than 1 year
- 99. cannot determine

D6 Active Ingredients of the Token Economy:

- 1. Identify behaviors that earn tokens
- 2. Assign value of token for each behavior
- 3. Identify schedule of when the token will be recieved
- 4. Identify what the token is: _____
- 5. Identify back up reinforcers: _____
- 6. Quantity of tokens for exchange of back up reinforcer:

- 7. Schedule for exchange of tokens for back up:

- 8. Does the TE include a cost response: YES or NO Describe:

SECTION E METHODOLOGICAL FEATURES AND QUALITY

E1. Type of design [DESIGN]

- 1. randomized controlled trials

- 2. quasi-experimental design
- 3. within-group pre-post test design
- 4 SCR design: _____
- 5. other: _____
- 99. cannot determine

E2. Graphs Included [GRAINC]

- 1. Yes
- 2. No [STOP]

E3. Reliability taken [RELIA]

- 1. Yes [if so How often? _____]
- 2 No

E4. Reliability Results [RELIARES]

- _____

E4. Who checked reliability? [CHEREL]

- 1. _____
- 99. cannot determine

Final Decision regarding this study

E5. Should this study be retained for further analysis? [INCLUDE]

- 1. yes
- 2. no
- 99. unsure based upon information obtained up to this point

**SECTION F
MEASURES**

One SECTION F should be completed for each outcome variable.

F1. Study ID ___ ___ ___ ___ [ID]

F2. Outcome number _____ [OUTID]

F3. Insert author's label for this outcome [LABEL]

Codes for Dependent Variable

F4. Construct measured [DV]

- 1. Environmental 1 _____
- 2. Academic 1 _____
- 3. Behavior 1 _____
- 4. Environmental 2 _____
- 5. Academic 2 _____
- 6. Behavior 2 _____

- 7. Environmental 3 _____
- 8. Academic 3 _____
- 9. Behavior 3 _____
- 10. Environmental 4 _____
- 11. Academic 4 _____
- 12. Behavior 4 _____
- 99. cannot determine

F5. Respondent or source of data [DVSOURCE]

- 1. Parent or caregiver report
- 2. Teacher report
- 3. Independent observer
- 4. Therapist (occupational, speech/language, etc.)
- 5. Child
- 6. other: _____
- 99. cannot determine or not reported

F6. Type of token [TOKEN]

- 1. Point
- 2. Sticker
- 3. coin
- 4. Ticket
- 6. other: _____
- 99. cannot determine or not reported

F7. Token Delivery [TOKDEL]

- 1. frequency: _____

F8. Back Up reinforcer [BUREIN]

- 1. Survey taken [yes / no/ cannot tell]
- 2. Menu of back up cost [yes / no/ cannot tell]
- 3. Frequency of purchase for back up reinforce _____

F9. Is information regarding validity provided? [VALID]

- 1. yes (e.g., inter-rater, internal consistency)
- 2. no
- 99. cannot determine or unclear

F10. Was data collected regarding maintenance of treatment effects over time (follow-up)?

[FOLLOW]

- 1. yes (proceed to next item)
- 2. no
- 99. cannot determine or unclear

F11. How much time (in months) passed between the end of the study and the collection of follow-up data? [FOLTIME]

99. cannot determine or not applicable

Codes for Data

F12. Baseline Data Points _____ [BDP]

F13. Baseline Mean _____ [BLM]

F14. Baseline Equated _____ [BEQU]

F15. Baseline SD _____ [BLSD]

F16. Intervention Data Points _____ [IDP]

F17. Intervention Mean _____ [INM]

F18. Intervention Mean Equated _____ [INEQ]

F19. Intervention SD _____ [INSD]

F20. Effect size (if calculated) _____ [ES]

F21. Standard error of effect size _____ [ESERR]

APPENDIX D

Operational definitions and coding procedures for each of these potential moderator variables are:

- (a) *Student Characteristic Variables:* student age, gender, disability category, and instructional setting.
- The *student age* variable has two potential levels (3-5; 6-18).
 - The *Disability* category is defined according to Individuals with Disabilities Education Improvement Act (IDEIA, 2004) special education eligibility categories (11). Students who are not identified as receiving special education services will be identified as general education or general education at-risk.
 - The *Instructional Setting* variable will be coded based on where the intervention takes place (e.g., special education or general education).
- (b) *Response Cost.* Response cost is defined as tokens being removed for behavior and has two levels (yes or no).
- (c) *Verbal reminder during token earning phase* is based on indications of whether the implementer verbally reminded the student about the expectations or goals set during the token earning phase. This moderator has two levels (yes or no).
- (d) *Is the outcome measure behavior or academic, in nature?* This variable was based on the indicated goal that was rewarded with a token. It had two levels.

- (e) *Active Ingredients*. Five levels: (1) visual chart of behaviors, (2) specification in when tokens will be earned, (3) determination of token, (4) visual chart of reinforcers and cost, and (5) access to back up reinforcers (five basic elements defined in Vannest, Reynolds, and Kamphaus, 2008).
- (f) Methodological quality of study. Scales (1-4) were developed for the assessment of methodological quality of SCR designs. All of the designs fell in three design features: Multiple baseline design (n=5), Reversal Design (n=18) and Changing Criterion design (n=1). This moderator had four levels.
- (b) *Response Cost*. Response cost is defined as tokens being removed for behavior and has two levels (yes or no).
- (c) *Verbal reminder during token earning phase* is based on indications of whether the implementer verbally reminded the student about the expectations or goals set during the token earning phase. This moderator has two levels (yes or no).
- (d) *Is the outcome measure behavior or academic, in nature?* This variable was based on the indicated goal that was rewarded with a token. It had two levels.
- (e) *Active Ingredients*. Five levels: (1) visual chart of behaviors, (2) specification in when tokens will be earned, (3) determination of token, (4) visual chart of reinforcers and cost, and (5) access to back up reinforcers (five basic elements defined in Vannest, Reynolds, and Kamphaus, 2008).

(f) Methodological quality of study. Scales (1-4) were developed for the assessment of methodological quality of SCR designs. All of the designs fell in three design features: Multiple baseline design (n=5), Reversal Design (n=18) and Changing Criterion design (n=1). This moderator had four levels.

VITA

Denise A. Soares

*Mailstop 4225; Harrington Tower 645
College Station, Texas 77843-4225
denisesoares@tamu.edu 979-845-6761 (W)*

Research Interest

There is reason to be concerned about current policy directions that have potential to affect students with emotional and behavioral disorders (EBD) and the teachers who teach them. My research interests encompass the effective use of teacher time, teacher roles and responsibilities, and academic/behavior interventions for student with EBD.

Education

Ph.D.	Texas A&M University	Graduation 2011 Advisor and Dissertation Chair: Kimberly J. Vannest, Ph.D. Dissertation: Effect Size and Moderators of Effects for Token Economy Interventions	Major: Education Psychology Minor: Special Education
M.Ed.	University of Houston	2003	Major: Administration and Supervision
B.S.	Southwest Texas State University	1989	Major: Elementary Education Minor: Special Education