REVISITING THE EFFECTS OF PRAISE ON STUDENT BEHAVIOR: A SYSTEMATIC REVIEW AND META-ANALYSIS OF THE SINGLE CASE RESEARCH

A Dissertation

by

HEATHER LEA HATTON

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	Mack D. Burke
Committee Members,	Mandy Rispoli
	Shanna Hagan-Burke
	Victor Willson
Head of Department,	Victor Willson

August 2015

Major Subject: Educational Psychology

ABSTRACT

The purpose of this dissertation was to: (1) conduct a systematic review of the single case research examining the effects of praise on student behavior, and (2) conduct a meta-analysis of the single-case research examining the effect of praise that meets the What Works Clearinghouse (WWC) standards for design quality and evidence of effects. In study one, the 28 included studies were coded for descriptive features to determine for whom and under what circumstances the effects of praise have been studied, and quality appraisal coding was conducted to determine if the studies met the WWC standards. For study two, four effect size metrics were calculated to determine the overall effects of praise. Additionally, the effects of nine moderator variables were examined. Overall, sufficient empirical evidence exists to recommend praise as an evidence-based practice in classroom settings. More specifically, praise can be recommended as an evidence-based practice for: (a) students with high incidence disabilities, (b) students in elementary classrooms, and (c) modifying social behaviors. There is also sufficient evidence to recommend praise be delivered: (a) contingent upon engagement in the target behavior, and (b) using a variable ratio schedule of reinforcement. The overall effects of praise on student behavior are moderate to strong across all four effect size metrics. The effects of moderator variables are complex. Implications for practice, areas of future research and limitations were addressed.

DEDICATION

This work is dedicated to my grandmother, Leona Luedeker, who gave me the gift of unconditional love and taught me the power of positive reinforcement.

ACKNOWLEDGMENTS

I would like to thank my committee; Dr. Burke, Dr. Rispoli, Dr. Hagan-Burke and Dr. Willson, for their support and guidance throughout the course of my doctoral program. Dr. Burke, thank you for patiently addressing every item on every list that I made. Dr. Rispoli, thank you reassuring me when I second-guessed myself. Dr. Hagan-Burke, thank you for encouraging this "Tier 3" student to persevere. Dr. Willson, thank you for helping me develop an appreciation of the methods used in this research.

In addition, I extend many thanks to the amazing colleagues who provided encouragement throughout this process. Specifically, thank you to Samar, Lisa, Jen, Nan, and Nancy who provided interrater reliability for both studies.

Throughout my doctoral program, I was never farther than a text message away from the support of friends who walked similar paths and could provide reassurance and a road map. Without Andy, Chris, and Gene there were a few times when I would have remained lost.

I want to acknowledge the amazing educators and students who motivated me to pursue the identification and dissemination of evidence-based practices in special education. In particular, I want to give a shout out to Raffy, Mr. Moore, and all of the students faculty and staff of Martin Middle School.

I also want to acknowledge the memories of those who are no longer here to witness this accomplishment. I'm not sure this is what my dad had in mind when he encouraged me get a teaching certificate "just in case," but I imagine he would be proud.

Jane Nethercutt, let me get my first taste of providing technical assistance and conducting program evaluation. In your honor, I drew a lot of pictures to explain this data. And, Michael, I wish I could call you and share this accomplishment with you.

Finally, I want to extend a thank you to Mom and Sir. As educators you set me on this path. Thank you for being role models.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Overview

The use of praise in classroom setting is a controversial issue (Maag, 2001). A significant number of empirical studies examining the effects of praise based on either operant conditioning theory or cognitive evaluative theory have been conducted. Additionally, several literature reviews have summarized the empirical findings (e.g., Bayat, 2011; Henderlong & Lepper, 2002; Kennedy & Willcutt, 1964). Contradictory findings in the empirical studies and reviews have arguably led to confusion regarding the use of praise (Maag, 2001). While current procedures for determining evidence-based practice from single-subject research could provide a unique contribution the debate (Kratochwill et al., 2013), no systematic review of the single-case literature has been conducted.

In addition to being controversial, praise is a complex intervention (Brophy, 1981). Several literature reviews suggest that the effects of praise may be moderated by: (a) characteristics of the individual receiving the praise, (b) characteristics of the setting in which praise is offered, (c) the topography of the praise statement, and (d) the topography of the behavior upon which praise is contingent (Bayat, 2011; Brophy, 1981; Henderlong & Lepper, 2002). While previous meta-analyses have examined variables moderating the effects of praise on intrinsic motivation (Cameron & Pierce, 1994; Deci, Koestner, & Ryan, 1999), no meta-analysis of the single-case studies examining the

effects of praise on behavior has been conducted. Identifying variables that moderate the effect of praise on behavior will improve recommendations regarding the appropriate use of praise in classroom settings.

Research Objective for Study One

The purpose of the study is to systematically: (a) identify studies using single-case designs to examine the effects of praise on student behavior; (b) identify the characteristics of the participants, settings, intervention topography, and outcome variable topography included in the existing research, and (c) evaluate the identified studies using the quality standards established by the WWC (2014). The review will specifically address the following research questions: (a) With whom and under what conditions have the effects of praise on student behavior been studied?, (b) Is there sufficient empirical support to recommend praise as an evidence-based strategy?, (c) For whom and under what conditions is there sufficient evidence to recommend praise as an evidence-based practice?

Research Objectives for Study Two

The purpose of the study is to conduct a meta-analysis of existing research employing single-case designs to: (a) estimate the magnitude of effect of praise on student behavior, and (b) examine variables that potentially moderate the effects of praise. The meta-analysis will specifically address the following the questions: (a) What is the overall effect of praise on student behavior?, and (b) For whom and under what conditions do the effects of praise generalize?

CHAPTER II

THE STATE OF THE EVIDENCE FOR PRAISE

Introduction

The use of praise in classroom setting is a controversial issue (Maag, 2001). The controversy arises from two parallel lines of research. In one line of research, praise is recommended as an effective behavioral intervention for encouraging academic and appropriate social behaviors, while preventing inappropriate social behaviors (Gable, Hester, Rock, & Hughes, 2009; Kennedy & Willcutt, 1964; Merrett & Houghton, 1989). In the other line of research, praise is sometimes denounced as a coercive teacher-student interaction that decreases intrinsic motivation and leads to learned helplessness (Dweck, 2007; Kohn, 2001). With such a large and contradictory body of evidence, it is particularly necessary to base conclusions on systematic literature reviews (Burns, 2012). In light of this, it is interesting that no previously published review or synthesis has systematically evaluated the single-case research literature on praise to determine if there is sufficient empirical evidence to recommend praise as an evidence-based practice in classroom settings.

The first line of research on praise began at the end of the nineteenth century, when Binet & Vaschide (1897) examined the effects of praise on student behavior. The Binet & Vaschide (1897) study is characteristic of the more than 30 studies conducted through the middle of the 20th century. These studies were conducted in classroom settings and employed quasi-experimental group designs. The outcome measures of

interest to researchers during this time included discrimination learning and motor tasks in school-aged children measured through direct observation (Kennedy & Willcutt, 1964) . In 1964, Kennedy and Willcutt suggested the field should consider the use of what they referred to as functional designs.

Accordingly, in 1968, three studies published in the inaugural issue of the *Journal of Applied Behavior Analysis* examined the effects of teacher behaviors, including praise, on student behavior (Hall, Lund, & Jackson, 1968; Hart, Reynolds, Baer, Brawley, & Harris, 1968; Thomas, Becker, & Armstrong, 1968). These studies are characteristic of the operant learning theory research being developed during this time in the emerging field of applied behavior analysis employing single-case designs. The outcome measures of interest to researchers include appropriate social behaviors, inappropriate social behaviors, and academic behaviors measured primarily through direct observation methods.

The second line of research originated in the second half of the twentieth century with Deci's (1971) study examining the effects of external rewards, including praise, on intrinsic motivation. This study is characteristic of the cognitive evaluative theory research employing experimental and quasi-experimental group designs in clinical settings. The outcome measure of interest to the researchers is intrinsic motivation reported as: (a) time on task during a free time condition measured by direct observation and (b) self-reports of motivation measured with a survey instrument.

A large quantity of empirical studies have been conducted under each line of research. As a result, several literature reviews have been published examining the

effects of praise (e.g., Bayat, 2011; Henderlong & Lepper, 2002; Kennedy & Willcutt, 1964), often reaching contradictory conclusions. Interpretation of these conclusions has arguably contributed to the controversy regarding the use of praise in classroom settings (Maag, 2001).

Previous Reviews

Narrative reviews examining the effects of praise are divided into three groups. First, four previous reviews examined the effects of praise on student behavior (Gable et al., 2009; Kennedy & Willcutt, 1964; Merrett & Houghton, 1989; Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008;). None of these reviews can be considered systematic literature reviews. Kennedy and Willcutt (1964) conducted a comprehensive review of studies published between 1897 and 1964, but did not provide details of the literature search or inclusion criteria. Merrett and Houghton (1989) limited studies included to those conducted in secondary schools. Gable et al. (2009) and Simonsen et al. (2008) included only a representative sample of studies. Further, none of the reviews conducted a quality appraisal of the studies included. All of the reviews conclude that praise is an effective practice, and Simonsen et al. (2008) suggested praise is an evidence-based practice.

Second, four previous reviews included studies that examined the effects of praise on students' intrinsic motivation (Cannella, 1986; Fair & Silvestri, 1992, Henderlong & Lepper, 2002; Morgan, 1984). Two of the reviews were comprehensive reviews of the literature (Fair & Silvestri, 1992; Henderlong & Lepper, 2002). Morgan (1984) limited his review to studies published between 1976 and 1982. Cannella (1986)

limited her review to studies with participants in elementary classrooms. As with the studies examining the effects of praise on behavior none of the reviews conducted a quality appraisal of the studies included. The reviews conclude that praise may enhance or undermine intrinsic motivation dependent on a variety of moderating variables (e.g., student demographics, type of praise). The reviewers suggest that praise should be used with caution.

Third, four previous reviews included studies that examined both the effects of praise on student behavior and also included studies that examined the effects of praise on intrinsic motivation (Bayat, 2011; Brophy, 1981; Emmer, 1988; Ferguson, 2013). None of these reviews were comprehensive reviews of the literature. Instead, the reviewers included a "representative selection" of published studies and did not provide specific inclusion criteria. None of the reviews conducted a quality appraisal of the primary sources included. The researchers concluded that praise is a complex process and that the effects of praise on behavior and motivation are moderated by a variety of variables (e.g., setting, interest in activity). The reviewers suggested that praise should be used with caution.

The variability of findings in the previous reviews allows critics to dismiss the effects of praise on student behavior and emphasize the possibility that praise could be coercive and detrimental to intrinsic motivation (Dweck, 2007; Kohn, 2001). In order to reach this conclusion, critics of praise have ignored large portions of the research on praise in favor of a few studies that align with their point of view (Maag, 2001). Current legislation and policy emphasizes implementation of evidence-based practices;

(Individuals with Disabilities Act, 2004; No Child Left Behind, 2002). Given the current focus on evidence-based practices, educators are often confused as to whether using praise is considered best-practice given the conflicting recommendations from the literature. Unfortunately, no previous reviews of the research examining the effects of praise have systematically determined if sufficient empirical evidence exists to recommend praise as an evidence-based practice in classroom settings.

Establishing Evidence-Based Practices Using Single Case Research

Single-case research can contribute, significantly, to the identification of evidence-based practices (Kratochwill et al., 2013; Wendt & Miller, 2012).Single-case designs are one of the strongest nonrandomized experimental designs (Shadish, Rindskopf, & Hedges, 2008). As a special case of time series designs, single case designs make a unique contribution to the understanding of intervention effects (Burns, 2012; Shadish, et al., 2008). Studies employing single-case designs allow the examination of behavior change in an individual over time (Scruggs, Mastropieri, & Castro, 1987). These designs typically have strong internal validity and identify strong intervention effects (Burns, 2012). However, it is difficult to generalize the effects of individual studies employing single-case designs to a larger population (Burns, 2012).

Horner et al. (2005) indicated that evidence-based practices can be identified using single case research if those practices have been: (a) examined in studies with acceptable methodological rigor and (b) replicated with significant effects across a number of participants. Following the Horner et al. (2005) recommendations, the WWC has recommended the 5-3-20 standard for determining evidence-based practice. The

WWC suggests that sufficient empirical evidence exists to recommend a practice as evidence-based practice if a minimum of five single case studies, including a minimum of 20 participants, have been conducted by a minimum of three independent research teams (Kratochwill et al., 2013).

A critical factor in determining whether a practice is evidence-based is the quality of the studies examining the effects of the practice (Maggin, O'Keefe, & Johnson, 2011). Conducting a quality appraisal of the existing studies is an important step in identifying evidence-based practices, and arguably, only high quality studies should be considered in the determination (Wendt & Miller, 2012). A quality appraisal ensures the methodological rigor of the included studies making the generalizability of the findings more reliable (Maggin, 2011).

Several quality standards have been put forward in the literature on determining the quality of empirical research (Wendt & Miller, 2012). The WWC (2014) has provided procedures and standards for identifying evidence-based practices from empirical research, including single-case designs that have been particularly influential in directing the field toward evidence-based practices. The WWC (2014) standards are separated into two sections. Initially, the methodological rigor of a study is evaluated using the design standards. If the study meets the design standards, with or without reservations, the evidence of effects is assessed. Based on visual analysis of the time series graph(s) included in the study, the intervention is rated as having strong, moderate or no evidence of effect on the outcome variable.

Purpose and Research Questions

A significant number of single-case studies have examined the effects of praise on student behavior (e.g., Darch & Gersten, 1985; Kazdin, 1973; Madsen, Becker, & Thomas, 1968, Sawyer, Luiselli, Ricciardi, & Gower, 2005). To date, no systematic review of these single case studies has been published. The purpose of the study is to systematically: (a) identify studies using single-case designs to examine the effects of praise on student behavior; (b) identify the characteristics of the participants, settings, intervention topography, and outcome variable topography included in the existing research, and (c) evaluate the identified studies using the quality standards established by the WWC (2014). The review will specifically address the following research questions: (a) With whom and under what conditions have the effects of praise on student behavior been studied?, (b) Is there sufficient empirical support to recommend praise as an evidence-based strategy?, (c) For whom and under what conditions is there sufficient evidence to recommend praise as an evidence-based practice?

Method

A systematic literature search was conducted to identify studies employing single-case designs that examined the effects of praise. The search consisted of three stages: (a) initial literature search, (b) initial inclusion screening, and (c) additional literature searches and screening.

Initial Literature Search

An electronic database search was conducted in three steps. First, two search strings were developed. One search string contained keywords associated with the

intervention (i.e., praise, verbal reinforcement, social reinforcement, positive feedback) joined with the Boolean operator *OR*. The second search string contained keywords associated with the relevant student outcomes (i.e., behavior, achievement, engagement) joined with the Boolean operator *OR*. The two search strings were combined using the Boolean operator *AND*. As a result, all records identified included at least one term from each search string.

The combined search string was entered into the following electronic databases: (a) Education Resources Information Center (ERIC), (b) PsycINFO, and (c) Academic Search Complete. In order to maximize the records identified, the search was configured to find the keywords in any field in a record. The searches were limited to peer-reviewed items. A total of 3,448 citations were identified. The identified citations were exported to RefWorks for inclusion screening. Duplicate citations were removed prior to the initial inclusion screening process.

Initial Inclusion Screening

In order to be included in the review, a study had to meet five inclusion criteria. First, all studies had to be published in English in a peer-reviewed journal. Publication in peer-reviewed journals was required to maximize the number of primary sources rigorous enough to meet the WWC (2014) standards for evidence-based practice. Second, all studies had to systematically manipulate praise as the independent variable. Third, all studies had to include a measure of student behavior, academic or social, as the dependent variable. Fourth, all studies had to employ a single-case experimental design (e.g., reversal, multiple baseline, alternating treatment). Fifth, the experiments in the

studies had to be conducted in typical school settings. A full rubric for inclusion screening is provided in Appendix A.

All studies identified for potential inclusion were reviewed in RefWorks. Initially, a rater reviewed the title and abstract in each record to determine if it met the inclusion criteria. If a determination could not be made, an electronic copy (PDF) of the study was downloaded and reviewed to make the determination. Once a study failed to meet any inclusion criterion, screening was stopped. The rater recorded, in RefWorks, whether or not the study met the inclusion criteria. If a study failed to meet the inclusion criteria, the criterion it failed to meet was also recorded. A total of 200 studies were retrieved for full text screening and 28 met the inclusion criteria.

Descriptive Coding Procedures

Included studies were coded for a variety of descriptive characteristics to determine with whom and under what conditions the effects of praise on student behavior has been studied. This coding was designed to gather additional information regarding: (a) participant and setting characteristics, (b) intervention characteristics, and (c) outcome characteristics. The full coding manual for descriptive characteristics is provided in Appendix B.

Participant and setting characteristics. Six items were coded related to participant and setting characteristics. These items included: (a) participant disability status, (b) participant ethnicity, (c) participant socioeconomic status, (d) participant gender, (e) educational setting, and (f) grade level

Intervention characteristics. Sixteen items were coded related to the characteristics of praise used in the studies. These items included: (a) whether a functional behavior assessment was conducted, (b) the primary role of the interventionist, (c) training provided to the interventionist, (d) the target(s) of the intervention, (e) training provided to target(s), (f) the schedule of reinforcement, (g) any additional antecedent manipulations, (h) any additional consequent interventions, (i) the type of contingency, (j) the specificity of the praise statement, (k) the inclusion of controlling language, (l) attribution to ability or effort, (m) participant awareness of the contingency, (n) participant interest in the activity, (o) treatment fidelity, and (p) social validity of the intervention.

Outcome characteristics. Four items were coded related to the characteristics of the outcome variable(s) measured in the study. These items included: (a) type of behavior, (b) direction of the expected change, (c) measurement type, and (d) data recording procedure.

WWC Standards Coding Procedures

The WWC (2014) standards were applied to cases identified within each of the included studies. A case was defined as a single independent variable, a single dependent variable, and one or more participants. For example, in a reversal or alternating treatment design, each unique combination of participant, independent variable, and dependent variable was appraised as a separate case. A case in a multiple baseline design across participants included multiple participants, but only contained a single independent variable and a single dependent variable.

Cases were appraised using the WWC (2014) evaluation standards to determine if praise has sufficient empirical support to be recommended as an evidence-based practice in classroom settings. The WWC (2014) evaluation process occurred in two stages. Initially the methodological rigor of each case was evaluated using the design standards. The evaluation was based on the methods described in the study. The methods must clearly demonstrate: (a) the independent variable is systematically manipulated by the researcher, (b) repeated measures of the outcome variable are taken over time, (c) interobserver agreement is collected for a minimum of 20% of all sessions, preferably for 20% of all sessions in each condition, (d) interobserver agreement meets the minimum threshold for accuracy for the statistic calculated, (e) the design provides at least three opportunities for replication of effect, and (f) each phase has sufficient data points to reliably demonstrate an effect. Possible ratings include: (a) meets the standards without reservation, (b) meets the standards with reservation, or (c) does not meet the standards.

Cases meeting the design standards, with or without reservations, were evaluated for evidence of effect based on visual analysis of each short-time series graph. The evidence standards required examination of the following within phase characteristics: (a) level, (b) trend, (c) variability. Additionally, the following between phase characteristics were examined: (a) immediacy of effect, (b) overlap, (c) consistency of data in a similar phase. Taking the visual analysis into consideration, the number of demonstrations of effect in the case is determined. Possible ratings include: (a) strong evidence of effects (three or more demonstrations of effect and no demonstrations of

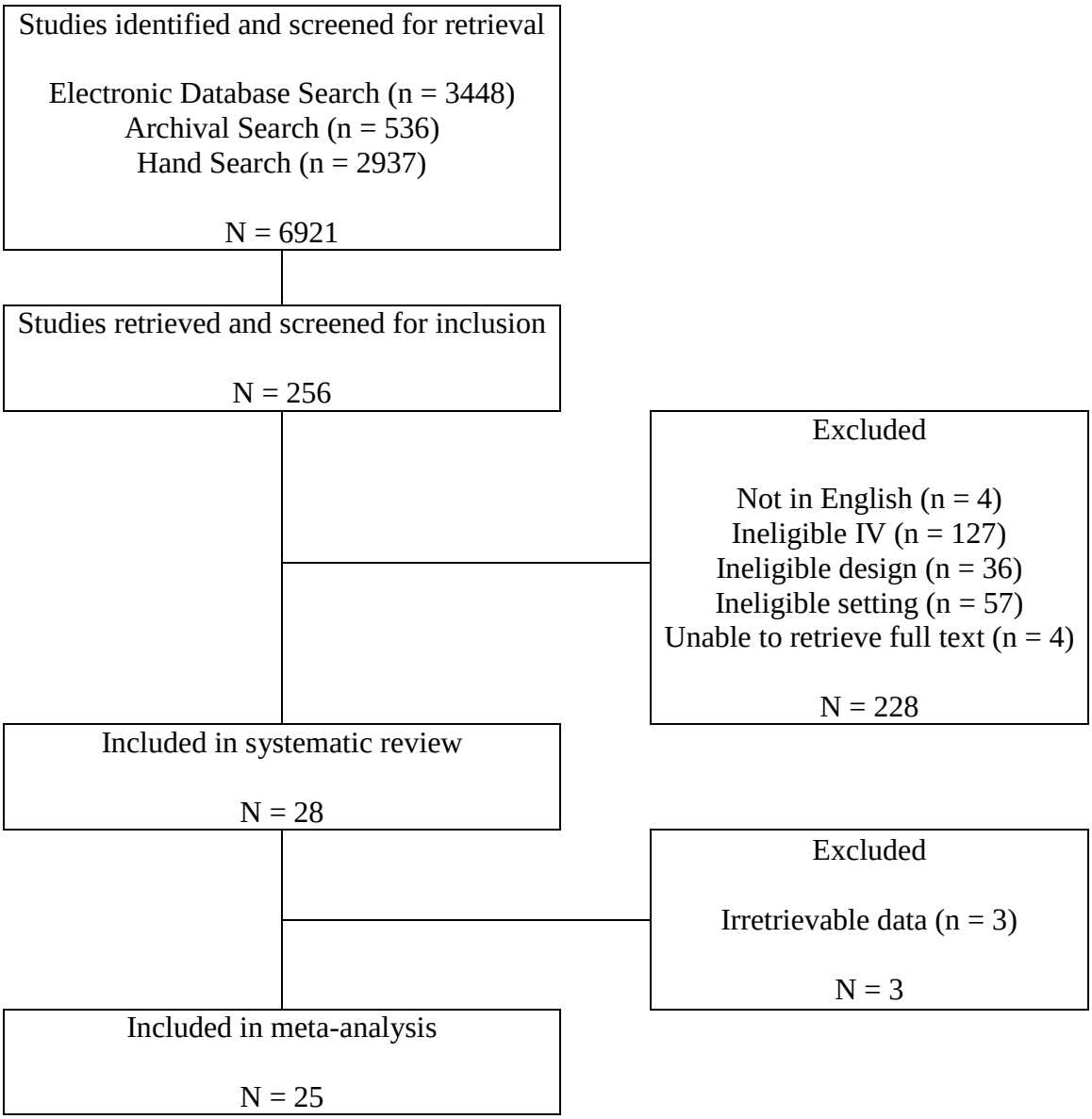
noneffect), (b) moderate evidence of effects (three or more demonstrations of effect and one or more demonstrations of noneffect, or (c) no evidence of effects (less than three demonstrations of effect).

Additional Literature Searches and Screening

After the initial literature search and inclusion screening were completed, an archival search and hand search were conducted. In the archival search, the references for all included studies were reviewed. Any citations appearing to meet the inclusion criteria were checked against the citations identified in the electronic database search. Duplicate citations were discarded. Remaining citations were screened using the same procedures used in the initial screening process. A total of 48 studies were retrieved for inclusion screening; however, none met the inclusion criteria.

In the hand search, tables of contents for the journals publishing two or more of the included studies were reviewed. The following journals were searched: (a) *The Journal of Applied Behavior Analysis*, and (b) *The Journal of School Psychology*. Tables of contents were reviewed from the inaugural issue to 2014. Any citations appearing to meet the inclusion criteria were checked against the citations identified in the electronic database search. Duplicate citations were discarded. Remaining citations were screened using the same procedures used in the initial screening process. A total of 58 studies were downloaded for inclusion screening; however, none of them met the inclusion criteria. Figure 1 details the number of citations found in each search, the number of duplicates removed, the number of studies excluded for each criterion, and the number of studies included in the review.

Figure 1. Results of the literature search and inclusion screening.



Reliability

Reliability was conducted for the electronic database search, inclusion, screening, descriptive coding, and WWC coding. The first author conducted the

electronic database search, the archival searches, and the hand searches. A graduate student with experience conducting systematic reviews replicated the electronic database search using the description provided in the methods section of this manuscript. The replication search identified an additional 26 studies. The difference is likely attributable to time elapsed between the searches. The additional citations were screened using the same procedures as in the initial electronic database search; however, none met the inclusion criteria.

The first author conducted the inclusion screening for all studies. A graduate student with experience in conducting systematic literature reviews received training on the inclusion criteria and conducted the reliability screening. Twenty-two percent of the studies retrieved and screened for inclusion were screened for reliability. The sample of studies screened for reliability was comprised of the 28 included studies and 28 excluded studies randomly selected from the studies retrieved for full text screening. Reliability was calculated by dividing the number of agreements by the total number agreements and disagreements. The percent agreement was 84%. Disagreements were resolved through discussion between the first author and the graduate student resulting in 100% final agreement across all studies.

The first author coded all included sources for descriptive characteristics and WWC standards. Twenty percent of studies were selected, randomly, for reliability coding. Graduate students with experience applying the WWC standards coded the selected sources. The same coders received training in the descriptive coding scheme and applied it to the selected studies. Reliability was calculated by dividing the number

of agreements by the total number agreements and disagreements. Percent agreement on the WWC standards was 95%. Percent agreement on the descriptive coding was 77%. Disagreements were resolved through discussion between the first author and the graduate student resulting in 100% final agreement across all codes.

Results

Twenty-eight studies were identified for inclusion in this systematic review. The studies were published between 1968 and 2014. Table 1 summarizes the characteristics of the studies. More than half of the studies (n = 17) of the studies were published between 1970 and 1989. Approximately one quarter of the studies were published between 2000 and 2015. Twenty-five percent of the studies (n = 7) were published in the Journal of Applied Behavior Analysis. The only other journal publishing multiple studies (n = 3) was the Journal of School Psychology. The most common designs employed in the studies were reversal (n = 16) and multiple baseline (n = 8). Two studies employed an alternating treatment design and two studies employed a case study design.

Descriptive Coding

The included studies were coded to determine with whom and under what conditions the effects of praise on student behavior have been examined.

Participant and setting characteristics. A total of 85 participants were included across the 28 studies. The effects of praise were examined with a wide variety of students, however the majority of participants in the studies were students with high incidence disabilities (n = 36). Ten participants were students with low incidence

Table 1. Summary of study features

Level	# of Studies	% of Studies	n	% of Participants	k	% of Cases
			Publication Year			
1960 – 1969	2	7%	5	6%	5	7%
1970 – 1979	8	29%	16	19%	16	22%
1980 – 1989	9	32%	33	38%	30	41%
1990 – 1999	1	4%	3	3%	1	1%
2000 – 2009	4	14%	6	7%	6	8%
2010 – 2015	4	14%	22	26%	16	22%
			Single Case Design			
AB	2	7%	3	3%	3	4%
ABAB	16	55%	48	56%	48	65%
MB	8	28%	25	29%	14	19%
AT	2	7%	9	10%	9	12%

Note: n = participants; k = cases.

disabilities and ten were typically developing students. For the remainder of the students' disability status was not reported (n = 26). More male participants (n = 34) were involved in the studies than females (n = 20). For the remaining participants, gender was not reported (n = 20) or was reported as a mixture of male and female participants without individually identifying participant gender (n = 11). For the majority of participants, ethnicity (n = 52) and socioeconomic status (n = 75) were not reported. In addition to involving a wide variety of participants, the studies were conducted across a range of settings. Most of the participants were evenly divided between general education (n = 34) and special education (n = 40) settings. Praise was provided in both settings for four students. The setting for the remaining seven students was not reported. Almost half of the participants were in early elementary classrooms (n = 37). However, participants were also in upper elementary (n = 11), middle school (n = 7), and high

Table 2. Summary of participant and setting characteristics.

Level	# of Studies	% of Studies	n	% of Participants	k	% of Cases
			Disability Status			
NR	9	32%	26	31%	17	23%
HI	10	36%	36	42%	34	46%
LI	6	21%	13	15%	13	18%
TD	5	18%	10	12%	10	14%
			Ethnicity			
NR	21	75%	52	61%	47	64%
W	4	14%	17	20%	12	16%
AA	4	14%	10	12%	9	12%
H	0	0%	0	0%	0	0%
AI/AN	0	0%	0	0%	0	0%
A	0	0%	0	0%	0	0%
NH/OPI	0	0%	0	0%	0	0%
Mixed	2	7%	6	7%	6	8%
			SES			
NR	26	93%	75	88%	66	89%
High	0	0%	0	0%	0	0%
Low	4	14%	6	7%	4	5%
Mixed	2	7%	4	5%	4	5%
			Gender			
NR	5	18%	20	23%	17	23%
M	16	57%	34	40%	31	42%
F	13	46%	20	24%	19	26%
Mixed	6	21%	11	13%	9	12%
			Instructional Setting			
NR	2	7%	7	8%	3	4%
Gen Ed	15	54%	34	40%	29	39%
SpEd	10	36%	40	47%	39	53%
Mixed	2	7%	4	5%	4	5%
			Grade Level			
NR	4	14%	8	9%	8	11%
PK-2	16	57%	37	44%	33	45%
3-5	3	11%	11	13%	9	12%
6-8	5	18%	7	8%	6	8%
9-12	2	7%	9	11%	9	12%
Mixed	2	7%	13	15%	10	14%

Note: n = participants; k = cases; NR = not reported; HI = high incidence disability, LI = low incidence disability, TD = typically developing, W = White; AA = African-American; H = Hispanic; AI = American Indian; AN = Alaskan Native; A = Asian, NH = Native Hawaiian; OPI = Other Pacific Islander, M = Male; F = Female; Gen Ed = general education; SpEd = special education.

school (n = 9) classrooms. In a few cases, multiple grade levels were reported but not referenced to individual students (n = 13), and in the remaining cases no grade level was reported (n = 8). A summary of the participant and setting characteristics at the study-, participant-, and case-level are reported in Table 2.

Intervention characteristics. The procedures for implementing praise were fairly consistent across all studies. A functional behavior assessment was conducted before implementing praise in only one study. In the majority of the studies, a teacher implemented the intervention (n = 23). The interventionist typically received brief training (n = 12) or no training (n = 11). In most studies, the interventionist directed praise at an individual student (n = 19). The target of the intervention was generally unaware of the contingency in place for receiving praise (n = 27). Most studies (n = 25) did not report if the praise was delivered during activity that was of high or low interest to the participants. Praise was typically implemented without any antecedent manipulations (n = 16) or consequences (n = 20). The most frequent antecedent manipulations paired with praise was prompting (n = 6) and the only consequence paired with praise was ignoring (n = 7). A majority of the studies employed an engagement contingency (n = 21). A variable ratio (n = 17) was the most common schedule of reinforcement.

In general, the studies did not report the language of the praise statements with enough detail to determine the specificity (n = 12), control (n = 22) or attribution (n = 21). Of the studies that reported specificity, most reported using behavior specific praise, alone (n = 6) or in combination with general praise (n = 7). Additionally, most of the

studies reporting specific language were informational (n = 5) instead of controlling. Studies where attribution could be determined were evenly split between ability-focused praise (n = 3) and effort-focused praise (n = 3).

The majority of the studies did not report treatment fidelity (n = 16) or social validity (n = 21). A summary of the intervention characteristics at the study-, participant-, and case-level are provided in Table 3.

Table 3. Summary of intervention characteristics

Level	# of Studies	% of Studies	n	% of Participants	k	% of Cases
Function-Based Intervention						
Y	1	4%	8	9%	8	11%
N	27	96%	77	91%	66	89%
Interventionist						
Teacher	23	82%	75	88%	64	86%
Peer	1	4%	8	9%	8	11%
Researcher	5	18%	10	12%	10	14%
Interventionist Training						
NR	2	7%	10	12%	10	14%
None	11	39%	28	33%	25	34%
Brief	12	43%	42	49%	36	49%
In-Depth	4	14%	13	15%	11	15%
Recipient						
Universal	7	25%	17	20%	17	23%
Targeted	4	14%	9	11%	7	9%
Individual	19	68%	59	69%	50	68%
Vicarious	1	4%	8	9%	8	11%
Expected						
Y	1	4%	1	1%	1	1%
N	27	96%	84	99%	73	99%
Activity						
NR	25	89%	77	91%	69	93%
HI	0	0%	0	0%	0	0%
LI	3	11%	8	9%	5	7%

Table 3. Continued

Level	# of Studies	% of Studies	n	% of Participants	k	% of Cases
			Antecedent Manipulations			
None	16	57%	54	64%	48	65%
Rules	3	11%	6	7%	6	8%
Prompt	6	21%	14	16%	11	15%
Instructional	3	11%	10	12%	8	11%
Multiple	1	4%	1	1%	1	1%
			Additional Consequences			
None	20	71%	56	66%	52	70%
Ignore	7	25%	24	28%	20	27%
	1	4%	5	6%	2	3%
			Contingency			
NR	2	7%	4	5%	4	5%
NC						
EC	21	75%	72	85%	61	82%
CC	3	11%	4	5%	4	5%
PC	1	4%	1	1%	1	1%
Mixed	1	4%	4	5%	4	5%
			Schedule of Reinforcement			
CR	1	4%	4	5%	2	3%
FR	3	11%	6	7%	6	8%
VR	17	61%	50	59%	46	62%
FI	3	11%	11	13%	8	11%
VI	2	7%	4	5%	2	3%
Mixed	2	7%	10	12%	10	14%
			Specificity			
NR	12	43%	37	44%	33	45%
BSPS	6	21%	19	22%	19	26%
General	3	11%	15	18%	8	11%
Mixed	7	25%	14	16%	14	19%
			Control			
NR	22	79%	66	78%	58	78%
Control	0	0%	0	0%	0	0%
Inform	5	18%	17	20%	14	19%
Mixed	1	4%	2	2%	2	3%
			Attribution			
NR	21	75%	66	78%	58	78%
Ability	3	11%	7	8%	7	9%
Effort	3	11%	10	12%	7	9%
Mixed	1	4%	2	2%	2	3%

Table 3. Continued

Level	# of Studies	% of Studies	n	% of Participants	k	% of Cases
			Treatment Fidelity			
Y	12	43%	52	61%	44	59%
N	16	57%	33	39%	30	41%
			Social Validity			
NR	21	75%	61	72%	55	74%
Y	5	18%	18	21%	16	22%
N	2	7%	6	7%	3	4%

Note: n = participants; k = cases; Y = yes; N = no; NR = not reported; HI = high interest; LI = low interest, NC = noncontingent; EC = engagement contingent; CC = completion contingent; PC = performance contingent; CR = continuous ratio; FR = fixed ratio; VR = variable ratio; FI = fixed interval; VI = variable interval; BSPS = behavior specific praise statements.

Outcome characteristics. Table 4 provides a summary of outcome characteristics at the study-, participant-, and case level. Overwhelmingly, the studies focused on increasing student behaviors (n = 22); either social (n = 21) or academic (n = 4). The outcomes are generally reported as either a frequency count of discreet behaviors (n = 10) or a percentage of intervals in which the target behavior occurred (n = 17). In most cases, the effects of the intervention were measured for individual participants (n = 20); however, some studies reported effects for small groups (n = 4) or the whole class (n = 5).

Table 4. Summary of outcome characteristics.

Level	# of Studies	% of Studies	n	% of Participants	k	% of Cases
			DV Direction			
Increase	22	79%	65	76%	57	77%
Decrease	10	36%	28	33%	25	34%

Table 4. Continued

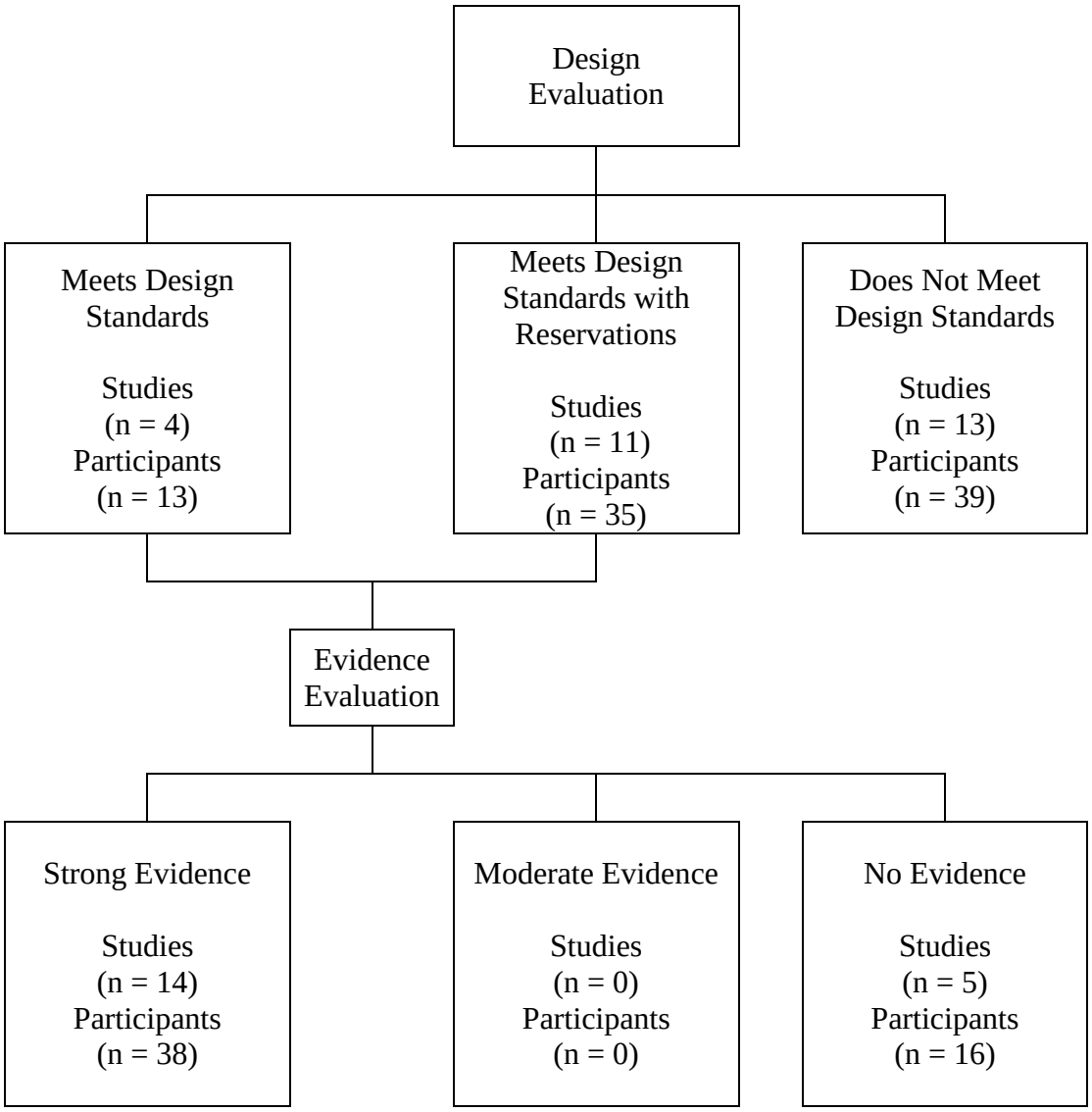
Level	# of Studies	% of Studies	n	% of Participants	k	% of Cases
		DV Type				
ASB	21	75%	64	75%	56	76%
ISB	8	29%	26	31%	23	31%
AB	4	14%	7	8%	7	9%
		DV Measure				
Frequency	10	36%	15	18%	15	20%
Interval	17	61%	58	68%	56	76%
Duration	2	7%	11	13%	4	5%
Latency	0	0%	0	0%	0	0%
Percent of Items	2	7%	6	7%	4	5%
		Unit of Analysis				
Whole Class	5	18%	9	11%	9	12%
Small Group	4	14%	9	11%	7	9%
Individual	20	71%	67	79%	58	78%

Note: n = participants; k = cases; ASB = appropriate social behavior; ISB = inappropriate social behavior; AB = academic behavior.

WWC Standards

Design standards. Figure 2 provides a summary of the results from the application of the WWC design and evidence standards. Half of the studies met the design standards with (n = 11) or without (n = 4) reservations. These studies included 48 participants and 43 cases. The remaining studies did not meet the standards for design quality (n = 13). Half of the studies that failed to meet the design standard did so due to a failure to meet the standard for interobserver agreement (n = 7). The other half failed to meet the standard for providing a minimum of three opportunities for replication of effects (n = 6).

Figure 2. Overview of the WWC design and evidence standards application.



Evidence standards. The studies that met the design standards, with or without reservations, were evaluated using the standards for evidence of effect. Evidence of effect was evaluated at the case-level. All 14 studies that met the design standards included at least one case demonstrating strong evidence of effect; however, five studies

also included cases (n = 12) that did not demonstrate any evidence of effects. The studies demonstrating strong evidence of effects included 38 participants (32 cases) and were conducted by 10 unique research groups.

Key Features of Studies Demonstrating Evidence of Effects

Table 5 summarizes the key features of studies demonstrating evidence of effects at the study-, participant-, and case-level. The participants involved in the studies were either identified as typically developing (n = 5) or were as students with high-incidence disabilities (n = 20). More participants were male (n = 17) than female (n = 6). Additionally, two studies conducted interventions on class groups with a mix of genders (n = 20). Additionally, two studies conducted interventions on class groups with a mix of genders (n = 7). The majority of participants were equally divided between general education (n = 17) and special education (n = 18) settings in early (n = 10) or upper (n = 11) elementary classrooms. Across the studies, praise was typically implemented without any additional antecedent manipulations or consequences (n = 9). Most studies were designed to employ an engagement contingency (n = 9). A variable ratio schedule of (n = 7) was the most common schedule of reinforcement. Praise statements delivered as an intervention were either behavior specific (n = 2) or a mix of behavior specific and general praise statements (n = 6). Studies examined the effects of praise on appropriate social behaviors (n = 8), inappropriate social behaviors (n = 6), and academic behaviors (n = 3) either alone or in combination.

Table 5. Key features of studies demonstrating evidence of effect.

Level	# of Studies	% of Studies	n	% of Participants	k	% of Cases
			Disability Status			
NR	4	29%	13	34%	7	22%
HI*	6	43%	20	53%	20	62%
LI	0	0%	0	0%	0	0%
TD	4	29%	5	13%	5	16%
Mixed	0	0%	0	0%	0	0%
			Gender			
NR	3	21%	7	18%	3	9%
M	6	43%	14	37%	14	44%
F	5	36%	6	16%	6	19%
Mixed	2	14%	7	18%	5	16%
			Instructional Setting			
NR	1	7%	1	3%	1	3%
Gen Ed	7	50%	17	45%	11	34%
SpED	5	36%	18	47%	18	56%
Both	1	7%	2	5%	2	6%
			Grade Level			
NR	3	21%	7	18%	7	22%
PK – 2*	6	43%	10	26%	10	31%
3 – 5*	3	21%	11	29%	9	28%
6 – 8	1	7%	1	3%	1	3%
9 - 12	1	7%	4	11%	4	12%
Mixed	1	7%	5	13%	1	3%
			Intervention Components			
Alone*	9	64%	25	66%	23	72%
Combined*	5	36%	13	34%	9	28%
			Schedule of Reinforcement			
NR	0	0%	0	0%	0	0%
CR	0	0%	0	0%	0	0%
FR	3	21%	6	16%	6	19%
VR*	7	50%	12	32%	12	38%
FI	2	14%	9	24%	5	16%
VI	1	7%	3	8%	1	3%
Mixed	1	7%	8	21%	8	25%
			Contingency			
NR	1	7%	1	3%	1	3%
NC	0	0%	0	0%	0	0%
EC*	9	64%	30	79%	24	75%
CC	2	14%	2	5%	2	6%
PC	1	7%	1	3%	1	3%
Mixed	1	7%	4	11%	4	13%

Table 5. Continued

Level	# of Studies	% of Studies	n	% of Participants	k	% of Cases
			Specificity			
NR	4	29%	8	21%	6	19%
BSPS	2	14%	12	32%	12	37%
General	2	14%	9	24%	5	16%
Mixed	6	43%	9	24%	9	28%
			DV Type			
ASB*	8	57%	21	55%	19	59%
ISB*	6	43%	20	53%	16	50%
AB	3	21%	5	13%	5	16%

Note. k = number of cases, *= evidence-based practice; NR = not reported; HI = high incidence disability, LI = low incidence disability; TD = typically developing, M = male; F = female; Gen Ed = general education; SpEd = special education; CR = continuous ratio; FR = fixed ratio, VR = variable ratio, FI = fixed interval, VI = variable interval; NC = noncontingent; EC = engagement contingent, CC = completion contingent, PC = performance contingent, BSPS = behavior specific praise statement, ASB = appropriate social behavior, ISB = inappropriate social behavior, AB = academic behavior.

Discussion

The purpose of this review was to determine if there is sufficient empirical support to recommend praise as an evidence-based strategy for managing student behavior in classroom settings based on the WWC protocol for evaluating single-case research. The following research questions were posed: (a) With whom and under what conditions have the effects of praise on student behavior been studied?, (b) Is there sufficient empirical support to recommend praise as an evidence-based strategy?, (c) For whom and under what conditions is there sufficient evidence to recommend praise as an evidence-based practice?

The first research question focused on the conditions under which the effects of praise on student behavior have been studied. Given the quantity of single case studies, it

is not surprising that the effects of praise have been studied with a wide variety of students in all educational settings. Praise has typically been provided by teachers to specific students contingent on a variety of student behaviors. Praise has been used with and without additional antecedent manipulations and consequence interventions. A variety of schedules of reinforcement and contingencies have been employed. The results clearly demonstrate the versatility of praise.

The second research question focused on determining if sufficient empirical evidence exists to recommend praise as an evidence-based practice in classroom settings. This review found sufficient empirical evidence from studies employing single case designs to recommend praise as an evidence-based practice in classroom settings. Praise has strong evidence of effects across varied student and setting characteristics. Additionally, some of the dimensions of praise (e.g., reinforcement schedule, contingency, specificity) are clearly associated with strong evidence of effects. Finally, these findings show strong evidence of the effects of praise on multiple forms of student behavior (e.g., appropriate social behavior, inappropriate social behavior, academic behavior).

The third research question focused on determining more specifically, for whom and under what conditions there is sufficient empirical evidence to recommend praise as an evidence-based practice. Cases were further reviewed to determine participant, setting, intervention, and outcome variable characteristics to which the effects of praise are most likely to generalize. Sufficient empirical evidence exists to recommend praise as an evidence-based practice for students with high-incidence disabilities and students

in elementary classrooms. Further, praise can be recommended as an evidence-based practice for increasing appropriate social behaviors and decreasing inappropriate social behaviors. Additionally, the results of this study show strong evidence of the effects of praise on typically developing students of both genders; however the evidence is not sufficient to recommend praise as an evidence-based practice under those conditions. Similarly there is strong evidence for the effects of praise in both general and special education settings, as well as in middle school and high school settings, but it cannot be recommended as an evidence-based practice in any of these settings. Finally, there is no evidence of the effects of praise on individuals with low-incidence disabilities.

The results of this study support the idea that praise is a complex reciprocal interaction (Brophy, 1981). There is sufficient empirical evidence to recommend the use of praise alone or in combination with antecedent manipulations and other consequences. Additionally, there is sufficient evidence to recommend the use of a variable ratio schedule of reinforcement and to employ engagement contingencies. While there is strong evidence of effects for fixed ratio, fixed interval, and variable interval schedules of reinforcement, none of them can be recommended as an evidence-based practice. The same is true for employing completion contingencies, performance contingencies, or a mix of contingencies. Further, there is strong evidence of effects for both behavior specific praise and general praise, alone or in combination. However, there is no evidence to support the use of continuous ratio reinforcement schedules or noncontingent praise in classroom environments.

Limitations

This study provides a systematic evaluation of the single case research examining the effects of praise on student behavior. However, the findings should be considered within the context of the following limitations. While a systematic literature search was conducted, it is possible that studies eligible for inclusion were omitted. Additionally, requiring studies to be published in peer-reviewed journals may create a positive bias, as studies with no evidence of effects may not have been considered for this study. While the application of the WWC standards ensures a minimum level of methodological rigor, the standards focus on internal validity and may not capture the quality of all aspects of a single case design (Wendt & Miller, 2012). Finally, the evidence of effect ratings cannot be interpreted as a magnitude of effect. Therefore, it is possible to have strong evidence of modest effect.

Implications for Research

This is the first systematic review of single case studies examining the effects of praise of student behavior. This work contributes to the existing knowledge of the effects of praise by establishing praise as an evidence-based practice in classroom settings and by providing detailed information regarding to whom and under what conditions the effects of praise generalize. However, additional research is warranted. Future studies should report student demographic characteristics (i.e., disability status, gender, socioeconomic status, and ethnicity) in more details. Conducting additional studies and reporting these characteristics will allow further synthesis of the findings and will lead to more precise recommendations regarding for whom and under which conditions praise

can be recommended as an evidence-based practice. Further, studies are needed examining the effects of praise on student behavior when the unit of analysis is small group or whole class. Additionally, future research should continue to examine the effects of praise on academic behavior.

Existing studies examining the effects of praise on student behavior have provided sufficient empirical evidence to make basic recommendations regarding the dimensions of praise used in behavioral interventions. Future, studies can help develop more precise recommendations by reporting implementation procedures in greater detail (i.e., inclusion of functional behavior assessment, the contingency employed, and the language used in the praise statements). Future studies should also compare and contrast different schedules of reinforcement, contingencies, and types of language used in praise statements.

Implications for Practice

This research supports use of praise as a behavioral intervention in classroom settings, in general. Teachers seeking behavioral interventions for elementary school students or students with high incidence disabilities should be strongly encouraged to implement praise. Across settings, teachers should be encouraged to implement praise using a variable ratio schedule of reinforcement and engagement contingencies. Individuals providing training to pre-service and in-service teachers should ensure that teachers understand that while praise cannot be recommended as an evidence-based practice in all settings, there are very few settings in which there is anything less than strong evidence of effect on student behaviors. In order to support appropriate

implementation of praise, teachers should receive training in the behavioral mechanisms of praise. Teachers should also receive ongoing coaching and support to ensure high fidelity implementation of praise in classroom settings.

CHAPTER III

THE EFFECTS OF PRAISE ON STUDENT BEHAVIOR: A META-ANALYSIS OF THE SINGLE CASE RESEARCH

Introduction

The body of research using single-case designs to examine the effects of praise on student behavior provides sufficient empirical support to recommend praise as an evidence-based practice. Previous reviews of the effects of praise suggest that several variables may moderate the effects of praise (Bayat, 2011; Brophy, 1981; Henderlong & Lepper, 2002). Variables that moderate the effects of praise on intrinsic motivation have been examined in meta-analyses (Cameron & Pierce, 1994; Deci et al., 1999). However, there has not been a corresponding examination of the variables that may moderate the effect of praise on behavior.

Research recommends teachers to use praise with individual students in order to achieve specific academic and behavioral outcomes (Conroy, Sutherland, Snyder, Al-Hendawi, & Vo, 2009; Musti-Rao & Haydon, 2011). As a positive reinforcer, praise increases rates of on-task behavior (Conroy et al., 2009; Partin, Robertson, Maggin, Oliver, & Wehby, 2010; Stormont & Reinke, 2009) and increases task engagement (Conroy et al., 2009; Partin et al., 2010). In turn, increasing on-task behavior and task engagement leads to increased work completion and increased correct responses (Conroy et al., 2009; Partin et al. 2010). Praise also increases rule following behavior (Marchant & Anderson, 2012) and compliance with directions (Partin et al., 2010). As a

result, students engage in fewer disruptive behaviors (Partin et al., 2010; Stormont & Reinke, 2009) and receive fewer office disciplinary referrals (Marchant & Anderson, 2012). Finally, praise is linked to an increase in intrinsic motivation (Conroy et al., 2009; Stormont & Reinke, 2009).

Additionally, researcher suggests teachers should use praise as a classwide behavior support. When a positive learning environment is established (Musti-Rao & Haydon, 2011) and, teachers can spend less time managing problem behavior and more time on instruction (Conroy et al., 2009; Musti-Rao & Haydon, 2011). Effective use of praise builds positive teacher-student relationships (Marchant & Anderson, 2012; Musti-Rao& Haydon, 2011) and improves overall classroom climate (Conroy et al., 2009; Partin et al., 2010). Additionally, using specific praise identifies peers engaging in appropriate behavior as models, which can prevent inappropriate behavior (Marchant & Anderson, 2012; Musti-Rao & Haydon, 2011; Stormont & Reinke, 2009).

In the last decade, consultants and coaches have focused on getting teachers to increase their rates of praise (Conroy et al., 2009; Marchant & Anderson, 2012; Musti-Rao & Haydon, 2011; Partin et al., 2010; Stormont & Reinke, 2009). Research indicates that all students in a classroom should be receiving praise regularly (Conroy et al., 2009). Some research recommends particular rates of praise or ratios of praise to redirection or reprimands (Musti-Rao & Haydon, 2011); however, students' need for praise may vary (Conroy et al., 2009).

Beyond the rate or ratio of praise statements delivered, several factors may function as moderators. For example characteristics of the individual receiving praise or

the setting in which praise is provided may moderate the effects of praise. Student characteristics that may moderate praise include: (a) student socioeconomic status, (b) ability level, (c) developmental level, (d) gender, (e) disability status, (f) English language proficiency, (g) age, (h) function of the behavior, (i) perception of praise as a reinforcer, and (j) previous reinforcement history (Conroy et al., 2009; Marchant & Anderson, 2012; Musti-Rao & Haydon, 2011; Partin et al., 2010). Praise may be more effective in classrooms where rules are established and reviewed often (Marchant & Anderson, 2012) and a positive climate has been established (Conroy et al., 2009), or in schools implementing School-wide Positive Behavior Interventions and Supports (PBIS; Musti-Rao & Haydon, 2011). The topography of the praise, itself, may also serve as a moderator. Examples of the components of praise that may impact the magnitude of its effect include: (a) consistency of use, (b) specificity of the praise statement, (c) contingency of the praise statement , and (d) attribution of the praise statement (i.e., ability or effort) (Brophy, 1981; Partin et al., 2010; Stormont & Reinke, 2009). Finally, the topography of the target behavior may moderate the effects of praise (Gable et al., 2009).

Previous Reviews

Two previous meta-analyses examine the effects praise on intrinsic motivation (Cameron & Pierce, 1994; Deci et al., 1999). Primary sources included in both of these meta-analyses employ group designs. Moderators examined include: (a) type of reward, (b) reward expectancy, (c) reward contingency, (d) value of target task, (e) controlling versus informational language. Both meta-analyses conclude that verbal rewards

improve intrinsic motivation, while tangible rewards decrease intrinsic motivation. Additionally, Deci et al. (1999) concluded that the language included in the praise statement moderates the effects of praise.

One previous meta-analysis examines the effects of nonaversive classroom procedures, including praise, on student behaviors (Skiba, Casey, & Center, 1985). Studies included in this meta-analysis employed single case designs. A regression-based approach was used to synthesize results across studies. The moderators examined included: (a) type of treatment, (b) topography of target behavior, (c) administrative arrangements, (d) interventionist, and (e) setting. The results of the meta-analysis indicated that feedback treatments, such as praise, are more effective: (a) for individuals, (b) for decreasing inappropriate social behaviors (e.g., disruptive behavior, off-task behavior), and (c) in special education settings.

The number of studies employing single-case designs to examine the effects of praise on student behavior has almost doubled since the Skiba et al. (1985) meta-analysis was conducted. Additionally, methodological procedures for determining evidence-based practices and conducting meta-analysis of single-case research have progressed. Therefore, an updated meta-analysis of the single case research examining the effects of praise on student behavior is needed to provide recommendations for implementation of praise in classroom settings.

Meta-Analysis of Single-Case Designs

Meta-analysis of studies employing single-case designs can provide a unique contribution to recommendations for implementing evidence-based practices (Burns,

2012). Meta-analysis allows researchers to systematically examine, evaluate, and synthesize the results of studies employing single-case designs, thus enabling researchers to examine the external validity of a particular intervention approach (Scruggs & Mastropieri, 1998; Shadish et al., 2008). Unlike narrative reviews, meta-analyses avoid treating all evidence equally and misrepresenting conclusions (Burns, 2012).

Purpose and Research Questions

Praise is a complex reciprocal interaction between teachers and students with many potential moderators (Brophy, 1981, Dweck, 2007; Gable et al., 2009; Shores, Gunter, & Jack, 1993). The purpose of the study is to conduct a meta-analysis of existing research employing single-case designs to: (a) estimate the magnitude of effect for praise on student behavior, and (b) examine variables that potentially moderate the effects of praise. The meta-analysis will specifically address the following the questions: (a) What is the overall effect of praise on student behavior?, and (b) For whom and under what conditions do the effects of praise generalize?

Method

The study extends the systematic literature review reported in Chapter II to determine if praise is an evidence-based practice. The studies included in this meta-analysis were identified using the methods described in the Chapter II. Three studies included in the systematic literature review were excluded because data could not be extracted from the published time series graphs. The meta-analysis was conducted in five stages: (a) phase contrast identification, (b) data extraction, (c) effect size estimation, and (e) moderator analysis.

Phase Contrast Identification

The studies included in the meta-analysis employed the following single-case experimental designs: (a) case study, (b) reversal, (b) multiple baseline, and (c) alternating treatment. For each time-series graph, individual phase contrasts were identified. For case studies, the baseline phase was contrasted with the intervention phase. For reversal designs, each baseline phase was contrasted with the adjacent intervention phases (i.e., A1 to B1, B1 to A2, and A2 to B2). For multiple baseline designs, the baseline phase was contrasted with the intervention phase for each participant. If a case study, reversal or multiple baseline design introduced components the intervention package sequentially, then the phase preceding the introduction of praise was used as the baseline phase. For alternating treatment designs, the baseline condition was contrasted with each treatment condition. If a study included a maintenance condition, then two additional phase contrasts were calculated: (a) intervention to maintenance and (b) baseline to maintenance.

Data Extraction

Data used in the calculations was extracted from the published short-time series graphs. In preparation for data extraction, the original graphs were digitally clipped from the PDF files using MS Word. The resulting images were saved as JPEG files. Each graph was opened using Plot Digitizer software. Within the Plot Digitizer interface, the X- and Y-axes were calibrated. Then, each data point was identified by clicking on it. The software then generate X and Y values for each data point. The extracted values

were copied and pasted into an MS Excel spreadsheet. In the spreadsheet, the data points were coded, as needed, for phase contrast calculations.

Effect Size Estimation

Percent of nonoverlapping data (PND). PND is a nonparametric, nonoverlap effect size regularly used in meta-analysis of single-case designs (Maggin et al., 2011; Parker, Vannest & Davis, 2011; Scruggs & Mastropieri, 1998). PND quantifies the nonoverlap of data between phases (Parker, Vannest & Davis, 2011). PND can be interpreted as the percentage of data points in a given phase that exceed the highest point in the previous phase (Parker, Vannest & Davis, 2011). Scruggs and Mastropieri (1998) provide the following interpretation guidelines for PND: (a) below 50%, ineffective intervention; (b) between 50% and 70%, questionable intervention effects; and (c) over 90%, very effective intervention. PND correlates strongly with the visual analysis of short time series graphs (Scruggs & Mastropieri, 1998).

Improvement rate difference (IRD). IRD is a nonparametric, nonoverlap statistic increasingly used in meta-analysis of single-case designs (Campbell, 2013). IRD quantifies the difference in improvement rates between two phases (Parker, Vannest, & Davis, 2011). IRD is an adaptation of risk-reduction or risk difference, used in medical intervention studies (Parker, et al., 2009). IRD can be interpreted as the percentage of improvement from one phase to another (Parker, et al., 2009). IRD has a known sampling distribution, which allows for the calculation of confidence intervals (Parker et al., 2009). Parker et al. (2009) suggest the following guidelines for interpreting IRD: (a) below .50, questionable effects; (b) between .50 and .70, moderate effects; and (c) above

.70, large effects. Like PND, IRD aligns with visual analysis of short time series graphs (Parker et al., 2009).

Tau-U. Tau-U is a nonparametric, nonoverlap effect size suggested for use in meta-analysis of single-case designs (Parker & Vannest, 2012). Tau-U quantifies the difference in trend and level between two phases (Parker, Vannest, Davis & Sauber, 2011). Tau-U is adapted from Kendall's rank correlation and the Mann-Whitney U test between groups (Parker, Vannest, Davis, & Sauber, 2011). Tau-U can be interpreted as the percent of data showing improvement (Parker, Vannest, Davis & Sauber, 2011). Like IRD, Tau-U has a known sampling distribution (Parker, Vannest, Davis, & Sauber 2011). While there are no published interpretation guidelines for Tau-U, it is reasonable to use the guidelines provided for PND because both metrics are interpreted as percentages.

Standardized mean difference (SMD). SMD is a parametric effect size routinely used meta-analysis of group designs that has been adapted to the meta-analysis of single case designs (Busk & Serlin, 1992). SMD quantifies the difference in the mean of two data sets, such as data in the baseline phase and data in the intervention phase (Busk & Serlin, 1992; Higgins & Green, 2011). SMD can be interpreted as the number of standard deviation units of difference between the data sets (Busk & Serlin, 1992). SMD is normally distributed, which allows for the calculation of confidence intervals. Cohen (1988) provided the following interpretation guidelines for SMD: (a) 0.2, small effects; (b) 0.5, medium effects; and (c) 0.8 large effects.

Phase contrast calculations. All four effect size estimates were calculated for all phase contrasts. Additionally, standard errors were calculated for IRD, Tau-U, and SMD estimates. The phase contrasts for PND were calculated using the SCDA package in the R platform. The phase contrasts for IRD were calculated in two steps. Initially, the extracted data was entered into an online calculator to identify the number of data points removed from each phase. Then, the proportions were entered into WINPepi to obtain IRD values and standard error. The phase contrasts for Tau-U and SMD were calculated using online calculators. The resulting effect size estimates and standard errors were input into an excel spreadsheet for moderator analysis. Study- and moderator-level aggregates were calculated.

Moderator Analysis

Descriptive coding from Chapter II was used to examine nine potential moderators to determine if the variables impact the effects of praise on student behavior. The variables were identified based on previous reviews of the literature and recommendations made to teachers in recent publications. The moderators examined included: (a) disability status, (b) gender, (c) educational setting, (d) grade level, (e) intervention components, (f) schedule of reinforcement, (g) contingency, (h) specificity, and (i) type of outcome variable. The full coding manual is provided in Appendix B. Statistically significant differences between levels of moderator variables was determined visually using nonoverlapping 83.4% confidence intervals.

Disability status. This moderator was divided into four levels. Participants with disabilities categorized by IDEA as high incidence (i.e., SLI, LD, EBD) were coded as

high incidence. Participants with disabilities categorized by idea as low incidence (i.e., ID, ASD, multiple disabilities) were coded as *low incidence.* Participants without disabilities were coded as *typically developing*. If the unit of analysis for the study was a group it was coded as *mixed* unless all participants were reported as belonging to one of the previously described categories.

Gender. This moderator was divided into three levels. When individual participant gender was reported they were coded as either *male* or *female*. If the unit of analysis for the study was a group it was coded as *mixed* unless all participants were reported as belonging to one of the previously described categories.

Educational setting. This moderator was divided into three levels. If all study activities took place in general education classrooms, the study was coded *general education.* If all study activities took place in special education classrooms, the study was coded *special education*. If study activities occurred in multiple settings, the study was coded *mixed*.

Grade level. This moderator was divided into five levels. Studies were coded based on the reported grade level of the participant(s) as *PK-2, 3-5, 6-8,* or *9-12*. If the unit of analysis was a group it was coded as *mixed* unless all participants were reported as belonging to one of the previously described categories.

Intervention components. This moderator was divided into two levels. If praise was the sole intervention implemented, the study was coded *alone.* If antecedent manipulations (e.g., stating rules, prompting behavior, altering instructional pace) or

additional consequent interventions (e.g., ignoring) were included in the intervention package, the study was coded *combination.*

Schedule of reinforcement. This moderator was divided into five levels. If the interventionist praised every occurrence of the target behavior, the study was coded *continuous ratio*. If the interventionist praised every *n*th occurrence of the target behavior, the study was coded *fixed ratio*. If the interventionist praised occurrences of the target behavior unpredictably, the study was coded *variable ratio*. If the interventionist praised the first occurrence of the target behavior after time specified amount of time has elapsed, the study was coded *fixed interval*. If the interventionist praised the first occurrence of the target behavior after varying amounts of time have elapsed, the study was coded *variable interval*.

Contingency. This moderator was divided into four levels. If the provision of praise was not related to a target behavior, the study was coded *noncontingent*. If the provision of praise was contingent on beginning, but not necessarily completing the target behavior, the study was coded *engagement contingent*. If the provision of praise was contingent on completing the target behavior, the study was coded *completion contingent*. If the provision of praise was contingent on matching or exceeding a certain performance criterion, the study was coded *performance contingent.*

Specificity. This moderator was divided into three levels. If the praise statements included an operational description of the target behavior, the study was coded *behavior specific*. If the praise statements do not include an operational description of the target

behavior, the study was coded *general*. If the praise statements include both behavior specific praise and general praise, the study was coded *mixed*.

Type of outcome variable. This moderator was divided into three levels. If the intent of the intervention was to increase a nonacademic target behavior, the study was coded *appropriate social behavior*. If the intent of the intervention was to decrease a nonacademic target behavior, the study was coded *inappropriate social behavior*. If the intent of the intervention was to demonstrate mastery of an academic skill, the study was coded *academic behavior*.

Results

Mean effect size estimates for this study were calculated across four indices. Figures 3, 4, 5 and 6 provide study-level aggregate effects for the PND, IRD, Tau-U, and SMD indices, respectively. An overall estimate was calculated using data from all 297 phase contrasts extracted from 76 cases published in 25 studies. Estimates were also calculated to compare with WWC coding for design quality and evidence of effects. Phase contrasts associated with cases demonstrating strong or moderate evidence of effects were separated by type of contrast (i.e., baseline to intervention, intervention to maintenance, baseline to maintenance, generalization) to compare main effects to maintenance and generalization effects. Finally, the phase contrasts associated with cases demonstrated strong or moderate evidence of effects were used to conduct a moderator analysis.

Results for the PND statistic indicate that when all 297 phase contrasts are included in the analysis 57% of data points in the treatment phase that exceed the highest data point

in the baseline phase. Results for the IRD statistic indicate a 75% improvement from treatment to baseline. Results for the Tau-U statistic indicate that 59% of data improve from treatment to baseline. Results for the SMD statistic indicate 5.39 standard deviation units of improvement from baseline to treatment. Studies that meet the WWC design standards without reservations have a PND of 47%, an IRD of 78%, a Tau-U of 58%, and a SMD of 4.47. Studies that meet the WWC design standards with reservations have a PND of 67%, an IRD of 73%, a Tau-U of 64%, and a SMD of 6.46. Studies that do not meet the WWC standards have a PND of 56%, and IRD of 68%, a Tau-U of 54%, and a SMD of 3.74. The effect size estimates associated with the design quality of the case is summarized in Table 6.

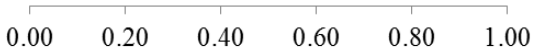
Table 6. Effect size estimates aggregated by WWC standards for design quality.

	WWC Rating			
	Meets without Reservations	**Meets with Reservations**	**Does Not Meet Standards**	**Overall Estimate**
Metric	**M (SE)**	**M (SE)**	**M (SE)**	**M (SE)**
PND	0.47 (--)	0.67 (--)	0.56 (--)	0.57 (--)
IRD	0.78 (0.01)	0.73 (0.02)	0.68 (0.02)	0.75 (0.01)
Tau-U	0.58 (0.03)	0.64 (0.03)	0.54 (0.04)	0.59 (0.02)
SMD	4.47 (0.14)	6.46 (0.08)	3.74 (0.13)	5.39 (0.06)

Note. Estimates are calculated from 297 phase contrasts taken from 76 cases published in 25 studies. M = mean; SE = standard error.

Figure 3. Study-level effects (PND)

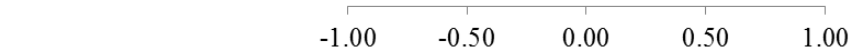
SID	PND	Range			Participants	Contrasts
		Min	ES	Max		
Omnibus		0.00	0.57	1.00	75	297
1		0.70	0.85	1.00	3	5
2		0.00	0.13	0.25	2*	2
3		0.40	0.60	0.80	2	2
4		0.25	0.82	1.00	1	6
5		0.75	0.95	1.00	4	12
6		0.00	0.30	0.60	2	2
7		0.38	0.93	1.00	2	16
8		0.00	0.44	1.00	2	64
10		0.00	0.28	0.75	1*	12
11		0.57	0.81	1.00	1	6
12		0.00	0.50	1.00	1	2
14		0.00	0.62	1.00	8	20
15		0.07	0.48	1.00	5	10
16		0.63	0.97	1.00	4	24
18		0.00	0.44	1.00	3	15
19		0.17	0.58	1.00	1*	2
20		0.00	0.49	1.00	3*	7
21		0.45	0.78	1.00	4	4
22		0.00	0.50	1.00	2	6
23		0.00	0.14	0.50	1	6
24		0.00	0.61	1.00	1	6
25		0.00	0.57	1.00	6	6
26		0.00	0.23	0.75	4	12
27		0.00	0.82	1.00	4*	24
28		0.00	0.25	1.00	8	26



Note: SID = study identification number; ES = effect size; *= group as unit of analysis.

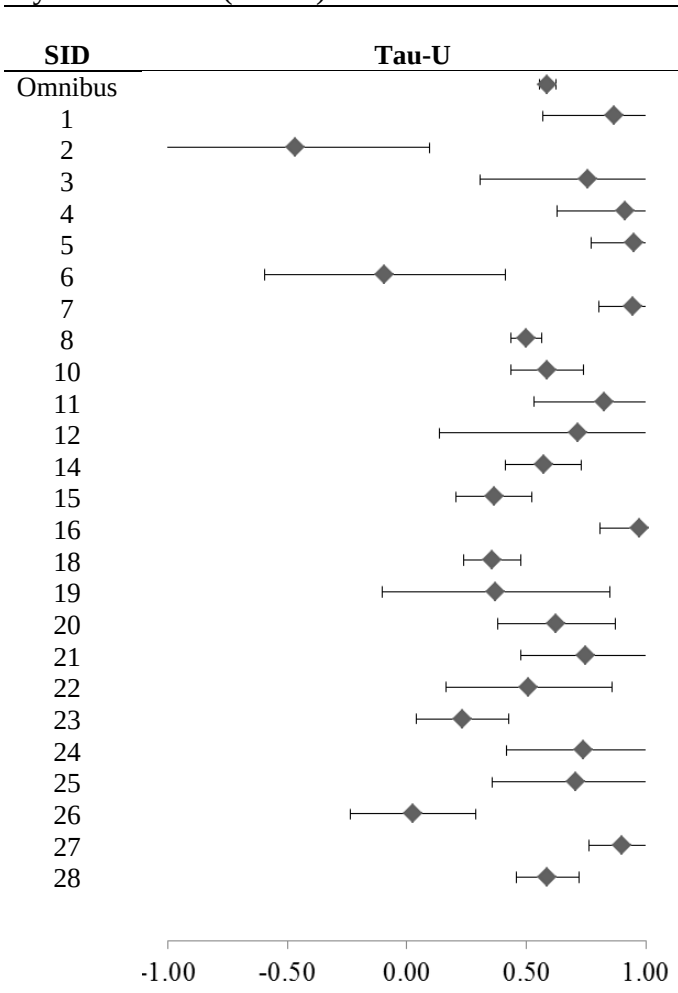
Figure 4. Study-level effects (IRD)

SID	IRD	95% Confidence Interval LL	ES	UL	Participants	Contrasts
Omnibus		0.72	0.75	0.77	75	297
1		0.76	0.87	0.98	3	5
2		0.27	0.59	0.92	2*	2
3		0.53	0.74	0.96	2	2
4		0.66	0.82	0.98	1	6
5		0.71	0.84	0.98	4	12
6		0.20	0.53	0.85	2	2
7		0.69	0.83	0.97	2	16
8		0.73	0.77	0.80	2	64
10		0.63	0.71	0.80	1*	12
11		0.43	0.65	0.87	1	6
12		-0.19	0.50	1.19	1	2
14		0.65	0.74	0.84	8	20
15		0.53	0.61	0.70	5	10
16		0.74	0.87	1.00	4	24
18		0.70	0.77	0.83	3	15
19		0.36	0.71	1.06	1*	2
20		0.63	0.76	0.89	3*	7
21		0.36	0.58	0.80	4	4
22		0.33	0.54	0.75	2	6
23		0.20	0.36	0.51	1	6
24		0.35	0.63	0.91	1	6
25		0.58	0.76	0.94	6	6
26		0.34	0.50	0.66	4	12
27		0.78	0.87	0.96	4*	24
28		0.67	0.74	0.81	8	26



Note: SID = study identification number; LL = lower limit; ES = effect size; UL = upper limit; *= group as unit of analysis.

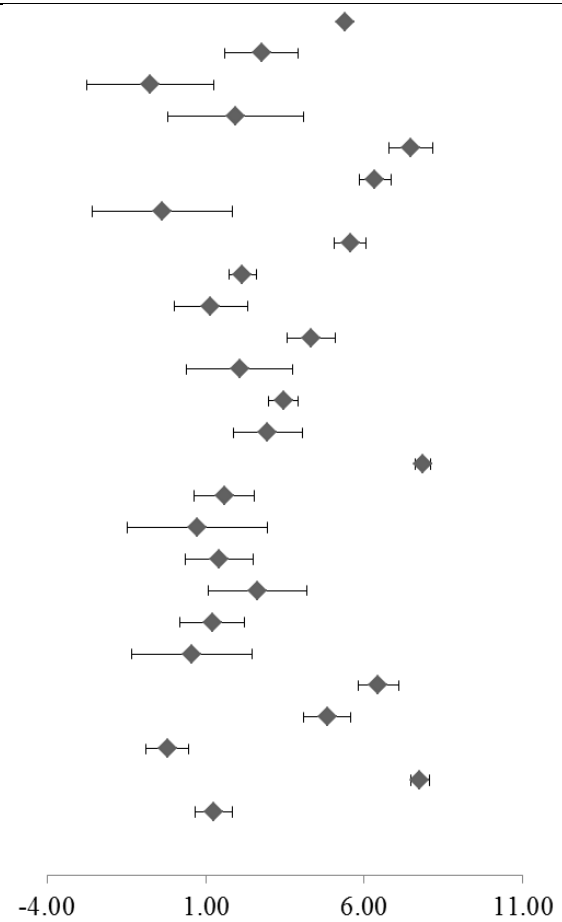
Figure 5. Study-level effects (Tau-U)



SID	Tau-U	95% Confidence Interval LL	ES	UL	Participants	Contrasts
Omnibus		0.55	0.59	0.62	75	297
1		0.57	0.87	1.17	3	5
2		-1.03	-0.47	0.09	2*	2
3		0.31	0.76	1.21	2	2
4		0.63	0.91	1.20	1	6
5		0.77	0.95	1.13	4	12
6		-0.59	-0.09	0.41	2	2
7		0.80	0.95	1.09	2	16
8		0.44	0.50	0.56	2	64
10		0.44	0.59	0.74	1*	12
11		0.53	0.83	1.12	1	6
12		0.14	0.72	1.30	1	2
14		0.41	0.57	0.73	8	20
15		0.20	0.36	0.53	5	10
16		0.81	0.97	1.14	4	24
18		0.24	0.36	0.48	3	15
19		-0.10	0.37	0.85	1*	2
20		0.38	0.63	0.87	3*	7
21		0.48	0.75	1.03	4	4
22		0.16	0.51	0.86	2	6
23		0.04	0.23	0.42	1	6
24		0.42	0.74	1.06	1	6
25		0.36	0.71	1.06	6	6
26		-0.23	0.03	0.29	4	12
27		0.76	0.90	1.04	4*	24
28		0.46	0.59	0.72	8	26

Note: SID = study identification number; LL = lower limit; ES = effect size; UL = upper limit; *= group as unit of analysis.

Figure 6. Study-level effects (SMD)



SID	SMD	95% Confidence Interval LL	ES	UL	Participants	Contrasts
Omnibus		5.27	5.39	5.51	75	297
1		1.61	2.77	3.93	3	5
2		-2.76	-0.75	1.27	2*	2
3		-0.20	1.94	4.08	2	2
4		6.79	7.48	8.17	1	6
5		5.85	6.35	6.85	4	12
6		-2.56	-0.35	1.86	2	2
7		5.05	5.56	6.07	2	16
8		1.72	2.16	2.60	2	64
10		0.02	1.17	2.32	1*	12
11		3.58	4.34	5.10	1	6
12		0.40	2.07	3.75	1	2
14		2.99	3.46	3.93	8	20
15		1.88	2.96	4.04	5	10
16		7.61	7.85	8.10	4	24
18		0.65	1.60	2.55	3	15
19		-1.48	0.74	2.97	1*	2
20		0.35	1.42	2.49	3*	7
21		1.10	2.64	4.19	4	4
22		0.18	1.21	2.24	2	6
23		-1.34	0.57	2.48	1	6
24		5.82	6.44	7.07	1	6
25		4.09	4.84	5.58	6	6
26		-0.88	-0.20	0.48	4	12
27		7.45	7.75	8.04	4*	24
28		0.66	1.24	1.83	8	26

Note: SID = study identification number; LL = lower limit; ES = effect size; UL = upper limit; *= group as unit of analysis.

The effect size estimates associated with the evidence of the effects of the cases is presented in Table 7. An overall effect size estimate was calculated. calculated from 212 phase contrasts extracted from 33 cases published in 15 studies that met the WWC standards for design quality with or without reservations. Results for the PND statistic indicate 58% of data in the treatment phase exceeded the highest data point in the baseline phase. Results for the IRD statistic indicate 76% improvement from treatment to baseline. Results associated with the Tau-U statistic indicate 60% improvement from baseline to treatment. Results associated with SMD statistic indicate 5.94 standard deviation units of improvement from baseline to treatment. Cases that showed strong evidence of effects, according to the WWC standards have a PND of 58%, an IRD of 76%, a Tau-U of 61%, and a SMD of 6.09. No cases demonstrated moderate evidence of effects. Cases that showed no evidence of effects have a PND of 51%, an IRD of 77%, a Tau-U of 51%, and a SMD of 1.88.

Table 7. Effect size estimates aggregated by WWC standards for evidence of effects.

	WWC Rating			
	Strong Evidence	**Moderate Evidence**	**No Evidence**	**Overall Estimate**
Metric	**M (SE)**	**M (SE)**	**M (SE)**	**M (SE)**
PND	0.58 (--)	--	0.51 (--)	0.58 (--)
IRD	0.76 (0.01)	--	0.77 (0.04)	0.76 (0.01)
Tau-U	0.61 (0.02)	--	0.51 (0.09)	0.60 (0.02)
SMD	6.09 (0.07)	--	1.88 (0.38)	5.94 (0.07)

Note. Estimates are calculated from 212 phase contrasts taken from 33 cases published in 15 studies. M = mean; SE = standard error.

The effect size estimates associated with each type of contrasts are summarized in Table 8. Baseline to intervention contrasts have a PND of 70%, an IRD of 79%, a Tau-U of 75%, and an SMD of 6.61. Intervention to maintenance contrasts have a PND of 12%, an IRD of 54%, a Tau-U of -31%, and a SMD of -0.48. Baseline to maintenance contrasts have a PND of 44%, an IRD of 76%, a Tau-U of 62%, and a SMD of 2.72. Generalization contrasts have a PND of 45%, an IRD of 75%, a Tau-U of 64%, and a SMD of 2.92.

Table 8. Effect size estimates aggregated by type of contrast.

	Baseline to Intervention k = 141	**Intervention to Maintenance k = 25**	**Baseline to Maintenance k = 25**	**Generalization k = 21**
Metric	**M (SE)**	**M (SE)**	**M (SE)**	**M (SE)**
PND	0.70 (--)	0.12 (--)	0.44 (--)	0.45 (--)
IRD	0.79 (0.01)	0.54 (0.04)	0.76 (--)	0.75 (0.03)
Tau-U	0.75 (0.03)	-0.31 (0.06)	0.62 (0.06)	0.64 (0.05)
SMD	6.61 (0.08)	-0.48 (0.33)	2.72 (0.29)	2.92 (0.46)

Note. Estimates are calculated from 212 phase contrasts taken from 33 cases published in 15 studies. k = number of contrasts; M = mean; SE = standard error.

Moderator Analysis

The effect size estimates associated with all moderator variables are summarized in Table 9. Moderators examined were (a) disability status, (b) gender, (c) educational setting, (d) grade level, (e) intervention components, (f) schedule of reinforcement, (g)

Table 9. Summary of moderator effects.

		PND		IRD		Tau-U		SMD	
Variable	**k**	**ES**	**SE**	**ES**	**SE**	**ES**	**SE**	**ES**	**SE**
Disability Status									
HI*	80	0.75	-	0.83	0.02	0.83^{2}	0.04	6.52^{2}	0.10
TD	17	0.69	-	0.79	0.03	0.67^{1}	0.08	5.53^{1}	0.28
Gender									
Male	40	0.61	-	0.82	0.02	0.74	0.05	4.72^{3}	0.21
Female	32	0.75	-	0.79	0.04	0.83	0.06	4.71^{3}	0.19
Mixed	27	0.80	-	0.85	0.04	0.89	0.07	$7.59^{1,2}$	0.15
Educational Setting									
Gen Ed	49	0.72	-	0.77^{3}	0.02	0.67^{3}	0.05	$7.03^{2,3}$	0.13
SpEd	64	0.73	-	0.77^{3}	0.03	0.78	0.05	$6.69^{1,3}$	0.10
Mixed	16	0.81	-	$0.89^{1, 2}$	0.03	0.91^{1}	0.06	$3.74^{1,2}$	0.41
Grade Level									
PK-2*	37	0.43	-	0.75^{4}	0.02	$0.58^{2, 3, 4}$	0.06	$1.97^{2, 3, 4}$	0.25
3-5*	35	0.76	-	0.81	0.04	0.82^{1}	0.07	7.48^{1}	0.12
6-8	6	0.82	-	0.82	0.08	0.91^{1}	0.15	7.48^{1}	0.35
9-12	24	0.82	-	0.87^{1}	0.05	0.90^{1}	0.08	7.75^{1}	0.15
Intervention Package									
Alone*	101	0.66	-	0.81^{2}	0.02	0.79^{2}	0.03^{2}	6.08	0.10
Combination*	40	0.80	-	0.74^{1}	0.03	0.63^{1}	0.06^{1}	7.36	0.12
Schedule of Reinforcement									
FR	24	0.88	-	0.80	0.05	$0.92^{2, 3}$	0.07	$6.20^{2, 3, 4}$	0.18
VR*	69	0.72	-	0.82^{3}	0.02	0.76^{1}	0.04	$6.63^{1, 3, 4}$	0.11
FI	29	0.75	-	0.74^{2}	0.03	0.66^{1}	0.08	$7.58^{1, 2, 4}$	0.15
VI	3	0.60	-	0.77	0.09	0.78	0.21	$2.02^{1, 2, 3}$	0.89
Contingency									
EC*	92	0.68	-	0.80	0.02	0.74^{4}	0.04	$6.01^{3, 4}$	0.11
CC	18	0.46	-	0.74	0.04	0.66^{4}	0.08	$5.81^{3, 4}$	0.30
PC	6	0.81	-	0.65	0.11	0.83	0.16	$4.34^{1, 2, 4}$	0.39
Mixed	24	0.97	-	0.87	0.07	$0.97^{1, 2}$	0.09	$7.85^{1, 2, 3}$	0.13
Specificity									
BSPS	40	0.61	-	0.79^{2}	0.03	0.77	0.06	7.06^{2}	0.14
General	17	0.78	-	$0.66^{1, 3}$	0.04	0.59	0.09	$5.97^{1, 3}$	0.25
Mixed	53	0.75	-	0.76^{1}	0.03	0.78	0.05	7.11^{2}	0.11
Outcome Type									
ASB*	70	0.84	-	0.85^{2}	0.02	0.84^{2}	0.04	$6.81^{2, 3}$	0.10
ISB*	50	0.44	-	0.72^{1}	0.02	$0.58^{1, 3}$	0.06	$3.27^{1, 3}$	0.20
AB	21	0.86	-	0.76	0.06	0.89^{2}	0.09	$8.07^{1, 2}$	0.15

Note. k = number of contrasts, ES = effect size, SE = standard error, *= evidence-based practice; HI = high incidence disability, TD = typically developing, FR = fixed ratio, VR = variable ratio, FI = fixed interval, VI = variable interval, EC = engagement contingent, CC = completion contingent, PC = performance contingent, BSPS = behavior specific praise statement, ASB = appropriate social behavior, ISB = in appropriate social behavior, AB = academic behavior.

contingency, (h) specificity, and (i) type of outcome. Appendix C contains forest plots for each moderator analysis.

Disability status. Across all four indices, students with high incidence disabilities experienced stronger effects than typically developing students. Students with high incidence disabilities have a PND of 75%, an IRD of 83%, a Tau-U of 83%, and an SMD of 6.52. Typically developing students have a PND of 69%, an IRD of 79%, a Tau-U of 67%, and a SMD of 5.53. Figure 8 shows the range of PND for each level of the moderator and the 83.4% confidence intervals for the IRD, Tau-U, and SMD. The difference between the levels is not statistically significant for the IRD index, but it is for the Tau-U index. The difference between the levels is statistically significant for the SMD index.

Gender. For the PND and Tau-U indices, female participants experienced stronger effects than male students. For the IRD and SMD indices, male participants experienced stronger effects that female students. Across all 4 indices, groups of mixed gender participants experienced the strongest effects. Male participants have a PND of 61%, an IRD of 82%, a Tau-U of 74%, and an SMD of 4.72. Female participants have a PND of 75%, an IRD of 79%, a Tau-U of 83%, and a SMD of 4.71. Cases in which the gender of the participants is mixed have a PND of 80%, an IRD of 85%, a Tau-U of 89% and a SMD of 7.59. Figure 9 shows the range of PND for each level of the moderator and the 83.4% confidence intervals for the remaining indices. There are no statistically significant differences between the effects for the IRD and Tau-U indices. For the SMD

index, effects for groups of participants with mixed difference is statistically significantly different than effects for either male or female participants

Educational setting. For the PND, IRD and Tau-U indices participants receiving intervention in both general education and special education settings experienced the strongest effects, followed by students in exclusively special education settings, and students in exclusively general education settings. For the SMD index, students in general education settings experienced the strongest effect follow by students in exclusively special education settings, and students receiving the intervention in both general and special education settings. Students in general education settings have a PND of 72%, an IRD of 77%, a Tau-U of 67%, and an SMD of 7.03. Students in special education settings have a PND of 73%, an IRD of 77%, a Tau-U of 78%, and a SMD of 6.69. Students receiving intervention in both general and special education settings have a PND of 81%, an IRD of 89%, a Tau-U of 91%, and a SMD of 3.74. Figure 10 shows the range of PND for each moderator level and the 83.4% confidence intervals for the remaining indices. For the results associated with the IRD index, the effects associated with intervention delivered in both general and special education settings is statistically significantly different from effects associated with intervention delivered in either setting independently. For the results associated with the Tau-U index, the effects associated with intervention delivered in both general and special education settings is statistically significantly different from intervention delivered solely general education settings; however, it is not statistically significantly different from intervention delivered

solely in special educations settings. For the SMD index, effects at each level are statistically significantly different from the effects at both of the other levels.

Grade level. Across all 4 indices participants in middle and high school experienced the strongest effects, followed by participants in upper elementary and participants in lower elementary. Participants in lower elementary grades have a PND of 43%, an IRD of 75%, a Tau-U of 58%, and a SMD of 1.97. Participants in upper elementary grades have a PND of 76%, an IRD of 81%, a Tau-U of 82%, and a SMD of 7.48. Participants in middle school grades have a PND of 82%, an IRD of 82%, a Tau-U of 91%, and a SMD of 7.48. Participants in high school grades have a PND of 82%, an IRD of 87%, a Tau-U of 90%, and a SMD of 7.75. Figure 11 shows the range of PND for each moderator level and the 83.4% confidence intervals for the remaining indices. For the IRD index, there is no statistically significant difference in the effects between the levels. For the Tau-U and SMD indices, there is a statistically significant difference in the effects for participants in early elementary grades as compared to all other levels.

Intervention package. For the IRD and Tau-U indices, stronger effects result when praise is the sole intervention. For the PND and SMD indices, stronger effects result when praise is used in combination with additional intervention components. Cases where praise was used alone have a PND of 66%, an IRD of 81%, a Tau-U of 79%, and a SMD of 6.08. Cases were praise was used in combination with additional intervention components have a PND of 80%, an IRD of 74%, a Tau-U of 63%, and a SMD of 7.36. Figure 12 shows the range of PND for each moderator level and the 83.4%

confidence intervals for the remaining indices. Across all indices the difference between the levels is statistically significantly different.

Schedule of reinforcement. The strength of effects associated with the schedule of reinforcement used varied across the four indices. Cases where a fixed ratio was used have a PND of 88%, an IRD of 80%, a Tau-U of 92%, and a SMD of 6.20. Cases where a variable ratio was used have a PND of 72%, an IRD of 82%, a Tau-U of 76%, and a SMD of 6.63. Cases where a fixed interval was used have a PND of 75%, an IRD of 74%, a Tau-U of 66%, and a SMD of 7.58. Cases where a variable interval was used have a PND of 60%, an IRD of 77%, a Tau-U of 78%, and a SMD of 2.02. Figure 13 shows the range of PND for each moderator level and the 83.4% confidence intervals for the remaining indices. For the IRD index, there is no statistically significant difference between the levels. For the Tau-U index, there is a statistically significant difference between the effects for cases where a fixed ratio was employed and cases where a fixed interval ratio was employed. For the SMD index, there is a statistically significant difference between effect for cases where a fixed interval was employed and all other levels. There is also a statistically significant difference between the effect for cases where a variable interval was employed and all other levels.

Contingency. Across all four indices, use of mixed contingencies was associated with stronger effects. The strength of effects associated with individual contingencies varied across the indices. Cases where an engagement contingency was employed have a PND of 68%, an IRD of 80%, a Tau-U of 74%, and a SMD of 6.01. Cases where a completion contingency was employed have a PND of 46%, an IRD of 74%, a Tau-U of

66%, and a SMD of 5.81. Cases where a performance contingency was employed have a PND of 81%, an IRD of 65%, a Tau-U of 83%, and a SMD of 4.34. Cases where multiple contingencies were employed have a PND of 97%, an IRD of 87%, a Tau-U of 97%, and a SMD of 7.85. Figure 14 shows the range of PND for each moderator level and the 83.4% confidence intervals for the remaining indices. For the IRD index, there is not statistically significant difference in effects between the levels. For the Tau-U index, there is a statistically significant difference between the effects for cases employing multiple contingencies and cases employing engagement contingencies. For the SMD index, there is a statistically significant difference between results for cases employing multiple contingencies and all other levels. There is also a statistically significant difference in effects between cases employing performance contingencies and all other levels.

Specificity. Across all of the indices, except PND, cases where behavior-specific praise or a mix of general and behavior-specific praise was used were associated with strong effects than cases where only general praise was used. Cases where behavior specific praise was used have a PND of 61%, an IRD of 79%, a Tau-U of 77%, and a SMD of 7.06. Cases were general praise was used have a PND of 78%, an IRD of 66%, a Tau-U of 59%, and a SMD of 5.97. Cases were a mix of general and behavior-specific praise was used have a PND of 75%, an IRD of 76%, a Tau-U of 78%, and a SMD of 7.11. Figure 15 shows the range of PND for each moderator level and the 83.4% confidence intervals for the remaining indices. For the IRD and Tau-U indices there is not statistically significant difference in effects between the levels. For the SMD index,

there is a statistically significant difference between the effect for general praise and all other levels.

Outcome type. Across all four indices appropriate social behavior and academic behavior exhibited stronger effects than inappropriate social behavior. Cases where the outcome variable was appropriate social behavior have a PND of 84%, an IRD of 85%, a Tau-U of 84%, and a SMD of 6.81. Cases where the outcome variable was inappropriate social behavior have a PND of 44%, an IRD of 72%, a Tau-U of 58%, and a SMD of 3.27. Cases where the outcome variable was academic behavior have a PND of 86%, an IRD of 76%, a Tau-U of 89%, and a SMD of 8.07. Figure 16 shows the range of PND for each moderator level and the 83.4% confidence intervals for the remaining indices. For the IRD index, there is a statistically significant difference between the effect on appropriate student behavior and inappropriate student behavior. For the Tau-U index there is a statistically significant difference between the effect on inappropriate social behavior and all other levels. For the SMD index the effects for all levels are statistically significantly different.

Discussion

The purpose of this meta-analysis was to determine the magnitude of the effect of praise on student behavior. The following research questions were posed: (a) What is the overall effect of praise on student behavior?, and (b) For whom and under what conditions do the effects of praise generalize?

The first research question focused on estimating the overall effects of praise on student behavior. The current study found that the overall effect of praise on student

behavior is moderate to strong across all effect size metrics. Although these results differ from the findings reported by some previous meta-analysis (Cameron & Pierce, 1994; Deci et al., 1999) they are consistent with those of Cherne (2008) and Skiba, et al. (1985). These findings may be explained by the different outcome measures associated with the reviews. Meta-analyses reporting minimal effects of praise focused on outcomes measures associated with motivation. Meta-analyses showing stronger effects, including the current study, synthesized effects based on measures of social and academic behavior.

The second research question focused on determining for whom and under what conditions the effects of praise generalize. On the whole, moderators associated with participants and setting characteristics suggest that praise can be used across educational environments without concern for adverse effects. In general, students benefit from receiving praise, however, some may experience stronger benefits than others. Disability status moderates the effect of praise on student behavior. Students with high incidence disabilities experience greater percentage of data improving and a larger shift in mean levels of behavior than typically developing students. In contrast, this study finds gender does not moderate the effects of praise on student behavior. Educational setting moderates the effect of praise on student behavior in a more complex manner. Praise results in a larger improvement in the mean level of behavior when implemented in either general education or special education settings, exclusively. However, the opposite is true for the percentage of data improving. Perhaps the most surprising moderator of the effects of praise is grade level. Contrary to expectations, students in

lower elementary grades experience the smallest change in level and the least improvement in data of all age groups. There are no statistically significant differences in effects between the other age groups. This finding may be the due to the fact that studies combining praise with additional reinforcers were excluded. Rather than interpreting this information as a recommendation to stop praise in early elementary settings, it should be seen as support for continuing to provide praise to students as they mature.

The moderating effects of the various characteristics of praise are complex. The presence of additional intervention components increases the change in level, but does not moderate the percentage of data improving. The schedule of reinforcement impacts the effects of praise in a variety of ways. Fixed interval schedules result in the largest improvement in mean difference, while variable intervals result in the smallest shift. Additional, the use of a variable interval schedule of reinforcement results in less improvement in the mean difference than other schedules. Fixed ratio schedules result in a better percentage of data improving between phases. These results should be interpreted with caution, however, because insufficient information exists to determine the frequency of praise in any of the schedules. The type of contingency employed also moderates the effect of praise on student behavior. Using multiple types of contingencies results in stronger improvements in the mean level of behavior than any one contingency in isolation. In opposition to the findings regarding the effect of praise on motivation, employing performance contingencies results in the lowest mean shift. Engagement contingencies and completion contingencies have similar outcomes. The specificity of

the praise statement also moderates the effects of praise on student behavior. General praise results in the lowest percentage of data improvement and the smallest shift in mean difference. There is no statistically significant difference between the use of only behavior specific praise and a combination of behavior specific and general praise.

The moderating effects of the type of behavior outcome are clear. Both appropriate social behavior and academic behavior show a strong increase in the percentage of data improving and the mean level of behavior when praise is implemented. By contrast, praise is least effective in decreasing inappropriate behaviors.

Limitations

While the present study provides additional information to guide the implementation of praise as an intervention to address student behavior, the results are subject to certain limitations. First, reporting of moderator variables was lacking in some primary sources. As a result, the sample size varied across moderator variables which may lead to inflation or deflation in the effect size estimates. Second, all of the phase contrasts included in moderator analysis were taken from studies with individual participants as the unit of analysis. Accordingly, these findings may not readily generalize to the classwide applications. Finally, no standard guidelines exist for selecting or interpreting effect sizes in the meta-analysis of single case studies. Therefore, conclusions drawn from the effect size estimates should be viewed with caution.

Implications for Research

While the results of this study indicate praise is effective for a wide range of participants across educational settings, there is no data available for students with low incidence disabilities. Future research should make a concerted effort to examine the effects of praise on the behavior of participants with such disabilities. Additionally, there was not enough data on participant socioeconomic status and ethnicity to analyze these variables as potential moderators. In future studies, researchers should take care to provide information on these variables. Lastly, all of the effects reported in this meta-analysis are drawn from studies where the individual is the unit of analysis. Future research needs to focus on the effects of praise as a universal behavior support.

Although this study provides preliminary evidence regarding the moderating effects of various characteristics of praise, future researchers should conduct studies that directly compare the various types of schedule of reinforcement and contingency. Future research should also examine the balance between behavior specific praise and general praise to determine a minimum threshold for strong effects. Additionally, there was insufficient data on the language of the praise statements to examine the effects of controlling statement or attributional statements. Single case research is ideally suited to directly compare these aspects of praise in alternating treatment designs.

Implications for Practice

Teachers should be encouraged to use praise across educational settings. Pre-service and in-service teachers would be well served to understand the nuances of praise and its effects on student behavior. Individuals supporting teachers in the

implementation of praise should consider recommending that teachers begin with a fixed interval schedule, unless the target behavior has not yet been acquired. These schedules are associated with high levels of effect and are easy to implement with fidelity if the teacher is provided a prompt (e.g. use of a Motivaider). In the case of teaching new skills, teachers could be coached to begin with a fixed ratio and then transition to a fixed interval as the student begins to exhibit higher levels of the target behavior. Teachers should also be coached to employ a combination of engagement and completion contingencies to improve student behavior outcomes.

CHAPTER IV

CONCLUSION

A controversy has arisen regarding the use of praise in classroom settings (Maag, 2001). The varying conclusions of previous narrative reviews (e.g., Bayat, 2011; Henderlong & Lepper, 2002; Kennedy & Willcutt, 1964) and meta-analyses (e.g., Cameron & Pierce, 1994; Deci et al., 1999) contributed to both sides of the debate. Additionally several variables may moderate the effectis of praise as a behavioral intervention, causing further uncertainty regarding best-practice for implementation (Bayat, 2011; Brophy, 1981; Henderlong & Lepper, 2002). To date, no previous review or meta-analysis has systematically evaluated the single case evidence base examining the effects of praise on student behavior.

Within this dissertation, two studies were conducted to systematically: (a) identify studies using single-case designs to examine the effects of praise on student behavior, (b) identify the characteristics of the participants, settings, intervention topography, and outcome variable topography included in the existing research examining the effects of praise on student behavior, (c) evaluate the identified studies using the quality standards established by the WWC (2014); (d) estimate the magnitude of effect of praise on student behavior, and (e) examine variables that potentially moderate the effects of praise.

In the first study, a systematic literature review and quality appraisal were employed to evaluate the state of the evidence for praise as a behavioral intervention in

classroom settings. The systematic literature review resulted in a summary of the characteristics of the single-case studies examining the effects of praise. Based on the WWC (2014) 5-3-20 standard, the single case evidence base provides sufficient empirical evidence, , to recommend praise as an evidence-based practice in classroom settings. Additionally, there is sufficient evidence to recommend praise as an evidence-based practice for: (a) students with high incidence disabilities, (b) in elementary school settings, (c) alone or in combination with other antecedent manipulations or consequences, (d) on a variable ratio schedule of reinforcement, and (e) contingent on student engagement in the target behavior.

In the second study, a meta-analysis was conducted to estimate the overall magnitude of the effects of praise on student behavior and to determine for whom and under what conditions the effects of praise generalize. The effect of praise on student behavior was strong to moderate across all four effect size indices. Moderator analysis demonstrated that the effects of praise are stronger for: (a) students with high incidence disabilities, (b) students in upper elementary and secondary classrooms, and (c) increasing appropriate social behavior and academic behavior. Further, the effects of praise are stronger when praise is: (a) provided on a fixed interval or fixed ratio schedule, and (b) includes behaviorally specific language.

Taken together, these studies demonstrate that sufficient empirical evidence exists to recommend praise as an evidence-based practice in classroom settings. In most contexts there is strong evidence of the effect of praise on student behavior and the

magnitude of those effects is moderate to large. However, additional research is necessary to examine the moderating effects of the topography of praise.

REFERENCES

"*" Indicates studies included in the systematic review and meta-analysis

Bayat, M. (2011). Clarifying issues regarding the use of praise with young children. *Topics in Early Childhood Special Education, 31,* 2, 121-128.

*Bellipanni, K. D., Tingstrom, D. H., Olmi, D. J., & Roberts, D. S. (2013). The sequential introduction of positive antecedent and consequent components in a compliance training package with elementary students. *Behavior Modification, 37,* 6, 768-89.

Binet, A., & Vaschide, E. (1897). Experiences de forces musculaire et de fond chez les jeunes garcons. *Année Psychology, 4*, 15-63.

*Blaze, J. T., Olmi, D. J., Mercer, S. H., Dufrene, B. A., & Tingstom, D. H. (2014). Loud versus quiet praise: A direct behavioral comparison in secondary classrooms. *Journal of School Psychology, 52,* 4, 349-360.

*Boyd, L. A., Keilbaugh, W. S., & Axelrod, S. (1981). The direct and indirect effects of positive reinforcement on on-task behavior. *Behavior Therapy, 12,* 1, 80-92.

Brophy, J. (1981). Teacher praise: A functional analysis. *Review of Educational Research, 51,* 1, 5-32.

Burns, M. K. (2012). Meta-analysis of single-case design research: introduction to the special issue. *Journal of Behavioral Education, 21,* 3, 175-184.

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill, & J. R. Levin (Eds.), *Single case research design and analysis: New directions for psychology and education* (pp. 187–212). Hillsdale, NJ: Erlbaum.

*Caldarella, P., Christensen, L., Young, K. R., & Densley, C. (2011). Decreasing tardiness in elementary school students using teacher-written praise notes. *Intervention in School and Clinic, 47,* 2, 104-112.

Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research, 64,* 3, 363-423.

Campbell, J. M. (2013). Commentary on PND at 25. *Remedial and Special Education, 34,* 1, 20-25.

Cannella, G. S. (1986). Praise and concrete rewards:Concerns for childhood education. *Childhood Education, 62,* 4, 297.

Cherne, J. L. (2008). *Effects of Praise on Student Behavior in the Classroom* (Doctoral dissertation). Retreived from http://library.tamu.edu/

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Conroy, M. A., Sutherland, K. S., Snyder, A., Al-Hendawi, M., & Vo, A. (2009). Creating a positive classroom atmosphere: teachers' use of effective praise and feedback. *Beyond Behavior, 18,* 2, 18-26.

*Cooke, T. P., & Apolloni, T. (1976). Developing positive social-emotional behaviors: a study of training and generalization effects. *Journal of Applied Behavior Analysis, 9,* 1, 65-78.

*Darch, C., & Gersten, R. (1985). The effects of teacher presentation rate and praise on LD students' oral reading performance. *British Journal of Educational Psychology, 55,* 3, 295-303.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 22*, 113-120.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*, 6, 627-668.

*DiCarlo, C. F., Reid, D. H., & Stricklin, S. B. (2003). Increasing toy play among toddlers with multiple disabilities in an inclusive classroom: A more-to-less, child-directed intervention continuum. *Research in Developmental Disabilities, 24*, 3, 195-209.

Dweck, C. S. (2007). The perils and promises of praise. *Educational Leadership, 65*, 2, 34-39.

Emmer, E. T. (1987). Praise and the instructional process. *The Journal of Classroom Interaction, 23*, 2, 32-39.

Fair III, E. M., & Silvestri, L. (1992). Effects of rewards, competition and outcome on intrinsic motivation. *Journal Of Instructional Psychology, 19*, 1, 3.

Ferguson, M. (2013). Praise: What does the literature say? What are the implications for teachers?. *Kairaranga, 14*, 2, 35-39.

*Ferguson, E., & Houghton, S. (1992). The effects of contingent teacher praise, as specified by Canter's Assertive Discipline programme, on children's on-task behaviour. *Educational Studies, 18*, 1, 83-93.

*Fox, J., Shores, R., Lindeman, D., & Strain, P. (1986). Maintaining social initiations of withdrawn handicapped and nonhandicapped preschoolers through a response-dependent fading tactic. *Journal of Abnormal Child Psychology, 14*, 3, 387-396.

*Gable, R. A., & Shores, R. E. (1980). Comparison of procedures for promoting reading proficiency of two children with behavioral and learning disorders. *Behavioral Disorders, 5*, 2, 102-7.

Gable, R. A., Hester, P. H., Rock, M. L., & Hughes, K. G. (2009). Back to basics: Rules, praise, ignoring, and reprimands revisited. *Intervention in School and Clinic, 44*, 4, 195-205.

*Hall, J. S., & Holmberg, M. C. (1974). The effect of teacher behaviors and food serving arrangements on young children's eating in a day care center. *Child Care Quarterly, 3*, 2, 97-108.

Hall, R. V., Lund, D., & Jackson, D. (1968). Effects of teacher attention on study behavior. *Journal of Applied Behavior Analysis, 1*, 1, 1–12.

Hart, B. M., Reynolds, N. J., Baer, D. M., Brawley, E. R., & Harris, F. R. (1968). Effect of contingent and non-contingent social reinforcement on the cooperative play of a preschool child. *Journal of Applied Behavior Analysis*, *1*, 1, 73–76.

Higgins JPT, Green S (Eds.). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Henderlong, J., & Lepper, M. R. (2002). The effects of praise on children's intrinsic motivation: A review and synthesis. *Psychological Bulletin, 128*, 5, 774-95.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71,* 2, 165.

Individuals With Disabilities Education Act, 20 U.S.C. § 1400 (2004).

*Kazdin, A. E. (1973). The effect of vicarious reinforcement on attentive behavior in the classroom. *Journal of Applied Behavior Analysis, 6,* 1, 71-78.

*Kennedy, C., Jolivette, K., & Ramsey, M. L. (2014). The effects of written teacher and peer praise notes on the inappropriate behaviors of elementary students with emotional and behavioral disorders in a residential school. *Residential Treatment for Children & Youth, 31,* 1, 17-40.

Kennedy, W. A., & Willcutt, H. C. (1964). Praise and blame as incentives. P*sychological Bulletin,* 62, 5, 323-332.

*Kirby, F. D., & Shields, F. (1972). Modification of arithmetic response rate and attending behavior in a seventh-grade student. *Journal of Applied Behavior Analysis, 5,* 1, 79-84.

Kohn, A. (2001). Five reasons to stop saying "Good Job.". *Young Children, 56,* 5, 24-28.

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34,* 1, 26-38.

Maag, J. W. (2001). Rewarded by punishment: reflections on the disuse of positive reinforcement in education. *Exceptional Children, 67,* 2, 173-86.

*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,* 2, 139-150.

Maggin, D. M., O'Keeffe, B. V., & Johnson, A. H. (2011). A quantitative synthesis of methodology in the meta-analysis of single-subject research for students with disabilities: 1985-2009. *Exceptionality, 19,* 2, 109-135.

*McNamara, E., Harrop, A., & Owen, F. (1987). The effect of group orientated classroom management procedures on individual pupils. *Educational Psychology, 7,* 3, 157-168.

Merrett, F., & Houghton, S. (1989). Does it work with the older ones? A review of behavioural studies carried out in British secondary schools since 1981. *Educational Psychology, 9,* 4, 287-310.

Morgan, M. (1984). Reward-induced decrements and increments in intrinsic motivation. *Review of Educational Research, 54,* 1, 5-30.

Musti-Rao, S., & Haydon, T. (2011). Strategies to increase behavior-specific teacher praise in an inclusive environment. *Intervention in School and Clinic, 47,* 2, 91-97.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

*O'Leary, K. D., Becker, W. C., Evans, M. B., & Saudargas, R. A. (1969). A token reinforcement program in a public school: a replication and systematic analysis. *Journal of Applied Behavior Analysis, 2,* 1, 3-13.

Parker, R. I., & Vannest, K. J. (2012). Bottom-up analysis of single-case research designs. *Journal of Behavioral Education, 21*, 3, 254-265.

Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children, 75*, 2, 135-150.

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*, 4, 303-22.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy, 42*, 2, 284-99.

Partin, T. C. M., Robertson, R. E., Maggin, D. M., Oliver, R. M., & Wehby, J. H. (2010). Using teacher praise and opportunities to respond to promote appropriate student behavior. Preventing school failure: *Alternative Education for Children and Youth, 54*, 3, 172-178.

*Pfiffner, L. J., Rosén, L. A., & O'Leary, S. G. (1985). The efficacy of an all-positive approach to classroom management. *Journal of Applied Behavior Analysis, 18*, 3, 257-261.

*Polvinale, R. A., & Lutzker, J. R. (1980). Elimination of assaultive and inappropriate sexual behavior by reinforcement and social-restitution. *Mental Retardation, 18,* 1, 27-30.

*Salzberg, R. S., & Salzberg, C. L. (1981). Praise and corrective feedback in the remediation of incorrect left-hand positions of elementary string players. *Journal of Research in Music Education, 29*, 2, 125-33.

*Sawyer, L. M. C., Luiselli, J. K., Ricciardi, J. N., & Gower, J. L. (2005). Teaching a child with autism to share among peers in an integrated preschool classroom: acquisition, maintenance, and social validation. *Education and Treatment of Children, 28,* 1, 1-10.

Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: issues and applications. *Behavior Modification*, *22*, 3, 221-42.

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: methodology and validation. *Remedial and Special Education, 8,* 2, 24-33.

*Serbin, L. A., Tonick, I. J., & Sternglanz, S. H. (1977). Shaping cooperative cross-sex play. *Child Development*, *48,* 3, 924-929.

Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-based Communication Assessment and Intervention, 2,* 3, 188-196.

Shores, R. E., Gunter, P. L., Jack, S. L. (1993). Classroom management strategies: Are they setting events for coercion? *Behavioral Disorders, 18,* 2, 92-102.

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: considerations for research to practice. *Education and Treatment of Children, 31,* 3, 351-380.

Skiba, R. J., Casey, A., & Center, B. A. (1985). Nonaversive procedures in the treatment of classroom behavior problems. *Journal of Special Education, 19*(4),

Stormont, M., & Reinke, W. (2009). The importance of precorrective statements and behavior-specific praise and strategies to increase their use. *Beyond Behavior*, *18*, 3, 26-32.

*Strain, P. S., & Timm, M. A. (1974). An experimental analysis of social interaction between a behaviorally disordered preschool child and her classroom peers. *Journal of Applied Behavior Analysis, 7*, 4, 583-90.

*Taylor, M. J., & Kratochwill, T. R. (1978). Modification of preschool children's bathroom behaviors by contingent teacher attention. *Journal of School Psychology, 16*, 1, 64-71.

Thomas, D. R., Becker, W. C., & Armstrong, M. (1968). Production and elimination of disruptive classroom behavior by systematically varying teacher's behavior. *Journal of Applied Behavior Analysis, 1*, 1, 35–45.

*Trollinder, D. M., Hee-sook, C., & Proctor, T. B. (2004). Use of delayed praise as a directive and its effectiveness on on-task behavior. *Journal of Applied School Psychology, 20*, 2, 61-83.

*Weaver, A. D., Watson, T. S. (2004). An ideographic investigation of the effects of ability- and effort-based praise on math performance and persistence. *The Behavior Analyst Today, 4*, 4, 381 – 389.

Wendt, O., & Miller, B. (2012). Quality appraisal of single-subject experimental designs: an overview and comparison of different appraisal tools. *Education and Treatment of Children, 35*, 2, 235-268.

What Works Clearinghouse (Ed.). (2014). *Procedures and standards handbook. Version 3.0*. What Works Clearinghouse. What Works Clearinghouse. Princeton, NJ.

*Workman, E. A., Kindall, L. M., & Williams, R. L. (1980). The consultative merits of praise-ignore versus praise- reprimand instruction. *Journal of School Psychology, 18,* 4, 373-380.

*Yawkey, T. D. (1971). Conditioning independent work behavior in reading with seven year old children in a regular early childhood classroom. *Child Study Journal*, *2,* 1, 23-34.

APPENDIX A

INCLUSION SCREENING CRITERIA

#	Definition	Examples	NonExamples	Rationale
1	peer-reviewed & published in English	English translation available		researcher only speaks/read English
2	IV – Praise	verbal teacher peer parent written positive peer reporting praise with informative language only consequence manipulation can separate effects of consequence manipulations praise is a consequence of the behavior in-situ teacher data collected for treatment integrity purposes	gestural praise is the DV, not the IV extensive informative feedback beyond the praise statement multiple consequence manipulations simultaneously (e.g., praise, , reprimand, token economy, self-monitoring, verbal/graphic feedback) praise is part of the training package and not offered when the behavior is exhibited in-situ CW-FIT The Good Behavior Game BEST in CLASS DBRC Check In Check Out teacher data collected as the DV	isolate the effects of praise limit confusion between tangible reinforcers and praise

#	Definition	Examples	NonExamples	Rationale
3	DV – Student Behavior	measure of observed behavior social behavior academic behavior task persistence	rating of intrinsic motivation classroom atmosphere animal behavior perception of ability perception of locus of control descriptive studies looking at naturally occurring praise, only preference of reinforcer	operant conditioning focuses on changing observable, measurable behaviors these measures may yield different outcomes that measures for intrinsic motivation, perception, etc.
4	Experimental – SCD	withdrawal multiple baseline multiple probe alternating treatment component analysis	RCT quasi-experimental pre-test/post-test descriptive longitudinal literature review meta-analysis opinion pieces editorials recommendation pieces	applied behavior analysis relies on single-case designs this literature base has not be reviewed previously
5	Setting – School-based	classroom common areas naturalistic activities	clinic home community contrived activity conducted in a school facility	evaluating using WWC standards which is focused on determining evidence-based practices for educators studies from other settings may not have external validity and generalize to school-based settings

#	Definition	Examples	NonExamples	Rationale
6	Include extractable data (Study 2 Only)	line graph raw data reported by unit of analysis	aggregated across unit of analysis means	extractable data is necessary in order to calculate effect sizes

APPENDIX B

DESCRIPTIVE CODING MANUAL

Feature	Code 1	2	3	4	5	6
			Study Characteristics			
Publication Year	1960 – 1969	1970 – 1979	1980 – 1989	1990 – 1999	2000 – 2009	2010 - 2015
Design	AB	ABAB	MB across participants	MB across settings	MB across behaviors	Alternating Treatment
Design Quality (WWC)	does not meet	meets with reservations	meets without reservations			
Evidence of Effects (WWC)	no evidence	moderate evidence	strong evidence			
			Participant & Setting Characteristics			
Participant Disability Status	high-incidence	low-incidence	typically developing	mixed		
	All individuals in the study are individual with disabilities categorized by IDEA as high incidence. Examples: SLI, LD, ED, ID (IQ 55-70)	All individuals in the study are individuals with disabilities categorized as low-incidence by IDEA. Examples: ID, HI, OI, VI, DB, Deafness, OHI, TBI, ASD, Multiple Disabilities	All of the individuals in in the study are typically developing. No individuals are identified as individuals with disabilities.	At least 1 individual in the study is identified as an individual with a disability AND at least 1 individual in the study is identified as typically developing.		
Participant Ethnicity	White	Black or African American	Hispanic	American Indian or Alaska Native	Asian	Native Hawaiian or Other Pacific

Feature	Code 1	2	3	4	5	6
						Islander
Participant SES	high	low	mixed			
	None of the participants in the study receive free or reduced price lunch.	All of the participants in the study receive free or reduced priced lunch or are identified by the researcher as low-income.	At least 1 participant receives free or reduced price lunch AND at least 1 participant does not receive free or reduced price lunch.			
Participant Gender	male	female	mixed			
Educational Setting	general education	special education	mixed			
	All study activities occur in general education classrooms.	All study activities occur in special education classrooms.	Study activities occur in both general education and special education classrooms.			
Grade Level	PK – 2	3 – 5	6 – 8	9 -12	mixed	
			Intervention Characteristics			
Function-Based Intervention	yes	no				
	An FBA	No FBA				

Feature	Code 1	2	3	4	5	6
	conducted prior to intervention. The results of the assessment indicate praise is a function-based intervention for the participant(s).	conducted. It is unknown if praise is a function-based intervention for the participant(s).				
Interventionist/Source	teacher The participant's classroom teacher delivered praise to the participant(s). Examples: teacher, co-teacher, teaching assistant, paraprofessional.	peer One or more of the participant's peers delivered praise to the participant(s). Examples: peer tutor, peer mentor, classmates	researcher The experimenter or a research assistant delivered praise to the participant(s).			
Interventionist Training	none The researcher provided no training for the interventionist. Examples: researcher	brief The researcher provided basic information to the interventionist. Examples:	in-depth The researcher provided training to the interventionist. Examples: researcher			

	Code					
Feature	**1**	**2**	**3**	**4**	**5**	**6**
	conducted intervention; interventionist told to provide praise without receiving any additional information or participating in any practice.	researcher explained the timing and/or type of praise to given; researcher explained the rationale for using praise as an intervention	provided opportunities for practice; researcher required interventionist to meet implementation criterion prior to beginning intervention			
Target	universal Interventionist directs praise at any participant without targeting specific students.	targeted Interventionist directs praise at a 2 or more targeted participants.	individual Interventionist directs praise at an individual participant.	vicarious Interventionist directs praise at an individual participant with the intent of changing another participant's behavior.		
Target Training	none Participants receive no information about the intervention	brief Participants are told that the intervention will be implemented.	in-depth Participants role play or practice the intervention conditions.			

Feature	Code 1	2	3	4	5	6
	prior to or during the implementation.					
Schedule of Reinforcement	continuous ratio The interventionist praises every occurrence of the target behavior.	fixed ratio The interventionist praises every *n*th occurrence of the target behavior.	variable ratio The interventionist praises occurrences of the target behavior unpredictably.	fixed interval The interventionist praises the first occurrence of the target behavior after time specified amount of time has elapsed.	variable interval The interventionist praises the first occurrence of the target behavior after varying amounts of time have elapsed.	
Antecedent Manipulations	review rules The interventionist verbally states the rules or expectations prior to the beginning of data recording.	prompt The interventionist increases the likelihood that the target will engage in the target behavior. Examples: physical prompt, verbal cue,				

Feature	Code 1	2	3	4	5	6
		visual cue, gestural prompt, modeling				
Additional Consequent Interventions	ignore					
Contingency	The interventionist ignores inappropriate behaviors. Examples: prior to intervention the interventionist relied on reprimands and during the intervention the interventionist is instructed to ignore inappropriate behaviors instead of delivering a reprimand noncontingent	engagement	completion	performance		

Feature	Code 1	2	3	4	5	6
		contingent The provision of praise is contingent on beginning, but not necessarily completing, the target behavior. Example: praise is provided for starting an assignment rather than waiting until the assignment is complete	contingent The provision of praise is contingent on completing the target behavior. Example: praise is withheld until all toys are put away in a center	contingent The provision of praise is contingent on matching or exceeding a certain performance criterion. Examples: score above 90%; task completion in less than 10 seconds		
Specificity	behavior specific The praise statements include an operational description of the target behavior. Examples: I see John placing the blocks gently into the tub.	general The praise statements do not include an operational description of the target behavior. Examples: Great job!	mixed The praise statements include both behavior specific praise and general praise.			

	Code					
Feature	**1**	**2**	**3**	**4**	**5**	**6**
Control	controlling The praise statement contains language conveying the expectation that the target behavior will continue or that the interventionist needs the participant to do well. Examples: That's great. You should keep coloring in the lines. If you keep up the good work, I can display your paper on the bulletin board for parents' night.	informational The praise statement is free of language conveying ongoing expectations or interventionist's needs. Examples: That's great. You are coloring in the lines.				
Attribution	ability	effort				

	Code					
Feature	**1**	**2**	**3**	**4**	**5**	**6**
	The praise statements contain language attributing the accomplishment of the target behavior to the participant's skill level. Examples: You are smart. You're really good at math.	The praise statement contains language attributing the accomplishment of the target behavior to the participant's effort. Examples: I can tell you tried really hard on this assignment. You worked hard in the science center				
Expected	yes	no				
	The participant is explicitly made aware of the contingency for receiving praise.	The participant is unaware of the contingency for receiving praise.				
Interest in Activity	high	low				
	The participant indicates that the target behavior	The participant indicates that the target behavior				

Feature	Code 1	2	3	4	5	6
	is appealing.	is not appealing				
Treatment Fidelity	yes Treatment fidelity is measured and inter-rater reliability is reported.	no Treatment fidelity is not measured.				
Social Validity	yes The interventionist(s) and/or participants find the intervention socially valid.	no The interventionist(s) and/or participants do not find the intervention socially valid.				
			Outcome Characteristics			
DV Type	appropriate social behavior The intent is to increase the target behavior. The behavior is not academic. Examples: on-task behavior, appropriate language, asking	inappropriate social behavior The intent is to decrease the target behavior. The behavior is not academic. Examples: off-task, out of seat, disruptive	academic behavior The target behavior demonstrates mastery of an academic skill. Examples: percent of question correct			

Feature	Code 1	2	3	4	5	6
	questions; percent of assignments completed					
Dependent Variable Direction	increase behavior	decrease behavior				
DV Measure	frequency The data recorded represent the number of time the target behavior occurred.	interval The data recorded represent the number or percentage of intervals in which the target behavior occurred or did not occur.	duration The data recorded represent the amount of time the participant engage in the target behavior.	latency The data recorded indicates the amount of time elapsed between the request for performance of the target behavior and the actual performance of the target behavior.		
Recording Procedure	pencil/paper	electronic				
			Effect Characteristics			
Unit of Analysis	whole class The outcome measure reflects the behavior of the entire class.	small group The outcome measure reflects the behavior of a portion of the	individual The outcome measure reflects the behavior of a	mixed Multiple outcome measures are collected.		

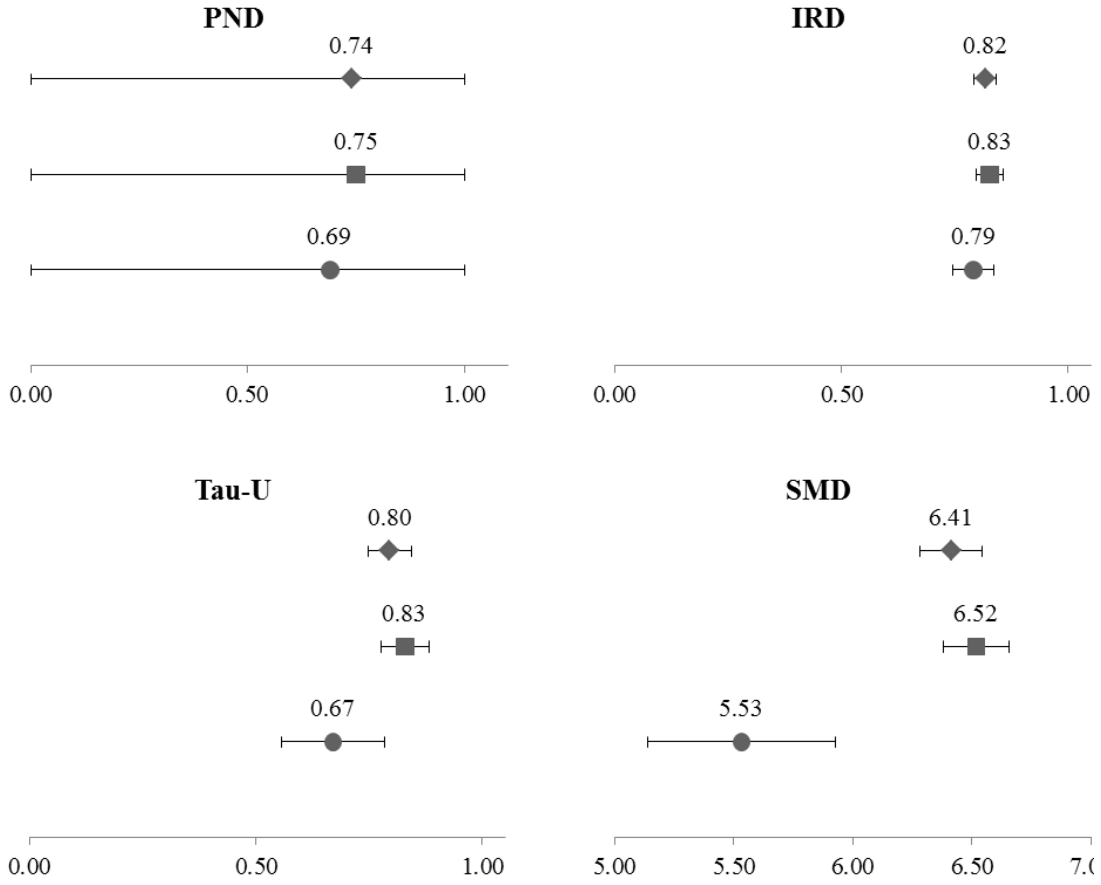
Feature	Code 1	2	3	4	5	6
	Examples: multiple baseline design across classrooms	class. Examples: intervention implemented with a reading group	single student. Examples: implementing praise as a behavior intervention for a single student			
Reported Effects Evaluation	effective alone Praise is the only consequence-based intervention employed in the study.	ineffective addition One, or more, consequent-based interventions are employed by the interventionist. Examples: alternating treatment comparing the effectiveness of praise and verbal rewards; multi-component intervention package	mixed			
Type of Contrast	baseline to intervention	baseline to nonadjacent intervention	intervention to intervention	baseline to maintenance	intervention to maintenance	intervention to nonadjacent

	Code					
Feature	**1**	**2**	**3**	**4**	**5**	**6**
						maintenance

APPENDIX C

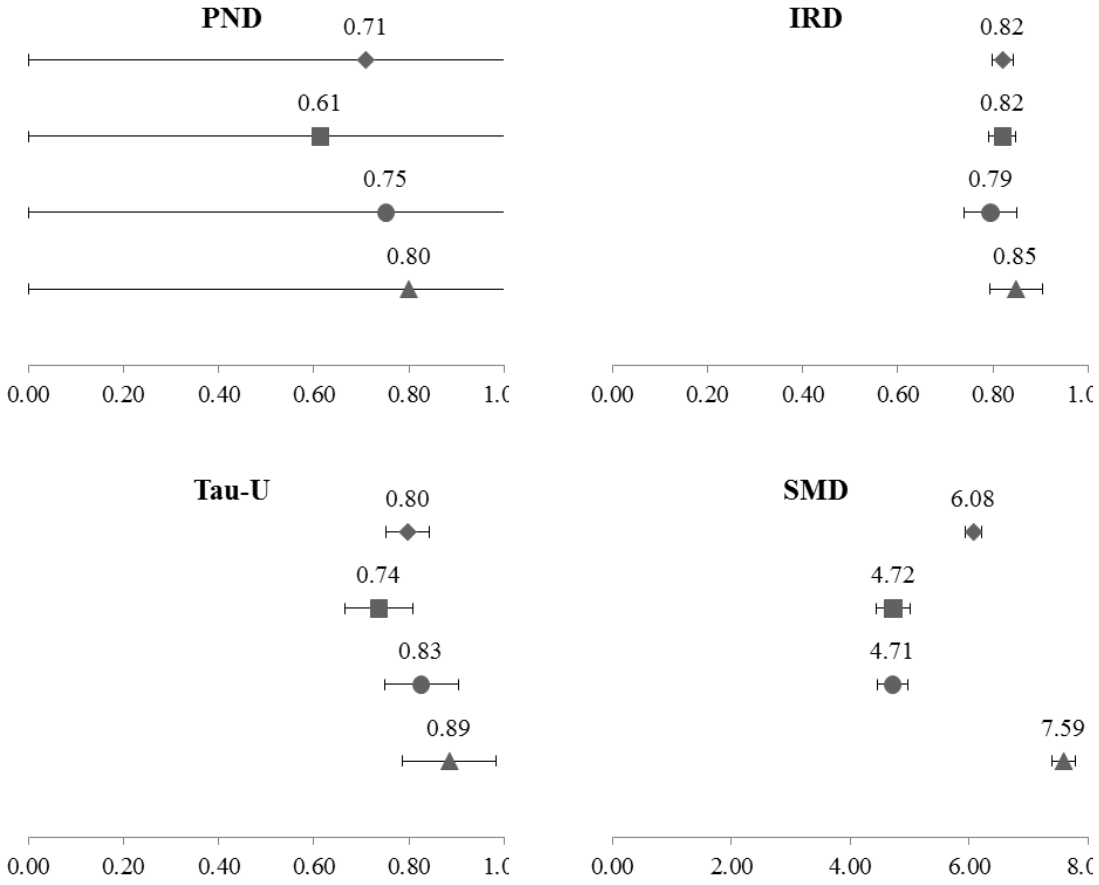
FOREST PLOTS FOR MODERATOR ANALYSES

Figure 7. Moderator analysis by disability status.



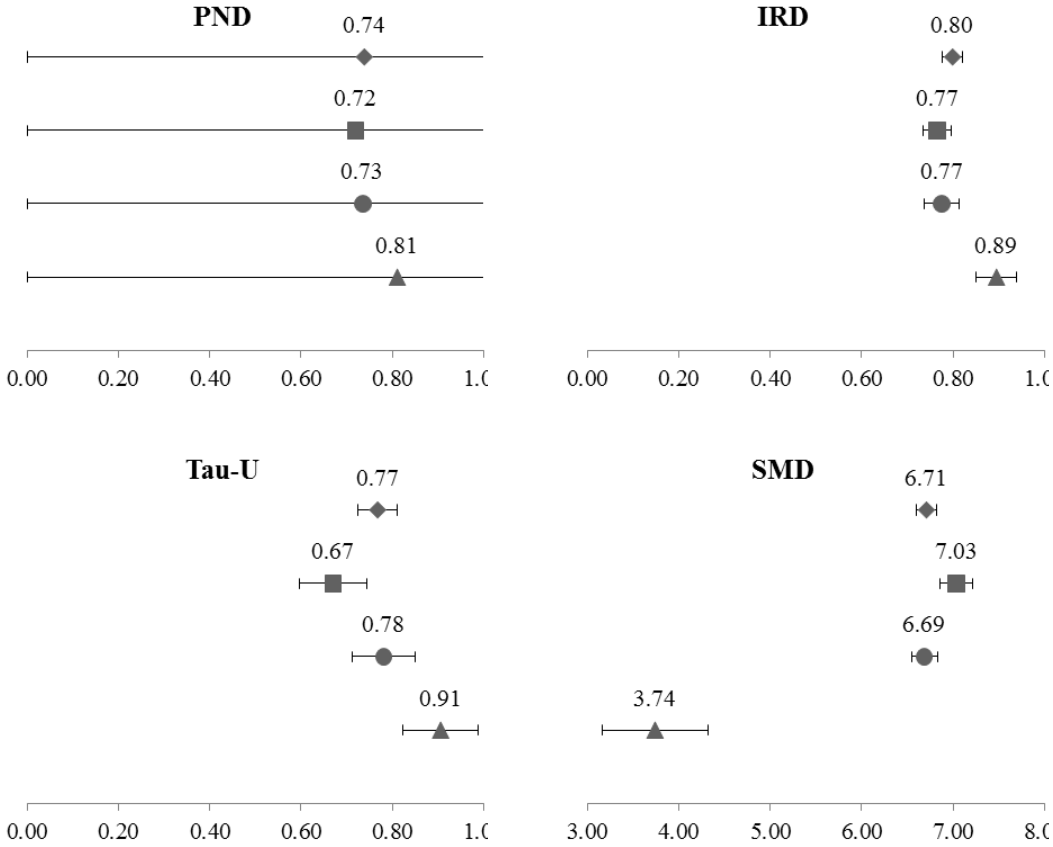
Note: ♦ = omnibus effect size estimate aggregated from 97 phase contrasts. ■ = effect size estimate for students with high incidence disabilities aggregated from 80 phase contrasts. ● = effect size estimate for typically developing students aggregated from17 phase contrasts. PND error bars show range. IRD, Tau-U, and SMD error bars show 83.4% confidence intervals. Nonoverlapping confidence intervals indicate a statistically significant difference between the effect size estimates at p = 0.05.

Figure 8. Moderator analysis by gender.



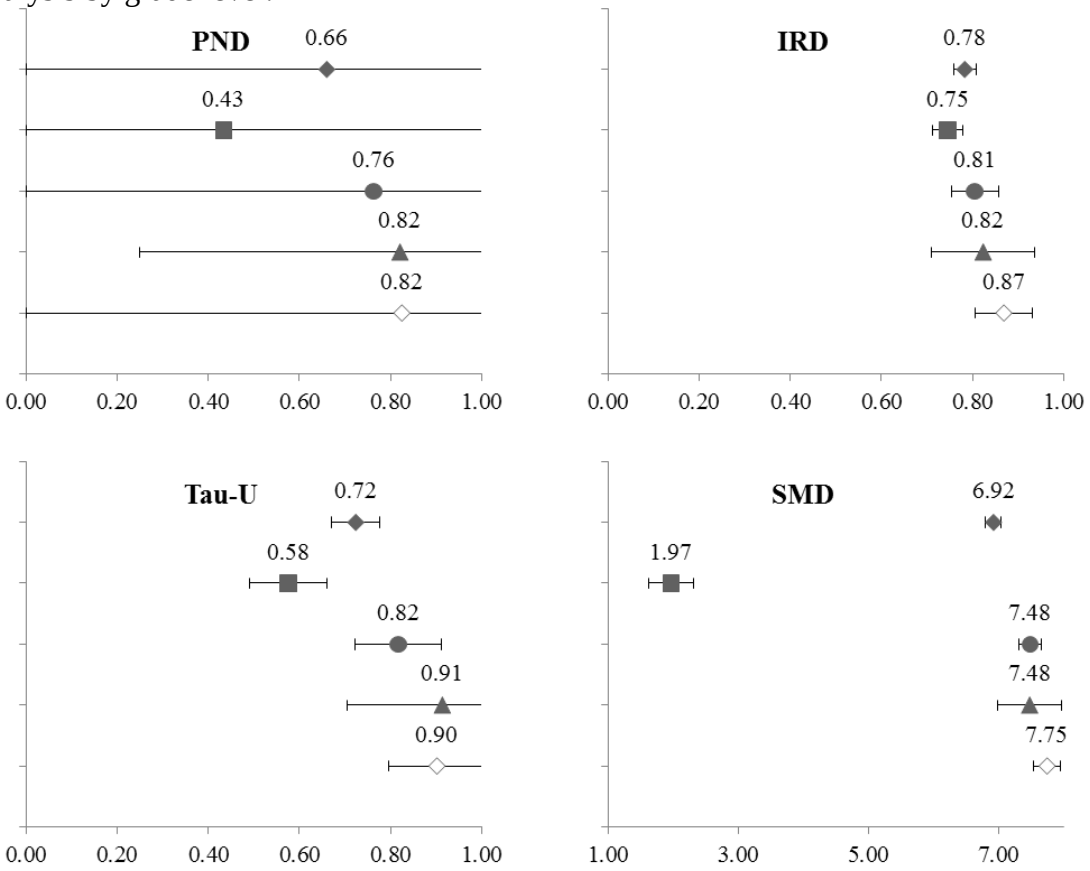
Note: ♦ = omnibus effect size estimate aggregated from 99 phase contrasts. ■ = effect size estimate for male participants aggregated from 40 phase contrasts. ● = effect size estimate for female participants aggregated from 32 phase contrasts. ▲ = effect size estimate for male and female students grouped together calculated from 27 phase contrasts. PND error bars show range. IRD, Tau-U, and SMD error bars show 83.4% confidence intervals. Nonoverlapping confidence intervals indicate a statistically significant difference between the effect size estimates at p = 0.05.

Figure 9. Moderator analysis by educational setting.



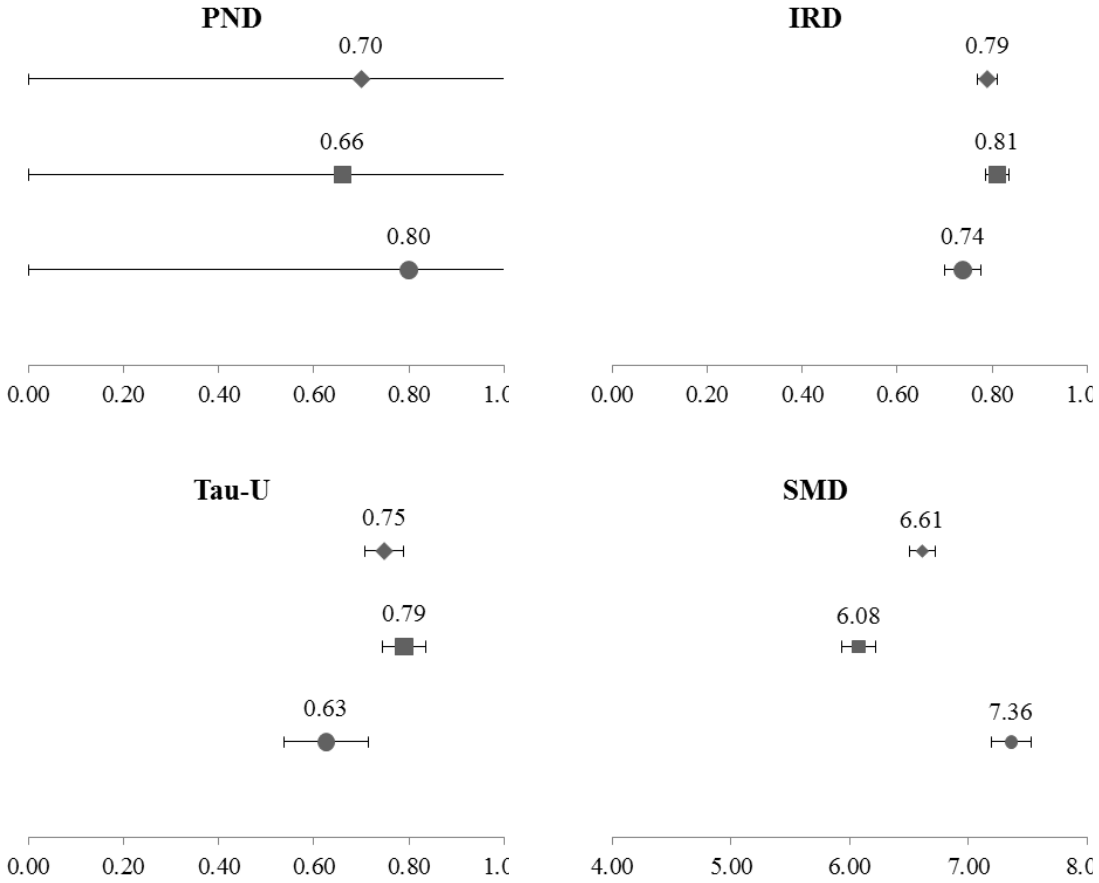
Note: ♦ = omnibus effect size estimate aggregated from 129 phase contrasts. ■ = effect size estimate for general education settings aggregated from 49 phase contrasts. ● = effect size estimate for special education settings aggregated from 64 phase contrasts. ▲ = effect size estimate for a combination of general education and special education settings calculated from 16 phase contrasts. PND error bars show range. IRD, Tau-U, and SMD error bars show 83.4% confidence intervals. Nonoverlapping confidence intervals indicate a statistically significant difference between the effect size estimates at p = 0.05.

Figure 10. Moderator analysis by grade level.



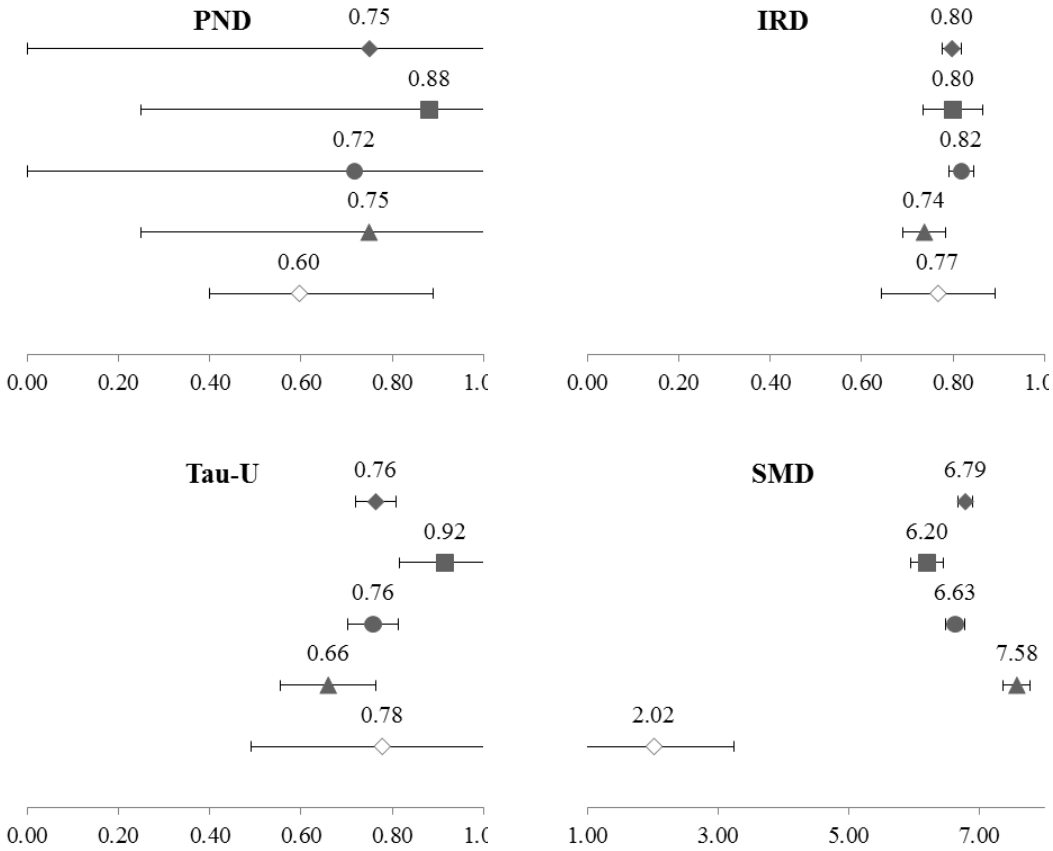
Note: ♦ = omnibus effect size estimate aggregated from 102 phase contrasts. ■ = effect size estimate for early elementary (PK -2) aggregated from 37 phase contrasts. ● = effect size estimate for upper elementary (3-5) aggregated from 35 phase contrasts. ▲ = effect size estimate for middle school (6-8) calculated from 6 phase contrasts. ◊ = effect soze estimates for high school aggregated from 24 contrasts. PND error bars show range. IRD, Tau-U, and SMD error bars show 83.4% confidence intervals. Nonoverlapping confidence intervals indicate a statistically significant difference between the effect size estimates at p = 0.05.

Figure 11. Moderator analysis by intervention package.



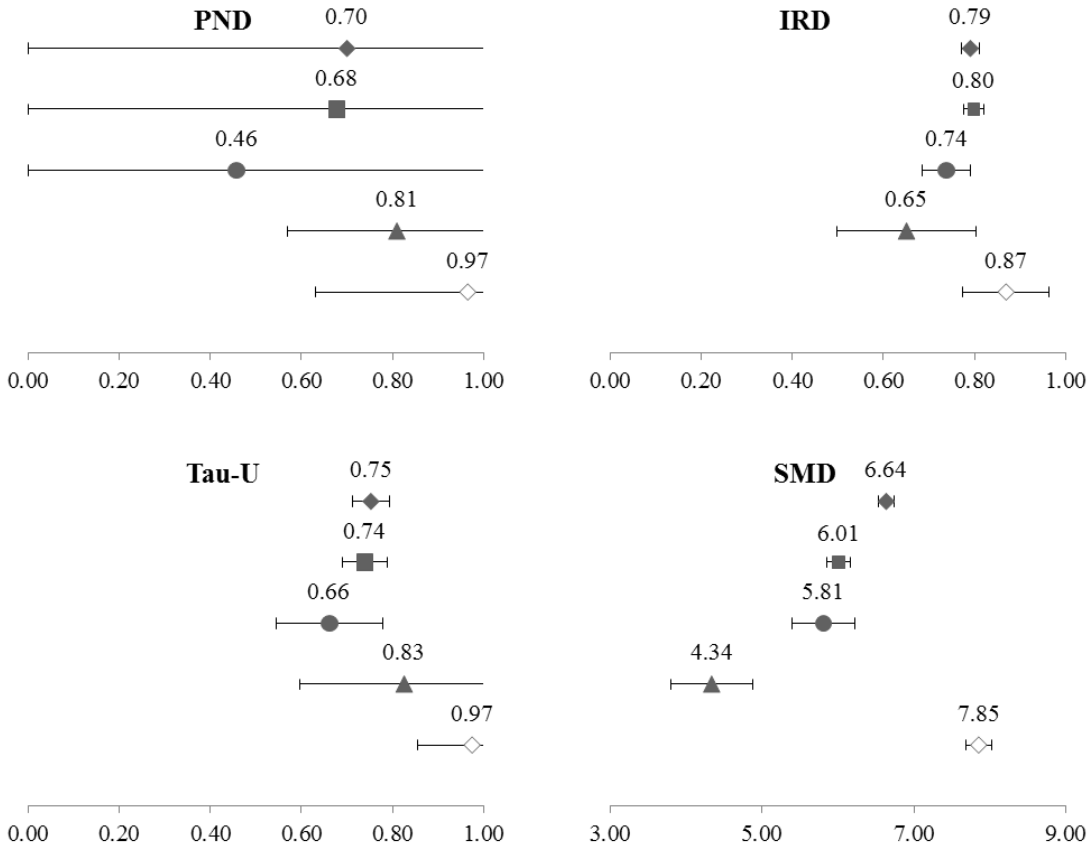
Note: ♦ = omnibus effect size estimate aggregated from 141 phase contrasts. ■ = effect size estimate for praise alone aggregated from 101 phase contrasts. ● = effect size estimate for praise in combination with antecedent manipulations or ignoring 40 phase contrasts. PND error bars show range. IRD, Tau-U, and SMD error bars show 83.4% confidence intervals. Nonoverlapping confidence intervals indicate a statistically significant difference between the effect size estimates at p = 0.05.

Figure 12. Moderator analysis by schedule of reinforcement.



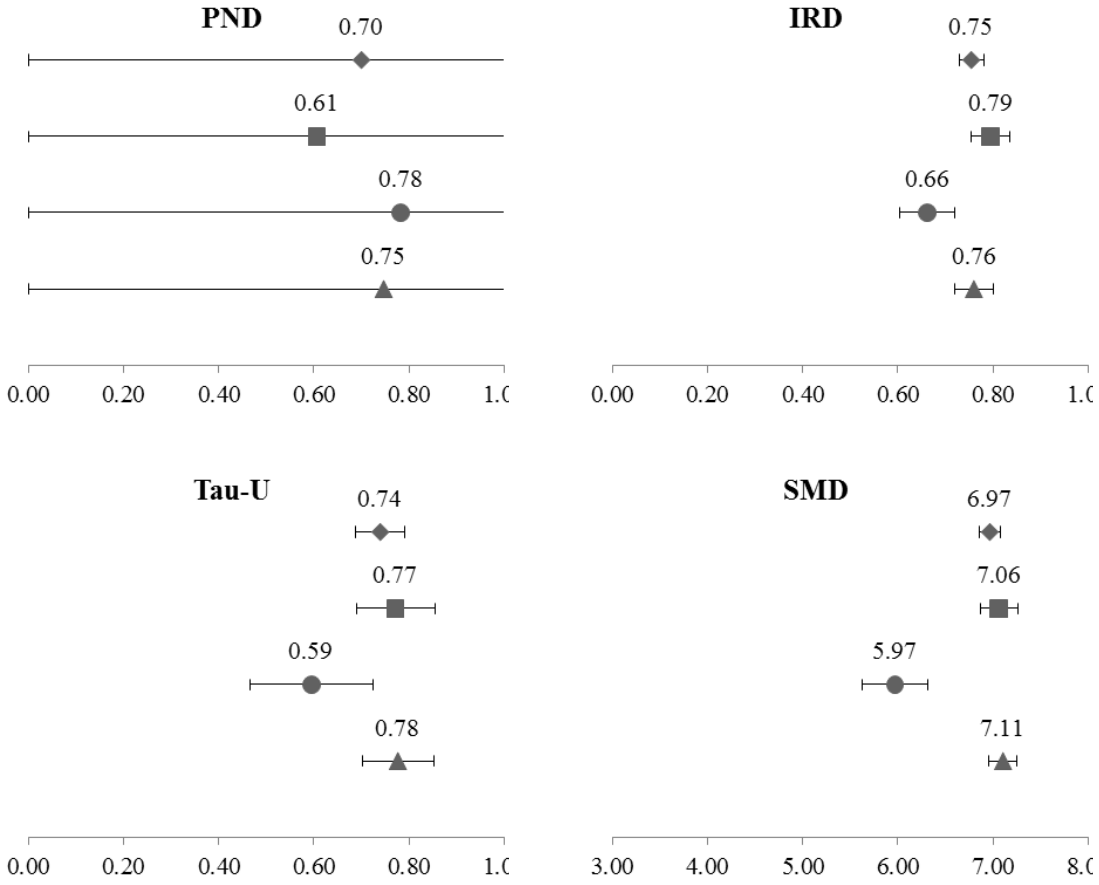
Note: ♦ = omnibus effect size estimate aggregated from 125 phase contrasts. ■ = effect size estimate for fixed ratio aggregated from 24 phase contrasts. ● = effect size estimate for variable ratio aggregated from 69 phase contrasts. ▲ = effect size estimate for fixed interval aggregated from 29 phase contrasts. ◊ = effect size estimates for variable interval aggregated from 3 contrasts. PND error bars show range. IRD, Tau-U, and SMD error bars show 83.4% confidence intervals. Nonoverlapping confidence intervals indicate a statistically significant difference between the effect size estimates at p = 0.05.

Figure 13. Moderator analysis by type of contingency.



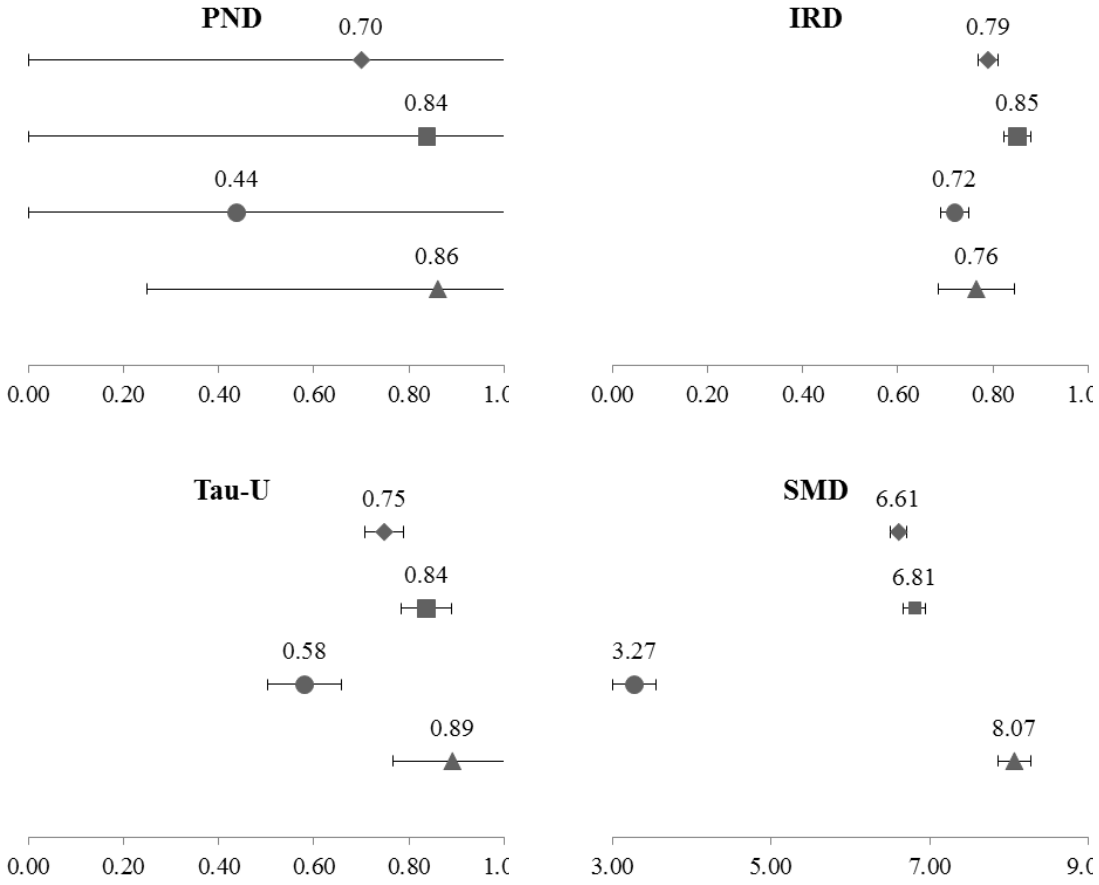
Note: ♦ = omnibus effect size estimate aggregated from 140 phase contrasts. ■ = effect size estimate for engagement contingent aggregated from 92 phase contrasts. ● = effect size estimate for completion contingent aggregated from 18 phase contrasts. ▲ = effect size estimate for performance contingent aggregated from 6 phase contrasts. ◊ = effect size estimates for multiple contingencies aggregated from 24 contrasts. PND error bars show range. IRD, Tau-U, and SMD error bars show 83.4% confidence intervals. Nonoverlapping confidence intervals indicate a statistically significant difference between the effect size estimates at p = 0.05.

Figure 14. Moderator analysis by specificity.



Note: ♦ = omnibus effect size estimate aggregated from 110 phase contrasts. ■ = effect size estimate for behavior specific praise aggregated from 40 phase contrasts. ● = effect size estimate for general praise aggregated from 17 phase contrasts. ▲ = effect size estimate for a combination of behavior specific and general praise calculated from 53 phase contrasts. PND error bars show range. IRD, Tau-U, and SMD error bars show 83.4% confidence intervals. Nonoverlapping confidence intervals indicate a statistically significant difference between the effect size estimates at p = 0.05.

Figure 15. Moderator analysis by type of outcome variable.



Note: ♦ = omnibus effect size estimate aggregated from 141 phase contrasts. ■ = effect size estimate for appropriate social behavior aggregated from 70 phase contrasts. ● = effect size estimate for inappropriate social behavior aggregated from 50 phase contrasts. ▲ = effect size estimate for academic behavior calculated from 21 phase contrasts. PND error bars show range. IRD, Tau-U, and SMD error bars show 83.4% confidence intervals. Nonoverlapping confidence intervals indicate a statistically significant difference between the effect size estimates at p = 0.05.