

**Using Classroom Observations to Describe and Model the Impact of Positive and Negative
Teaching Behaviors on Classroom Disruptive Behavior**

by

Justin N. Coy

Bachelor of Science, University of Pittsburgh, 2011

Master of Education, University of Pittsburgh, 2012

Submitted to the Graduate Faculty of the
School of Education in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

SCHOOL OF EDUCATION

This dissertation was presented

by

Justin N. Coy

It was defended on

October 30, 2020

and approved by

Douglas E. Kostewicz, PhD, Faculty, Department of Instruction and Learning

Jesse E. Dvorchak, PhD, Senior Associate, edCount, LLC

Steven R. Lyon, PhD, Emeritus Faculty, Department of Instruction and Learning

Rachel E. Robertson, PhD, Assistant Professor, Department of Instruction and Learning

Copyright © by Justin N. Coy

2020

Using Classroom Observations to Describe and Model the Impact of Positive and Negative Teaching Behaviors on Classroom Disruptive Behavior

Justin N. Coy, PhD

University of Pittsburgh, 2020

Positive, proactive classroom management strategies support academic and behavioral student outcomes while fostering positive teacher-student relationships. However, teachers often cite struggles with classroom management and challenging student behavior as key reasons they ultimately leave the field. Additionally, pre-service teachers often fail to receive substantive training in effective classroom management. Experiments within the present study sought to better understand the foundational role of teachers' positive and negative verbal interactions with students. Experiment 1 utilized descriptive and inferential statistics to better understand the current rate of teachers' positive and negative verbal interactions regarding student behavior, as well as the influence of specific teacher behaviors on classroom disruptive behavior. Experiment 2 evaluated the effectiveness of a low-intensity treatment package (training, performance feedback, and reflective goal-setting) to adjust teachers' verbal interactions with students. Results from Experiment 1 show teachers used nearly five times as many negative interactions as positive, with significant differences across teachers and specific behaviors. Teachers' negative statements were also two times longer than their positives, on average. Teachers appeared to rely on unique negative 'crutches' – individual collections of specific negative behaviors. Both criticisms and attention to junk statements significantly influenced the rate of classroom disruptive behaviors. Experiment 2 findings indicate the treatment package helped one participant make significant

changes over baseline (increased positive interactions and reduced negative interactions). Results from this study support the need for additional large-scale descriptive studies of teacher interactions and coercives, as well as an exploration of the wide variability of teachers' positive and negative interaction rates across available research.

Table of Contents

Preface.....	xiv
1.0 Introduction.....	1
2.0 Literature Review	7
2.1 Purpose of Current Review	7
2.2 Method.....	8
2.2.1 Article Analysis	9
2.2.2 Article Quality Assessment and Data Extraction.....	10
2.3 Results.....	11
2.3.1 Research Question One – Who and Where	11
2.3.1.1 Teachers.....	11
2.3.1.2 Unit of Analysis	15
2.3.1.3 Settings.....	16
2.3.2 Research Question Two – Target Teacher and Student Behaviors.....	16
2.3.2.1 Teacher Behaviors	17
2.3.2.2 Student Behaviors	17
2.3.3 Research Question Three – Intervention Components.....	18
2.3.3.1 Research Design	21
2.3.4 Research Question Four – Empirical and Social Validity Results	21
2.3.4.1 Baseline	22
2.3.4.2 Teacher Trainings.....	22
2.3.4.3 Performance Feedback.....	22

2.3.4.4 Self-Monitoring	23
2.3.4.5 Treatment Packages	23
2.3.4.6 Negative Teacher Behaviors	24
2.3.4.7 Follow-up and Maintenance.	24
2.3.4.8 Student Behaviors	24
2.3.4.9 Social Validity.	25
2.4 Discussion	26
3.0 Method	30
3.1 Setting	30
3.2 Participants	30
3.3 Materials.....	32
3.4 Dependent Variables	32
3.4.1 Teacher Behavior	33
3.4.2 Classroom Disruptive Behavior	33
3.5 Experimental Design	35
3.6 Data Collection and Analysis.....	36
3.6.1 Descriptive Statistics	36
3.6.2 Inferential Statistics	37
3.6.2.1 Analysis of Variance.....	37
3.6.2.2 Multiple Linear Regressions	39
3.6.3 Qualitative Analysis of Teachers' Vocal Statements	40
3.6.4 Assessing the Effectiveness of Intervention Package via Brief Case-Study..	40
3.6.4.1 SCC Metrics	40

3.6.4.2 t-tests	42
3.7 Independent Variable.....	43
3.7.1 Training.....	43
3.7.2 Performance Feedback Emails	44
3.7.3 Reflective Goal-Setting	45
3.8 Procedures.....	45
3.8.1 Pre-Study Observation	45
3.8.2 Observational Data/Baseline	46
3.8.3 Training.....	46
3.8.4 Performance Feedback Emails	47
3.9 Reliability, Believability, and Procedural Integrity	47
3.10 Social Validity	48
4.0 Results	49
4.1 Rates and Characteristics of Teachers' Positive and Negative Vocal Interactions	49
4.1.1 Positive Interactions.....	50
4.1.2 Negative Interactions	52
4.2 Classroom Disruptive Behavior	54
4.2.1 Multiple Linear Regression.....	55
4.3 Qualitative Analysis of Teachers' Vocal Statements.....	61
4.3.1 Criticisms	61
4.3.2 Arguing	67
4.3.3 Logic/Lecture.....	67
4.3.4 Questioning.....	68

4.3.5 Sarcasm	68
4.3.6 Pleading/Despair	68
4.3.7 Threats	68
4.3.8 Force	69
4.3.9 Attention to Junk Behavior	69
4.3.10 Other Negatives	69
4.4 Effectiveness of Intervention Package via Brief Case-Study	69
4.4.1 Baseline	72
4.4.2 Intervention	72
4.4.3 Social Validity	73
5.0 Discussion	76
5.1 Coercion in the Classroom	79
5.2 Intervention Effectiveness	81
5.3 Limitations	82
5.4 Directions for Future Research	83
5.5 Implications for Practice	84
Appendix A Literature Review Descriptive Study Characteristics	85
Appendix B Study Materials	86
Bibliography	99

List of Tables

Table 1. Teacher Demographics Across Reviewed Studies.....	12
Table 2. Student/Classroom Demographics Across Reviewed Studies	13
Table 3. School Demographics Across Reviewed Studies	14
Table 4. Experimental Methodologies Across Reviewed Studies	19
Table 5. Participating Teacher Demographics.....	31
Table 6. Participating Classroom Characteristics	32
Table 7. Teacher and Student Dependent Variable Definitions.....	34
Table 8. Frequency of Teachers’ Positive Vocal Interaction Behaviors per 16-min. Observation	51
Table 9. Number of Words Associated with Teachers’ Positive Vocal Interaction Behaviors	51
Table 10. Frequency of Teachers’ Negative Vocal Interaction Behaviors per 16-min. Observation	52
Table 11. Number of Words Associated with Teachers’ Negative Vocal Interaction Behaviors	53
Table 12. Frequency of Classroom Disruptive Behaviors per 16-min. Observation	55
Table 13. Pairwise Correlations Between Classroom Disruptive Behavior and Teacher Behaviors	57
Table 14. Frequency Regression Model Results.....	60
Table 15. Word Regression Model Results.....	60

Table 16. Themes and Illustrative Quotes from Qualitative Analysis of Teacher Behaviors
..... 62

Table 17. Within-Condition and Between-Condition Measures of Adam’s Positive and Negative Vocal Interaction Frequencies 71

Table 18. Within-Condition and Between-Condition Measures of Words Associated with Adam’s Positive and Negative Vocal Interactions 71

Table 19. Within-Condition and Between-Condition Measures of Classroom Disruptive Behavior 71

List of Figures

Figure 1. Kernel Density Estimations of Classroom Disruptive Behaviors.....	56
Figure 2. Scatterplots of Classroom Disruptive Behavior and Correlated Teacher Behaviors	58
Figure 3. SCCs for Adam’s Positive and Negative Verbal Interactions (Frequency and Words) and Classroom Disruptive Behavior	70
Figure 4. Adam’s Completed Social Validity Measure	74
Appendix Figure A1. Literature Review Descriptive Study Characteristics	85
Appendix Figure B1. University of Pittsburgh Institutional Review Board Study Approval	86
Appendix Figure B2. Teacher Recruitment Letter.....	87
Appendix Figure B3. Full Teacher Consent (page 1)	88
Appendix Figure B4. Full Teacher Consent (page 2)	89
Appendix Figure B5. Full Teacher Consent (page 3)	90
Appendix Figure B6. Parent Notification Letter	91
Appendix Figure B7. Data Collection Form.....	92
Appendix Figure B8. Procedural Fidelity Checklist – Trainings.....	93
Appendix Figure B9. Procedural Fidelity Checklist – Emails.....	93
Appendix Figure B10. PowerPoint Training.....	94
Appendix Figure B11. PowerPoint Training, con’t	95
Appendix Figure B12. Performance Feedback Email Template.....	96
Appendix Figure B13. Social Validity Measure (close-ended questions).....	97

Appendix Figure B14. Social Validity Measure (open-ended questions)..... 98

Preface

I'd like to first thank my advisor, Dr. Douglas Kostewicz for his continuous support throughout my long educational journey. Thank you for pushing me to always think analytically and to consider problems through my behavior analytic lens.

Thank you to my committee members: Dr. Jesse Dvorchak, Dr. Steve Lyons, and Dr. Rachel Robertson. Your feedback and support helped me sculpt an amazing dissertation, even after everything flipped on its head.

A huge thank you to my amazing cohort mates: Dr. Rachel Schwartz, Dr. Rachel Schles, Dr. Sarah Westerfield, and Dr. Miguel Ampuero. Your friendship and support were instrumental in my success. I'm excited for our future collaborative work! Thank you to Olivia Enders for being my secondary coder, research buddy, and sounding board.

Finally, thank you to Jeremy DeLuca, my partner, best friend, citation checker, and biggest cheerleader. Words cannot express my gratitude for your patience, love, and support through my studies. I can't wait to see what the future holds for us!

1.0 Introduction

Decades of research establishes the evidence base for a variety of positive teaching behaviors, including increasing classroom structure and teaching expectations (Simonsen et al., 2008). Positive, proactive classroom management strategies promote student academic achievement, increase appropriate student behavior, and help build positive relationships between teachers and students (MacSuga-Gage et al., 2012). Struggles with classroom management and challenging student behavior are commonly cited as key stressors of teacher and main reason general education and special education teachers leave the field (Allday et al., 2012; Simonsen et al., 2017; Walter et al., 2006). However, many pre-service training programs fail to provide substantive training around classroom management or addressing student behaviors (Reinke et al., 2011; State et al., 2011). Additionally, in-service teachers at every grade level request additional training and support in these arenas (Reinke et al., 2008). Helping teachers build or change effective instructional teaching behaviors, however, often requires intensive researcher support. Researchers should strive to produce meaningful behavioral change as efficiently as possible. Altering the minute-by-minute interactions between teachers and students provides a potentially efficacious avenue for researchers to meaningfully change classroom instruction and environment.

Teachers' interactions with students can be generally classified as *positive* (e.g., praise or approval), *negative* (e.g., reprimand or coercion), or *neutral* (e.g., instruction, academic or routine directions, ineffectual statements). The positive vocal and/or non-vocal teacher behaviors intended to increase specific student behaviors can be described as positive teacher interactions (PTIs). Research to date has most commonly focused on praise when describing classroom PTIs. Praise is

an easy, effective, and valuable skill for teachers. The term praise describes “any verbal statement or gesture indicating teacher approval of a desired student behavior” (Reinke et al., 2008, p. 319). Researchers distinguish general praise from behavior specific praise. General or generic praise (GP) identifies statements conveying approval, but lacking specificity (e.g., “*Good job, Jacob*”). Behavior-specific praise (BSP) includes an acknowledgment of the desired student behavior (e.g., “*Jacob, thank you for sitting in your seat.*”; Kalis et al., 2007). A broader conceptual view of PTIs includes affirmations of correct responses, expressions of appreciation, and physical touch or gestures (Sabey et al., 2019). Vocal and non-vocal PTIs provide appropriate behavioral models for students and help teachers positively reinforce appropriate behaviors, thus increasing their future likelihood (Cooper et al., 2020).

Negative teacher interactions (NTIs) describe vocal and/or non-vocal teacher behaviors intended to decrease specific challenging behaviors. NTI behaviors vary across studies, but typically serve to criticize, reprimand, and/or disapprove of student behavior(s). Examples of NTIs include: (a) disapproving/corrective statements (e.g., “no,” “Don’t do that,” etc.); (b) verbal reprimands; (c) rude or coercive statements, such as arguing with student(s), using sarcasm, or lecturing student(s) about behavior; (d) error correction; (e) providing attention to inappropriate behavior; and (f) physical restraint or punitive gesture (Cook et al., 2017; Mrachko et al., 2017; Sabey et al., 2019)

The minute-by-minute positive interactions between teachers and students play an important role in students’ classroom experience (Hughes, 2011; McCormic et al., 2013). Teachers using more positive interactions establish supportive classroom environments which promote student academic and behavioral improvements (Jenkins et al., 2015; Markelz & Taylor, 2016; Royer et al., 2018). While it is likely that this ratio of positive to negative interactions influences

classroom environment and relationships, simply telling teachers to change their ratio does not provide enough support to enact and maintain these changes. Identifying effective, research-based behavioral guidelines or ratios might provide teachers a goal to work towards. Commonly cited metrics for rates of positive teaching behaviors include: (a) four positives for every negative interaction, (b) five positives to one negative, (c) six praises per 15-minutes, and (d) one BSP statement every two minutes (Cook et al., 2017; Myers et al., 2011; Sutherland et al., 2000). While these provide a goal for teachers, some important limitations exist. First, the origins of these recommendations are “nebulous and difficult to track” (Sabey et al., 2019, p 155). While the earliest mentions of *teacher approval and disapproval* date back to White (1975), ‘ideal’ teacher positive-to-negative ratios (PNRs) did not exist until the 1990s. Second, while some empirical support exists for the recommendations generally, none has compared the effects of one PNR to another. Empirical studies do support increasing teachers’ ratio of positive to negative interactions. Improving teachers’ ratio decreases student disruptive behavior and increases academic engagement when analyzed through descriptive statistics and visual analysis (Caldarella et al., 2020; Cook et al., 2017). Third, teachers ‘natural’ PNRs and PTI behaviors fall well below these recommendations as seen when reviewing existing literature. Even with intensive support, teachers are unable to meet these recommendations during interventions (Cook et al., 2017). Fourth, the recommendations do not specify important contextual attributes, including timeframe and quality. For example, is it better to achieve a 4:1 ratio every ten minutes, or for the entire class period? Does it matter if the PTIs are directed to the whole class, while every NTI is directed to the same group of students?

Researchers within the reviewed studies utilized a variety of intervention components to increase teachers use of PTI behaviors. In general, a positive relationship appears between

increased researcher involvement in the teachers' behavior change and the effectiveness of the intervention. Trainings alone were ineffective at promoting meaningful, sustained improvements in PTI behavior. This supports previous research showing the ineffectiveness of short and/or one-time trainings (Yoon et al., 2007). It appears that experiments with increased researcher interaction with participants (e.g., not relying on participant self-monitoring alone or within treatment packages) promoted more immediate and sustained improvements in PTI behaviors.

Reviewed studies including performance feedback (PF) focused treatment packages helped most teachers demonstrate immediate and sustained improvements in their PTI behaviors (e.g., Horton, 1975; Mrachko et al., 2017; O'Handley et al., 2018; Simonsen et al., 2017). Substantial research has identified PF as an effective, evidence-based way to improve teacher behavior (Cornelius & Nagro, 2014; Fallon et al., 2015). PF typically involves brief in-person or remote interactions between an intervention agent (researcher, colleague, supervisor, etc.) and a participant (Schles & Robertson, 2019). The information shared encompasses aspects of the participants' performance (Hattie & Timperley, 2007). Meta-analyses establish PF as an evidence-based practice for increasing participants' intervention fidelity and promoting behavior change with various education professionals (teachers, paraeducators, grade-level teams, etc.; Fallon et al., 2015). However, the effectiveness of interventions including performance feedback differed across experiments. Additionally, no clear link existed between frequency (e.g., daily, every few days, or weekly) or mode (e.g., in-person or email) of PF and its effect on targeted PTI behavior(s).

PF research often include a variety of additional intervention components (Fallon et al., 2015). For example, intervention agents can provide teachers with graphed performance data and emailed PF or provide verbal PF after teaching participants to self-monitor their own behavior change. Allowing teachers to self-reflect on researcher-provided PF, however, is a potentially

effective strategy rarely included in PF research (Barton et al., 2011). Self-reflection describes “the ability of teachers to reflect upon practice in a critical way so as to *actively improve current practice*” (Watts & Lawson, 2009, p. 610, emphasis in original). Behaviorally, self-reflection equates to self-management, promoting teachers’ evaluation and modification of their own behavior (Simonsen et al., 2010).

Self-reflection research typically evaluates permanent products left behind after reflection has occurred, such as journal entries or transcripts of conversations (Gelfuso & Dennis, 2014). A potentially effective way to promote self-reflection following PF would be to have the teachers self-identify their own instructional strengths and needs. Such ‘reflective goal-setting’ allows the teacher to review performance feedback information (behavioral data, researcher comments, etc.), reflect on their instruction, and self-identify their instructional successes and opportunities for improvement. Goal-setting is an effective addition to PF (Cavanaugh, 2013), however, the goals are usually established a priori or by researchers. As a novel addition to education literature, having teachers identify their own strengths/needs may increase their commitment and actualized behavior change.

This study posits a first step at exploring the quality of student-teacher interactions through quantitative and qualitative analyses of teachers’ positive and negative vocal interactions related to student behavior. The following research questions guided this study: (1) What are the natural rates and characteristics of teachers’ positive and negative vocal interactions with middle school students in a private school? (1a) What are the common characteristics or features of teachers’ vocal statements? (1b) How does teachers’ use of specific positive and negative behaviors relate to classroom disruptive behavior(s)? (2) What is the preliminary effectiveness of a low-intensity intervention package including training, performance feedback emails, and reflective goal-setting

on a teachers' positive and negative vocal interactions with students? (2a) How socially valid is the intervention in altering teachers' positive and negative interactions with students?

2.0 Literature Review

The results from a series of reviews support the effectiveness of praise and other PTIs and highlight important trends in the growing PTI literature base. First, increasing teacher praise promotes meaningful changes in student behavior for students with and without disabilities (Jenkins et al., 2015; Moore et al., 2019). Increases in PTIs promote student on-task behavior, attention, and compliance and reduce student inappropriate behaviors and tardiness (Markelz & Taylor, 2016; Royer et al., 2018; Simonsen et al., 2008). Second, general and special education teachers from kindergarten through high school use significantly more general praise than BSP (Jenkins et al., 2015). Despite decades of research, teachers use this effective strategy with varying frequency, consistency, and success (Reinke et al., 2008). Third, overall PTIs decrease as students' age. Early elementary teachers demonstrate higher rates of PTIs than later-aged teachers and teacher praise prompts greater change in younger students (Markelz & Taylor, 2016; Royer et al., 2018). Fourth, reviews of high-quality praise research establish praise as a potentially evidence-based practice for classroom management (Royer et al., 2018; Moore et al., 2019).

2.1 Purpose of Current Review

Published reviews of PTIs often focus on identifying the current rate with which teachers use praise or on assembling studies to evaluate the evidence base in totality. To date, no review has examined the methodological characteristics of empirical studies of PTIs. The purpose of this study is to systematically review how single-subject researchers have attempted to experimentally

increase teachers' use of PTIs, with a specific focus on their unit of analysis, intervention components, and subsequent empirical and social validity results. The driving research questions include:

1. With whom and where have researchers conducted PTI studies?
2. What target teacher and student behaviors have been targeted within PTI studies?
3. What intervention components have researchers utilized to increase teachers' PTIs?
4. What empirical and social validity results occurred in response to the intervention?

2.2 Method

For this systematic literature review, the author conducted a search of the ERIC and PsycINFO databases, as well as Google Scholar. Keywords for the literature search included various and relevant combinations of *teacher*, *behavior*, *praise*, *approval*, *general praise*, and *behavior specific praise*. The initial database search for this review took place in October 2017, manual searches of article reference sections, review of articles citing historic/seminal praise articles (e.g., White, 1975), and the search engines' "related articles" feature identified additional potential articles. Studies must have been published within peer-reviewed journals (i.e., no dissertations, chapters or sections of books, etc.) and in English, without any limit on the year of publication. Two subsequent database searches took place in June 2018 and October 2018 to identify newly published articles missing from the initial search and all duplicates were removed. In all, the researcher reviewed 705 unique potential articles.

2.2.1 Article Analysis

The author individually reviewed the titles, abstracts, and entire articles (if necessary) within Microsoft Excel[®]. The following inclusion/exclusion criteria were applied to the title, abstracts, and articles during the review: (1) articles explicitly evaluated an intervention to actively manipulate PTI behaviors; surveys assessing social validity of praise, student approval, perception, preference of praise/recognition, or intervention approval measures) met exclusion criteria ($n = 245$, 33.4%); (2) articles presented results of an empirical study resulting in quantitative data for analysis; articles not presenting the results of empirical studies (e.g., practitioner articles, strategies for using praise, or describing assessment or observation protocol) met exclusion criteria ($n = 89$, 12.1%); (3) articles presented results of an empirical study focused on teachers as behavior change agents; research studies not focused on in-service teacher behavior (e.g., measuring positive interactions of parents, pre-service teachers, managers, peers, caregivers, coaches, therapists, professors, undergraduates as teachers, paraprofessionals, or instructors/tutors) met exclusion criteria ($n = 96$, 13.1%); research collecting PTI behavior data to establish the effectiveness of an intervention (including School-Wide Positive Behavior Supports, the Good Behavior Game, etc.) also met exclusion criteria ($n = 40$, 5.5%); research studies including PTI behaviors as part of a treatment or intervention package or as a treatment phase/condition also met exclusion criteria ($n = 111$, 15.7%); and (4) researchers attempted to change a dimension (rate, quality, type, etc.) of PTI behaviors; studies that solely quantified a dimension of PTI behavior(s) but made no attempt to change the dimension met exclusion criteria ($n = 92$, 12.6%).

After applying the aforementioned inclusion/exclusion criteria, 68 empirical studies remained (9.3% of original sample). The author applied a secondary set of exclusion criteria to focus the final analysis: (1) studies not conducted within elementary general education and

inclusive classrooms (kindergarten to fifth grade; $n = 33$, 4.5%); (2) international research ($n = 7$, 1.0%); and (3) studies not employing a single-subject research design ($n = 6$, 0.8%). The sample included 15 studies (2.1% of the original sample).

2.2.2 Article Quality Assessment and Data Extraction

The researcher read and coded each article based on published quality indicators for single-subject research (Cook et al., 2015). The *Council for Exceptional Children's Standards for Evidence-Based Practices* provides twenty-one quality indicators (QI) to examine the methodological soundness of single-subject studies (Cook et al., 2015). The QIs focus on descriptions of the participants, context and setting, intervention agent(s), intervention/practice, implementation fidelity, internal validity, dependent variables, and data analysis. In order to review as many methodologically sound articles as possible, the researcher included articles meeting at least 80% of the QIs. Only eight articles (1.1% of the original sample) met this criterion and were included in the final review sample. The researcher reread each article and coded descriptive characteristics (Figure A1) of each article into a Microsoft Excel[®] workbook for analysis, including: (a) participant demographics; (b) experimental methodology; (c) data collection; (d) accuracy, reliability, and believability; (e) procedural fidelity; and (f) intervention results.

2.3 Results

Tables 1-5 present descriptive characteristics of the reviewed studies. Blank cells represent information missing from the manuscript.

2.3.1 Research Question One – Who and Where

Tables 1-3 report the teacher, student/classroom, and school demographic information provided within the reviewed articles.

2.3.1.1 Teachers

In all, 28 classroom teachers (general education or inclusion) participated in the reviewed studies. The teachers had a wide range of experience (from 1-25 years), but almost half of them ($N = 12$; 48.0%) were ‘early career’ teachers with less than five years of experience. Of the studies that reported it, 21 of the 25 teachers (84.0%) were female and most (90%) were Caucasian or White. More than half of the teachers (61.1%) had a Master’s degree and 38.9% had a Bachelor’s degree. Allday et al. (2012) reported their three classroom teachers were “highly qualified” (pp. 89-90) but did not report their actual degrees/certifications. The five teachers from Horton (1975) and O’Handley et al. (2018) were the only two studies to report age, with an average age of 32.8 years, ranging from 26 to 47 years old.

Table 1. Teacher Demographics Across Reviewed Studies

Study	Num. of Participants	Age	Gender	Race/Ethnicity	Certifications	Teaching Experience (in years)
Allday et al., 2012	3		F, F, F	C, C, C	HQ, HQ, HQ	
Cossairt et al., 1973	3					4, 2, 3
Horton, 1975	2	28, 47	M, M		M, M	4, 15
Mrachko et al., 2017	4		M, M, F, F		MA, MS, MS, BS	14, 1, 14, 5
O’Handley et al., 2018	3	31, 26, 32	F, F, F	C, AA, C	M, BS, BS	1, 4, 8
Reinke et al., 2008	4		F, F, F, F	W, W, W, W		25, 14, 13, 15
Simonsen et al., 2017	6		F, F, F, F, F, F		M, B, M, M, M, M	15, 2, 13, 7, 5, 11
Thompson et al., 2012	3		F, F, F		BA, BA, BS	11, 13, 1
<i>REPORTED</i>	<i>8/8 100%</i>	<i>2/8 25%</i>	<i>7/8 87.5%</i>	<i>3/8 37.5%</i>	<i>5/8 62.5%</i>	<i>7/8 87.5%</i>

Note. Information missing from manuscripts is shown with a blank cell. Teacher demographic information for multiple teachers is provided in the order presented within manuscripts; Gender – F: Female, M: Male; Race/Ethnicity – C: Caucasian, AA: African-American, W: White; Certifications – HQ: “Highly qualified,” M: Master’s degree, MA: Master of Arts, MS: Master of Science, BS: Bachelor of Science; B: Bachelor’s degree, BA: Bachelor of Arts

Table 2. Student/Classroom Demographics Across Reviewed Studies

Study	Unit of Analysis	Grade(s)	Student Ages	Student Genders	Student Races/Ethnicities	Num. of Students in Classroom
Allday et al., 2012	D, T, T	2, K, 1	8, 5, 6, 7, 7	4 M, 1 F	5 C	---
Cossairt et al., 1973	SG, SG, SG	4, 4, 3		5 M, 4 F		---
Horton, 1975	C, C	4, 4				
Mrachko et al., 2017	C, C, C, C	5, 5, 5, 5		11 M, 15 F (x4)		26 (x4)
O’Handley et al., 2018	C, C, C	5, 5, 1		9 M, 5 F; 10 M, 8 F; 7 M, 8 F	14 AA; 18 AA; 15 AA	14, 18, 15
Reinke et al., 2008	C, C, C, C	1, 2, 2, 5				
Simonsen et al., 2017	C, C, C, C, C, C	K, 3, 3, 3, K, 3				19, 17, 17, ~20, 20, 20
Thompson et al., 2012	D, D, D	4, 4, 3	8, 10, 11	3 M	3 W	---
<i>REPORTED</i>	<i>8/8 100%</i>	<i>8/8 100%</i>	<i>2/8 25%</i>	<i>5/8 62.5%</i>	<i>3/8 37.5%</i>	<i>3/8 37.5%</i>

Note. Information missing from manuscripts is shown with a blank cell. Student/classroom demographic information for multiple student participants is provided in the order presented within manuscripts; Units of Analysis – D: Dyad (teacher and 1 student), T: Triad (teacher and 2 students), SG: Small Group (teacher and 4 students), C: Class-wide; Grade(s) – K: Kindergarten; Genders – M: Male, F: Female; Race/Ethnicity – C: Caucasian, AA: African-American, W: White

Table 3. School Demographics Across Reviewed Studies

Study	# of Schools	Geographic Region	Urbanicity	SES	Performance	School Statistics						
						Minority	FRL	S:T	ESL	SE	SWBS	
Allday et al., 2012	2	Southwest US										
Cossairt et al., 1973	2	Kansas City, KS		“Low”								
Horton, 1975	2			“ED”								
Mrachko et al., 2017	1	Southwest PA	Rural		Met AYP							x
O’Handley et al., 2018	1					96%	80%					✓
Reinke et al., 2008	2	Pacific Northwest				23%	76%					
						11%	44%					
Simonsen et al., 2017	2	New England	Suburban		“Excelling” [*]	20%	1%	13:1				
					“Transitioning”	85%	52%	13:1				
Thompson et al., 2012	3	Western US	Suburban				22%		2%	14%		
							61%		14%	19%		
							27%		4%	13%		
<i>REPORTED</i>	8/8 100%	6/8 75%	3/8 37.5%	2/8 25%	2/8 25%	3/8 37.5%	4/8 50%	1/8 12.5%	1/8 12.5%	1/8 12.5%	2/8 25%	

Note. Information missing from manuscripts is shown with a blank cell. School demographic information for multiple schools is provided in the order presented within manuscripts; SES – ED: “economically disadvantaged”; Performance – AYP: annual yearly progress, * identification based on state school performance report; Minority: percent minority students; FRL: free reduced lunch percentage; S:T: student-teacher ratio; ESL: English as a second language percentage; SE: special education percentage; SWBS: school-wide behavior system – x: no system in place, ✓: adequate PBIS implementation

2.3.1.2 Unit of Analysis

In all, more than 340 students participated in the reviewed eight studies – one study (Reinke et al., 2008) did not report the number of students within each of their four participating classrooms. Researchers investigated methods for increasing PTIs with individual students (teacher-student dyads), small groups of students (triads with two students, table groups of four students), or all students within the classroom (class-wide). Most researchers (five studies; 62.5%) explored teachers' positive interactions with their class as a whole (*class-wide*). Researchers in three studies (37.5%) observed teachers' positive interactions with dyads, triads, or small groups of students (*specific students*).

Class-wide. The five articles with a class-wide unit of analysis presented research conducted within 19 teachers' classrooms throughout all elementary grades (kindergarten to fifth). Latter elementary years (third, fourth, and fifth grades) made up a majority of the participating classrooms ($N = 13$; 68.4%), averaging 21.5 students per room. Early elementary classrooms (kindergarten to second grade) had an average of 18 students. Only two studies (Mrachko et al., 2017; O'Handley et al., 2018) provided information on the academic and behavioral needs of students in their research classrooms. For example, Mrachko et al. (2017) reported five students had individualized education plans or behavior plans and their disability categories. Twelve students across three classrooms in O'Handley et al. (2018) received behavioral supports through the schools' response to intervention (RTI) process.

Specific students. The three publications with a unit of analysis focused on specific student(s) shared the results of research conducted with nine teachers and 20 students through kindergarten to 4th grade. Most of the individual students participating in these studies were reported to be white males with histories of aggression, disruption, or noncompliance (Allday et

al., 2012), “low attending and instruction-following behaviors” (Cossairt et al., 1973, p. 90) or being “noncompliant and disruptive in class” (Thompson et al., 2012, p. 526).

2.3.1.3 Settings

The eight studies were conducted in a total of 15 public elementary schools across the country including the Midwestern US, New England, Pacific Northwest, Southwestern Pennsylvania, Southwest US, and Western US. The studies were conducted within suburban (Simonsen et al., 2017; Thompson et al., 2012) and rural (Mrachko et al., 2017) elementary schools of varying socio-economic and characteristics. The schools had varying free/reduced lunch rates, averaging 45.3% (range: 1-80%). The reviewed studies were conducted in schools with adequate implementation of a school-wide behavior system (O’Handley et al., 2018) as well as elementary schools without any school-wide behavior system (Mrachko et al., 2017, Thompson et al., 2012).

2.3.2 Research Question Two – Target Teacher and Student Behaviors

The studies included 30 dependent variables (DVs) – 22 positive and negative teacher behaviors and eight positive and negative student behaviors (see Table 4).

2.3.2.1 Teacher Behaviors

Researchers studied various PTI behaviors, including BSP (Allday et al., 2012; Horton, 1975; Mrachko et al., 2017; O’Handley et al., 2018; Reinke et al., 2008; Simonsen et al., 2017; Thompson et al., 2012), general verbal praise (Allday et al., 2012; Cossairt et al., 1973; Horton, 1975; Mrachko et al., 2017; Reinke et al., 2008), and approving gestures (Mrachko et al., 2017). The seven experiments including BSP provided similar

definitions, however, the most technical, and likely seminal, definition of BSP comes from Horton (1975): BSP has “one component that differentiated it from other praise statements, the identification of the student behavior to be praised, which was attached to a word or group of words that connoted approval or praise” (p. 312). Researchers in four studies (50.0%) also collected data on teachers’ use of negative teaching behaviors, including reprimands/corrections (Allday et al., 2012; Reinke et al., 2008), negative comments/criticisms (Mrachko et al., 2017), praise for non-attending (Cossairt et al., 1973).

2.3.2.2 Student Behaviors

Researchers in five studies (62.5%) collected concurrent data on positive and/or negative student behaviors. Positive student behaviors included ‘on-task’ (Allday et al., 2012; Thompson et al., 2012), ‘attending’ (Cossairt et al., 1973), and ‘appropriate engagement’ (O’Handley et al., 2018). Negative student behaviors included students being ‘disruptive’ (Cossairt et al., 1973; O’Handley et al., 2018; Reinke et al., 2008) and ‘off task’ (Allday et al., 2012; Thompson et al., 2012).

2.3.3 Research Question Three – Intervention Components

Table 4 presents the independent variable (IV) components and research design (including maintenance) for each reviewed study. All of the studies included multi-component IVs for increasing PTIs with specific students or classrooms as a whole. IV components are discussed below by frequency.

Most of the studies ($n = 7$; 87.5%) reported some form of an initial teacher training (e.g., workshops, in-service, professional development, consultation, discussions). Each training occurred during a one-time session lasting approximately 30 minutes (range: 15 minutes to 2 hours). These trainings often included: (a) definition and/or explanation of specific PTI behavior(s); (b) time devoted to rehearsal/role-play of using the PTI behavior(s); (c) discussing examples and non-examples; (d) discussing the PTI behaviors' research-based, effects, and/or benefits; (e) steps or components of implementing the behavior; and/or (f) discussion of baseline levels of teacher behavior(s).

Researchers in a majority of studies ($n = 5$, 62.5%) provided PF to their participants. Three experiments (37.5%) conducted in-person feedback sessions either daily (Cossairt et al., 1973 and Thompson et al., 2012) or weekly (O'Handley et al., 2018). During these sessions, outside researchers shared and/or discussed teacher behavior data (Cossairt et al., 1973; Thompson et al., 2012), set and/or reviewed behavioral goals (O'Handley et al., 2018; Thompson et al., 2012), and reviewed student behavior data (Cossairt et al., 1973; O'Handley et al., 2018). Three experiments (37.5%) included emailed PF for their teachers either daily (Mrachko et al., 2017) or weekly (Allday et al., 2012; Thompson et al., 2012). The emails included: a discussion of teacher behavior data or performance (Allday et al., 2012; Mrachko et al., 2017), goal achievement or underachievement (Allday et al., 2012), review of student behavioral data (Allday et al., 2012),

Table 4. Experimental Methodologies Across Reviewed Studies

Study	Dependent Variables		Independent Variables						Design	Maint.	
	Teacher	Student	Train	GS	PF	VPF	SM	Indiv.			
Allday et al. (2012)	BSP; GP; BSC; GC; TSP; TSC	On-task; Off-task	•		•					MBL-P	
Cossairt et al. (1973)	P for attending or non-attending	Attending; Disruptive bx	•		•					MBL-P	•
Horton (1975)	BSP; GA		•	•		•		•		MBL-S	
Mrachko et al. (2017)	BSP; GP; G; C; R		•	•	•	•				MBL-P	•
O’Handley et al. (2018)	BSP; R	Appropriate engage.; Disruptive bx	•	•	•	•				MBL-P	•
Reinke et al. (2008)	BSP; GP; R	Disruptive bx				•		•	•	MBL-P	•
Simonsen et al. (2017)	BSP		•	•		•		•		MBL-P	•
Thompson et al. (2012)	BSP	On-task	•	•	•			•	•	MP	

Note. Dependent Variables – BSP: behavior-specific praise, GP: general praise, BSC: behavior-specific correction, GC: general correction, TSP: target student praise, TSC: target student correction, P: praise, GA: general approval, G: gesture indicating approval, C: coercion, R: reprimands, Bx: behavior, Engage: engagement.; Independent Variables – Train: teacher training, GS: goal setting, PF: performance feedback, VPF: visual performance feedback, SM: self-monitoring, Indiv: individualized interventions; Design – MBL-P: multiple-baseline across participants, MBL-S: multiple-baseline across [academic] subjects, MP: multiple probe, Maint: maintenance.

specific instances of successful and unsuccessful implementation of BSP (Mrachko et al., 2017), and praise for data collection and improved BSP (Thompson et al., 2012).

Five experiments (62.5%) included the provision of visual performance feedback (VPF) on their participants' behavior. VPF included participants examining their own self-generated graphs (Horton, 1975; Simonsen et al., 2017), discussion of researcher-provided graphs of positive and negative teaching behaviors (Mrachko et al., 2017; O'Handley et al., 2018), and/or corresponding researcher-provided graphs of student behavior (O'Handley et al., 2018). Reinke et al. (2008) did not describe the VPF they provided to their participants.

Researchers in five studies (62.5%) incorporated goal setting to help increase PTIs. Goals were set based upon individual teacher baseline performance (Mrachko et al., 2017; Thompson et al., 2012), predetermined by the researchers (Horton, 1975; O'Handley et al., 2018), or simply chosen by the participating teacher (Simonsen et al., 2017). Only four studies (50.0%) reported their PTI behavior goal – at least one BSP per minute (Horton, 1975; Mrachko et al., 2017), one BSP every two minutes (O'Handley et al., 2018), or a 50% increase from baseline BSP levels (Thompson et al., 2012).

Four experiments (50.0%) incorporated participant self-monitoring in their intervention. Participants in two studies (25.0%) were taught to collect audio (Horton, 1975) or video recordings (Thompson et al., 2012) of their teaching, select a sample of time, and self-score, record, and analyze their data. Reinke et al. (2008) had teachers record their intervention implementation through a provided procedural checklist of intervention components. Participants in Simonsen et al. (2017) established a *self-management plan* similar to a behavior contract; the teachers set their own behavioral goal, reinforcers, etc., and self-reinforced successful increases in BSP.

Researchers in two studies (25.0%) provided individualized (Reinke et al., 2008) or tiered (Thompson et al., 2012) interventions, tailoring them for each participant. Consultants and participants in Reinke et al. (2008) collaborated to develop a menu of intervention options for teachers; the teacher selected their intervention(s) and consultants provided ongoing implementation support. Thompson et al. (2012) followed an RTI approach, with three tiers of supports: school-wide training, self-monitoring, and self-monitoring plus coaching. Increases or decreases in BSP frequency from baseline determined tier placement and applicable supports (Thompson et al., 2012).

2.3.3.1 Research Design

Most of the experiments ($n = 6$, 75.0%) used a multiple baseline (MBL) design across participants. Horton (1975) utilized a MBL across academic subjects (e.g., mathematics, language arts, and reading for the same teacher) and Thompson et al. (2012) employed a multiple probe across participants design. Five experiments (62.5%) included a follow-up or maintenance phase in their experimental designs. These served as checks to the long-term effects of the independent variable. The median follow-up time was 2-3 weeks, with a range of immediately following termination of independent variable (O’Handley et al., 2018) to one month (Reinke et al., 2008).

2.3.4 Research Question Four – Empirical and Social Validity Results

The researcher visually analyzed each data-display for meaningful within- and between-phase changes in teacher behavior, as well as the immediacy of the effect, data overlap, and consistency of data pattern across similar phases (Kratochwill et al., 2013). Generally, the behavior

change interventions effectively increased PTIs. The effects of each intervention varied between the studies, components, and participants.

2.3.4.1 Baseline

As expected, teachers naturally engage in PTI behaviors at varying rates. More than half of the participating teachers rarely provided the measured PTI behavior (praise, BSP, approval statements, etc.). The remaining teachers provided positive statements or praise to students between once every minute (Reinke et al., 2008) to once every five to ten minutes (O’Handley et al., 2018).

2.3.4.2 Teacher Trainings

Although nearly all of the studies providing an initial training or professional development session for teachers about increasing their PTIs with students, only two studies (25.0%) presented data on the effectiveness of trainings alone (Cossairt et al., 1973; Thompson et al., 2012). Trainings alone were ineffective at producing meaningful and sustained changes in PTI behaviors in five teachers.

2.3.4.3 Performance Feedback

Two studies (25.0%) presented the results of how providing just PF affected teachers’ PTI behavior (Allday et al., 2012; Cossairt et al., 1973). The addition of PF produced immediate increases in PTI behavior for most teachers. This increase was then followed by variable responding at levels consistently above baseline. Two teachers demonstrated no immediate change in their PTI behavior, with only minor increases in the following days. Another teacher quickly increased the use of praise, however, it quickly returned to baseline level (zero).

2.3.4.4 Self-Monitoring

Teachers' use of self-monitoring (collecting and recording their own behavior data) also had mixed results in the two studies (25.0%) which presented data on its effectiveness alone (Reinke et al., 2008; Thompson et al., 2012). Self-monitoring helped two teachers demonstrate immediate, sustained improvement over their baseline PTI levels (one each in Reinke et al., 2008 and Thompson et al., 2012). However, self-monitoring alone did not help most teachers in making meaningful, sustained improvements (level, trend, or variability) in teacher praise or BSP.

2.3.4.5 Treatment Packages

Half of the studies reported intervention results for their entire treatment packages. Results for the two most common treatment packages follow. A treatment package including training, goal setting, and PF (Mrachko et al., 2017; O'Handley et al., 2018) helped most teachers demonstrate immediate increases in PTI behavior. Over time this treatment package promoted increases in PTI behaviors (praise, BSP, etc.) to moderately stable, higher levels with little overlap with baseline levels. The behavioral metrics afforded by graphing on the standard celeration chart allowed Mrachko et al. (2017) to precisely quantify their teachers' behavioral changes. Teachers' demonstrated up to a 32% weekly increase in their use of PTIs (Mrachko et al., 2017).

Another common treatment package included training, goal setting, VPF, and self-monitoring. The implementation subject-specific VPF significantly increased two teachers' use of BSP in their mathematics, language arts, and reading classes (Horton, 1975). Smaller BSP increases were seen in both of their science/health and social studies classes (Horton, 1975). A "targeted professional development" (Simonsen et al., 2017, p. 40) led to immediate (albeit varied) increases in BSP rates for their six participating teachers. Teachers maintained these gains for approximately one to two months with only moderate reductions over time (Simonsen et al., 2017).

2.3.4.6 Negative Teacher Behaviors

Only two studies (25.0%) presented data on teachers' negative behaviors, including teachers' negative statements and/or reprimands (Mrachko et al., 2017; O'Handley et al., 2018). Three teachers in Mrachko et al. (2017) reduced their negative statements during intervention between 5% and 53%, while one teacher increased her use of negative statements 14% weekly. Teachers in O'Handley et al. (2018) provided reprimands at lower rates than BSP during intervention, though increases in BSP did not seem to affect teachers' rate of reprimands.

2.3.4.7 Follow-up and Maintenance.

Teacher levels of PTI behaviors typically decreased during follow-up data collection. Teacher behavior seemed to maintain immediately following IV removal (O'Handley et al., 2018). Most teachers had reductions in PTI behavior in the weeks (Cossairt et al., 1973; Mrachko et al., 2017; and Simonsen et al., 2017) and month (Reinke et al., 2008) after intervention removal and repeated follow-up measures were highly variable.

2.3.4.8 Student Behaviors

Increases and decreases in student behavior mirrored changes in teacher behavior. Students' rates of attending and on-task behaviors (Allday et al., 2012; Cossairt et al., 1973; Thompson et al., 2012) and appropriate engagement behaviors (O'Handley et al., 2018) increased concurrently with increased use of PTI behaviors. Additionally, students engaged in reduced rates of disruptive behaviors (O'Handley et al., 2018; Reinke et al., 2008) as teachers increased their use of PTI behaviors. In a correlation analysis, Allday et al. (2012) found a statistically significant relationship ($r = 0.62$, $p < 0.001$) between the rate of teacher BSP and student on-task behaviors.

2.3.4.9 Social Validity.

Six studies (75.0%) included measures of social validity, often relying on a survey or other Likert scale measures at the completion of the study (see Table 6). Three studies (Mrachko et al., 2017; Simonsen et al., 2017; Thompson et al., 2012) reported the rating scale used by participants (e.g., 6-point scale from 1 “*Strongly Disagree*” to 6 “*Strongly Agree*”; Thompson et al., 2012). Three studies did not report their rating scale (Allday et al., 2012; O’Handley et al., 2018; Reinke et al., 2008). Four studies provided the participants with researcher-developed surveys (Allday et al., 2012; Mrachko et al., 2017; Reinke et al., 2008; Thompson et al., 2012), while two provided a research supported and empirically validated measure (O’Handley et al., 2018; Simonsen et al., 2017). The social validity surveys focused primarily on the teachers’ opinion of the IV, including: (a) the effectiveness of the intervention to increase PTIs (O’Handley et al., 2018; Reinke et al., 2008; Simonsen et al., 2017; Thompson et al., 2012); (b) the acceptability or favorability of intervention package (Allday et al., 2012; Mrachko et al., 2017; O’Handley et al., 2018; Simonsen et al., 2017); (c) the appropriateness, adequacy, or intrusiveness of the intervention (Reinke et al., 2008; Simonsen et al., 2017; Thompson et al., 2012); and (d) the extent to which participants would recommend the intervention to others teachers (Mrachko et al., 2017; Simonsen et al., 2017). The social validity questionnaires also asked participants if they perceived a change in student behavior as a result of increasing their BSP (Thompson et al., 2012) and if they are likely to continue using BSP in the future (Allday et al., 2012).

Only two studies (25.0%) included open-ended social validity questions (Mrachko et al., 2017; Thompson et al., 2012). Mrachko et al. (2017) asked their participants to describe which components of intervention package were most helpful, to provide suggestions to improve the training, and if they would recommend the training as professional development for other teachers.

Thompson et al. (2012) did not include their specific open-ended questions, but “two participants also completed a section inviting comments. The researcher encouraged them to be candid in their responses” (p. 533).

2.4 Discussion

The minute-by-minute positive interactions between teachers and students play an important role in students’ classroom experience (Hughes, 2011; McCormic et al., 2013). PTIs include a range effective and efficient classroom management behaviors (i.e., praise, BSP, approving gestures, etc.). Teachers using these behaviors more often establish supportive classroom environments which promote student academic and behavior change (Jenkins et al., 2015; Markelz & Taylor, 2016; Royer et al., 2018). The present study sought to review the reporting and methodological characteristics of PTI literature. The eight reviewed studies used intervention treatment packages (often training, goal-setting, PF, and/or self-monitoring) to effectively increase teachers’ positive interactions with students. Increases in PTI behaviors, including BSP, GP, and approval statements, mirrored positive changes in student behavior. Trends within the reviewed literature provide points for discussion and directions for future PTI research.

PTIs encompass a variety of effective classroom management behaviors, commonly including GP and BSP. Patterns emerged regarding the measurement of PTI behaviors. Although none of the researchers cited seminal or previous literature definitions of their dependent variables, the authors reported similar operational definitions. Definitions of praise commonly included “indication of teacher approval” (Reinke et al., 2008, p. 319) and “conveys general reference to a

desirable behavior” (Allday et al., 2012, p. 88). BSP definitions included “a specific desired social or academic behavior exhibited by the student” (Thompson et al., 2012, p. 528). The convergence of operational definitions within this review is a positive sign for PTI research. Consistent operational definitions increase the consistency between studies and aides in comparing results across studies.

A majority of the operational definitions included the explicit requirement of a verbal or audible statement or behavior. Researchers appear to have prioritized vocal praise behavior(s) over other forms of PTIs. This explicit focus on vocal behavior supports accurate and reliable data collection. However, educators can interact positively with students through more than just verbal praise statements. Researchers within the review handled non-vocal PTI behaviors inconsistently. Two studies included approving gestures within larger categories of positive behavior (Mrachko et al., 2017) and general praise (Reinke et al., 2008). Researchers within the remaining studies excluded these behaviors from data collection explicitly (O’Handley et al., 2018) or through their omission within operational definitions. Approving gestures (e.g., thumbs up), physical touch (e.g., pat on the back), and other non-vocal PTI behaviors will likely act as a positive, primary reinforcer for most students (Cooper et al., 2007). The empirical study of non-vocal PTI behaviors provides an extension of currently available research. Additionally, more inclusive and functionally-equivalent PTI behaviors will allow teachers to vary the form of their PTIs and may help teachers to get closer to the ‘ideal’ or ‘research recommended’ levels. These ratios, however, are merely suggestions and have not yet been empirically tested against each other.

Researchers within the reviewed studies utilized a variety of intervention components to increase teachers use of PTI behaviors. In general, a positive relationship appears between increased researcher involvement in the teachers’ behavior change and the effectiveness of the

intervention. Trainings alone were ineffective at promoting meaningful, sustained improvements in PTI behavior. This supports previous research showing the ineffectiveness of short and/or one-time trainings (Yoon et al., 2007). It appears that experiments with increased researcher interaction with participants (e.g., not relying on participant self-monitoring alone or within treatment packages) promoted more immediate and sustained improvements in PTI behaviors. These methodologies overwhelmingly included PF, which inherently include more consistent interactions with participants. As a result, participants likely receive more access to feedback and opportunities for professional reflection based on ‘expert’ input or suggestions. The effectiveness of interventions including PF differed across experiments. No clear link exists between PF frequency (e.g., daily, every few days, or weekly) or mode (e.g., in-person or email) of and its effect on targeted PTI behavior(s). The *quality* of the PF provided to teachers likely plays an important role in how teachers respond to and incorporate the feedback. Future research should investigate efforts to increase the meaningfulness of PF and interventions on PTIs in general.

All behavior analytic research strives to promote socially-significant change in the participants’ lives. The effectiveness of an intervention is demonstrated through functional relations and high measures of social validity (Snodgrass et al., 2018). Researchers commonly relied on post-study questionnaires as their sole source of critical social validity data. Consistent with previous reviews (Snodgrass et al., 2018), these one-time social validity questionnaires typically focused on the acceptability and effectiveness of the treatment package. Unfortunately, these methodologies often miss important aspects of the participants’ experience during the study.

Teaching teachers to increase their use of PTIs is likely a personally challenging process for the teachers. Collecting and reporting more rigorous and valid social validity measures may increase the believability of empirical findings. The reviewed articles provide some methods to

increase the rigor and validity of social validity findings. Researchers should consider collecting social validity data during intervention (see Reinke et al., 2008). This mid-study data provides in-the-moment input from the participant about their experience as behavior change agents (Snodgrass et al., 2018). Inviting participant input during a study would likely provide the researcher important information about their perceptions of the procedures and feasibility of behavioral targets (Snodgrass et al., 2018).

Researchers should also attempt to incorporate psychometrically validated social validity questionnaires or instruments. Using these measures reduces the commonly cited concerns of researcher-developed social validity instruments, namely their small sample sizes and lack of important accuracy or validity measures. O’Handley et al. (2018) used the Behavior Intervention Rating Scale (BIRS) developed by Elliott and Von Brock Treuting (1991) and Simonsen et al. (2017) used an adapted version of the Intervention Rating Profile-15 (IRP-15; Martens et al., 1985).

Finally, researchers should allow participants to provide oral or written responses to open-ended questions about their experiences (see Mrachko et al., 2017; Thompson et al., 2012). Participants’ responses to open-ended questions verify the findings of self-reported social validity scores. Open-ended questions should focus on the participants’ experience during treatment or intervention and perception of resulting behavior change(s).

3.0 Method

The state of Pennsylvania mandated all schools move to remote instruction in March 2020 due to the COVID-19 pandemic. As a result, study data collection concluded after 38 days. Experiment 1 consisted of observational teaching data for all four teachers across 37 days of instruction. Resulting data was analyzed through quantitative (via descriptive and inferential statistics) and qualitative methodologies. Experiment 2 consisted of a brief case-study involving the first teacher (Adam) during 13 days of baseline and seven days of intervention. Data was analyzed through inferential statistics and metrics afforded by Standard Celeration Charts (SCC).

3.1 Setting

This experiment was conducted within general education classrooms in a private, University-affiliated laboratory school. The school enrolls over 400 students from kindergarten through eighth grade. Most students are White (55.3%) or multiracial (22.9%), with an 11:1 student-teacher ratio.

3.2 Participants

Four general education classroom teachers participated in this study: two language arts teachers and two science teachers. School administration received recruitment materials, including narrative study description, teacher recruitment letter (Figure B2), and full consent document

(Figure B3) to distribute among potential participants. Administration recommended members of the middle school team would benefit from participation in this study. The researcher scheduled an in-depth follow-up conversation with interested teachers via email to discuss the study in more depth and obtain consent to participate. All participating teachers identified a class period which they needed support managing disruptive student behavior. A parental notification letter (Figure B4) was emailed to each classroom parent by the participating teacher. No classroom parents contacted the researcher with questions or concerns about the study.

Tables 5 and 6 present the teacher and classroom demographics for the four teachers participating in Experiment 1. Adam is a White male, Chris is a Black male, and Evelyn and Katie are both White females. The participant’s average age was 32.5 years. Adam, Evelyn, and Katie had PA Instructional I teaching certifications. Katie, Chris, and Adam taught for between two and four years; Evelyn had 23 years of teaching experience, all of which at the study school. The teacher-identified sections included 68 students, an average of 17 students per class. Sixteen students (23.53%) had academic support plans, and three (4.41%) had active behavior support plans. Adam was the first participant to enter intervention, as such, he was the only participant in Experiment 2.

Table 5. Participating Teacher Demographics

Teacher	Age	Gender	Race/Ethnicity	Teaching Certifications	Years Teaching	
					Study School	Total
Adam	29	Male	White/Caucasian	7-12 Biology	4	5
Chris	30	Male	Black	None	4	4
Evelyn	45	Female	White/Caucasian	7-12 Biology	23	23
Katie	26	Female	White	7-12 English Language Arts	2	2

Table 6. Participating Classroom Characteristics

	Number of Students	Academic Support Plans	Behavior Support Plans
Adam – 6 th Grade Science	15	4	1
Chris – 7 th Grade ELA	19	3	0
Evelyn – 8 th Grade Science	17	3	1
Katie – 6 th Grade ELA	17	6	1

3.3 Materials

The four participants received a Zoom Q8 2.3K HD camcorder, 64 GB SanDisk™ memory card, and tripod to video-record the entirety of their lessons with the class section. A provided portable audio-recorder collected an independent audio record of the teachers' vocal instructional behavior. Data collection materials (Figure B5) and procedural fidelity checklists (Figures B6 and B7) were created by the researcher. The researcher created all teacher training materials (Experiment 2) from existing trainings/lectures, advisor guidance, and information from textbooks and online.

3.4 Dependent Variables

Both experiments included the same dependent variables (DV) of rate and number of words associated with teachers' positive and negative vocal interaction behaviors and frequency of classroom disruptive behavior(s).

3.4.1 Teacher Behavior

The primary DV of this study was the frequency of teachers' vocal positive and negative interaction behaviors with students per 16-minute continuous observation. Positive vocal interactions (PVI) were defined as any general praise (GP; e.g., "You all are doing a great job on your project"), behavior-specific praise (BSP; e.g., "Alexandra, thank you for opening your book after finishing the question"), or other positive statement delivered to an individual student, group of students, or the entire class. Teacher PVI research often includes multiple types of praise or approval statements rather than a single topography (Mrachko et al., 2017; Reinke et al., 2013). Negative vocal interactions (NVI) were defined as vocal behavior indicating teacher disapproval or reprimand. Examples of negative vocal statements include coercives (e.g., criticism, sarcasm, pleading, etc.; Latham, 1998, Sidman, 1989), positive or negative statements in response to inappropriate student behavior (e.g., "Now I just said we were not calling out answers"), or other negative statement directed to individual student, group of students, or the entire class. The inclusion of multiple NVI behaviors (including coercives) allows for a fine-grained analysis of teacher's negative behavior(s). Table 7 presents operational definitions for all dependent variables. The researcher also transcribed all words associated with each positive and negative interaction.

3.4.2 Classroom Disruptive Behavior

Data was also collected on disruptive student behavior across each classroom based on a sampling procedure described in the Procedures. Student behavior data (frequency of vocalizations without permission and out-of-seat without permission), condensed across behaviors and students, yielded a total rate of 'classroom disruptive behaviors' per 16-minute observation.

Table 7. Teacher and Student Dependent Variable Definitions

	Behavior	Definition	Examples
Teacher Positives	General Praise	Vocal statement indicating praise or approval without direct mention of specific appropriate behavior	“Good job,” “Nice work.”
	Behavior-Specific Praise	Vocal statement indicating praise or approval with direct mention of specific appropriate behavior	“Thank you for raising your hand,” “I’m glad you have your notes.”
	Other Positive	Other positive vocal statement	“That is correct,” “I agree with that.”
Force	Criticism	Vocal statements which chastise, judge, or highlight a fault in a student’s behavior or performance	“I shouldn’t have to tell you to take your binder out again.”
	Arguing	Vocal statements of disagreement or debate	“I saw you do that,” “Yes you did.”
	Logic/Lecture	Vocal statements using reasoning to convince or support requests	“I’m trying to show you how to do this so you can do it on your own.”
	Questioning	Vocal statements asking rhetorical and/or repeated questions	“Are you listening?” “Why do I have to keep asking you?”
	Sarcasm	Vocal statements of irony, mocking, or teasing	“So now it’s okay to talk,” “Oh, now you speak up.”
	Pleading/Despair	Vocal statements including guilt or appeals for cooperation	“I said please stop,” “I’m tired of waiting for you all to focus.”
	Threats	Vocal statements insinuating or explicating stating forthcoming negative consequences	“I’m about to start switching seats.”
		Vocal statements or direction including forceful language or yelling	“I said take out this work now.”

	Attention to Junk Behavior	Vocal statements in response to identified junk behavior (vocalizations without permission; out of seat without permission)	“Sarah, have a seat,” “You’re muted Kevin.”
	Other Negative	Other negative vocal statements in relation to student behavior (does not include academic corrections)	“John and Billy separate,” “Hey, focus up here.”
Student Behavior	Vocalizations without Permission	Vocalizations (call outs, side conversations, noises, etc.) not in response to teacher opportunity to respond	
	Out of Seat without Permission	Leaving the desk area without teacher permission	

3.5 Experimental Design

This investigatory study collected observational data on teachers’ natural use of PVI and NVI behaviors within their classrooms (Experiment 1). The researcher viewed and coded total of 1,120 minutes (or 18.67 hours) of classroom instruction across 70 recordings. A brief case-study (e.g., AB design; Johnston & Pennypacker, 2009) also assessed the effects of the treatment package for one teacher (Adam) in Experiment 2. Despite limitations, an AB design provides an initial exploration of the effectiveness of training, performance feedback emails, and reflective goal-setting. Adam entered intervention after at least five days with stable or worsening progress (e.g., stable positives and negatives, or decreasing positives and increasing negatives).

Intervention entrance criteria for subsequent participants was set at: (1) stable or worsening progress, (2) the prior participant was in intervention for at least five days, and (3) the prior participant achieved at least a $\times 1.2$ celeration of PVS frequency. Intervention exit criteria was set

at a $\times 1.3$ improvement index change (see Data Collection and Analysis) after at least 10 days of intervention. All participants would exit intervention after 15 days to not overburden the teachers with an extended intervention.

3.6 Data Collection and Analysis

The researcher coded 16-minute segments of instruction, using a random number generator to identify the coding starting point (e.g., start coding at minute 12 after instruction began). Teachers' instructional recordings averaged approximately 51 minutes. Raw frequency data on all dependent variables was collected during all observations. Additionally, the researcher transcribed and recorded the number of words associated with each PVI or NVI statement. Data were also visually displayed on Standard Celeration Charts (SCC). SCCs allow researchers and practitioners to display multiple behaviors (e.g., PVIs and NVIs) on a single, standardized display. This standardized data display prevents scalar manipulations and other possible graph construction errors (Kubina et al., 2017; Lindsley, 2005). Successive calendar days on the horizontal axis allows for an accurate picture of behavior change over time (Datchuk & Kubina, 2011). A logarithmically scaled vertical axis allows for the recording of behavior frequencies from once per day to 1,000 times per minute (Datchuk & Kubina, 2011).

3.6.1 Descriptive Statistics

Mean and standard deviation were calculated for each teacher and student DV across the entire sample ($n = 70$ baseline observations) and for individual teachers. Average words per

instance (WPI) was calculated for each teacher DV following each observation as $\frac{\text{number of DV words}}{\text{DV frequency}}$, with all values of zero removed from the calculation.

3.6.2 Inferential Statistics

Analysis of the observational dataset (Experiment 1) via inferential statistics assessed the presence of significant differences between teachers' use of PVI and NVI behaviors (analysis of variance) and the effect of those behaviors on classroom disruptive behavior (multiple linear regressions)

3.6.2.1 Analysis of Variance.

Multiple one-way between-subjects analysis of variance (ANOVA) were conducted to examine teachers' use of PVI and NVI behaviors. An ANOVA compares the means and variability of a continuous measure (specific teacher behavior) between three or more independent samples (e.g., the four participating teachers). For example, ANOVA quantify the significance of teachers' differential use of criticism or sarcasm in response to student behavior. ANOVA compare the grand mean, or overall sample mean (μ , or "mu"), and teacher's individual means (μ_{T1} , μ_{T2} , μ_{T3} , and μ_{T4}). As with other statistical analyses, ANOVA test the probability with which we can reject, or fail to reject, a null hypothesis (e.g., no effect or difference). The null hypothesis was all individual means are the same; or, there were no significant differences in teacher's use of the behavior ($H_0: \mu = \mu_{T1} = \mu_{T2} = \mu_{T3} = \mu_{T4}$). The alternative hypothesis was at least one individual mean was different from the grand mean ($H_a: \mu_k \neq \mu_l$, where μ_k and μ_l represent any two teacher means).

ANOVA calculate *dispersion* or distribution variability of behavior between and within groups. Each participating teacher is considered their own group. Between-group variability compares teacher means (μ_{T1} , etc.) to the grand mean (μ). As between-group variability increases, teachers' means grow further apart. The mean square (MS) coefficient reflects the sum of each squared deviation for each group, weighted by their sample size, divided by the degrees of freedom (*df*). Within-group variability compares teacher data points to their individual mean. As within-group variability increases, teachers' distributions spread. For within-group variability, the MS coefficient reflects the sum squared deviation of each value from its respective sample mean, divided by the *df*. Comparing these MS coefficients produces an *F*-statistic which, along with its associated *p* value, describes the significance of differences between sample means. *F* statistics near 1.0 or $p > 0.05$ fail to reject the null hypothesis, indicating no difference across teachers. Higher *F* statistics (with $p \leq 0.05$) reject the null hypothesis, indicating significant differences across teachers. Higher *F* statistics result when the between-group variance is larger than the within-group variance.

Using a study DV as an example, ANOVA first calculate the average use of PVIs across all teachers during baseline (grand mean). ANOVA then compare each teacher's individual PVI distribution to the grand mean (e.g., between-group variability). Within-group variability compares each teachers' unique variability in their use of PVI. If the variability of PVI was larger between teachers than within teachers, we would reject the null hypothesis. This would indicate statistically significant differences in teachers' use of PVIs in their classroom.

3.6.2.2 Multiple Linear Regressions

Multiple linear regressions estimate the relationship between a continuous outcome measure (e.g., classroom disruptive behavior) and two or more nominal predictor variables (e.g., teachers' PVI and NVI behaviors). Linear regressions plot a line of best-fit through the scatterplot of a continuous criterion variable (dependent variable) and independent variable (predictor variable). This regression line, therefore, represents a predicted value (\hat{Y} , or "y-hat") for every value of the predictor variable. The change in the predicted value for every one-unit change in the predictor variable represents the slope of the regression line (β , or "beta"). The standard error (SE) calculates the average amount of error (difference) between the predicted and observed criterion values. The ratio between β (slope) and SE produces a computed t -statistic, which, along with its associated p value, describes the significance of relationships between the variables.

For example, a linear regression could estimate the relationship between classroom disruptive behavior and frequency of attention to junk across all baseline data. A line of best fit through a scatterplot of the variables would produce an estimated number of classroom disruptive behaviors (\hat{Y}) for every attention to junk frequency value. The line slope (β coefficient) would then represent the expected number of additional disruptive behaviors for every additional attention to junk behavior statement.

Multiple linear regressions allow for the inclusion of two or more continuous predictor variables (e.g., all teacher DVs). Multiple linear regressions are appropriate when data fit several assumptions: (1) normal distribution of outcome, or *normality*; (2) presence of linear relationships between the outcome and each predictor, or *linearity*; (3) residuals (error values) are approximately normally distributed at each value of the outcome, or *homoscedasticity*; (4) absence of significant

outliers; and (5) absence of high correlations between predictors, or *multicollinearity*. Tests for these assumptions are presented in the results.

3.6.3 Qualitative Analysis of Teachers' Vocal Statements

During coding, the researcher transcribed the statement(s) associated with an instance of each teacher DV. The researcher compiled the transcripts for each DV into individual data files prior to analysis (holistic coding; Saldaña, 2016). The researcher identified universal and idiosyncratic themes while proceeding deductively through each data file. Concept codes (Saldaña, 2016) were generated to describe the unique words and characteristics of the statements.

3.6.4 Assessing the Effectiveness of Intervention Package via Brief Case-Study

The researcher assessed the effectiveness of this brief case-study of one teacher through metrics afforded by charting with a SCC and within-subject *t*-tests.

3.6.4.1 SCC Metrics

SCCs allow for the calculation of objective behavior metrics within conditions (i.e., level, celeration, bounce, and improvement index). *Level* refers to the average rate of responding during observation. Level was calculated as geometric mean, with values of zero removed (Clark-Carter, 2005). *Celeration* refers to the change in behavioral frequency per week and are typically shown on SCCs as dark, solid lines. Celeration is calculated as the behavior frequency over time divided by time, or behavior per minute per week (Kennedy, 2005; Mrachko et al., 2017). Acceleration describes increasing frequency per week. For example, a $\times 1.25$ (“multiply-by one point two five”

or “times one point two five”) represents ‘a 25% increase weekly.’ Deceleration describes decreasing frequency per week. For example, a $\div 1.4$ (“divide-by one point four”) represents ‘a 40% decrease weekly.’ The *bounce* value quantifies the behaviors’ variability over time. Behaviorally, variability decreases when people display the target behavior more consistently and results from tighter stimulus control (Kubina, 2018). Suggested bounce envelopes (or stimulus control guidelines) are: $\times 1.0$ to $\times 3.0$ – very strong stimulus control; $\times 3.0$ to $\times 6.0$ – strong stimulus control; $\times 6.0$ to $\times 10.0$ – moderate stimulus control; and $\times 10.0$ and above – weak, inconsistent stimulus control (Kubina, 2018). *Improvement index* (II) quantifies behavioral improvement across time by comparing the concurrent celerations of corrects (e.g., PVI) to incorrects (e.g., NVI; Kubina, 2019). II is calculated by comparing the growth (acceleration) of corrects to the decay (deceleration) of incorrects within the same condition. If the signs match, divide the larger value by the smaller. If they differ, multiply the values together. Then assign the sign of change to the value (e.g., multiply-by if progress improved or divide-by if progress deteriorated).

SCCs also generate between conditions behavior-change metrics (i.e., level change, celeration multiplier, bounce change, and II change). *Level change* represents the difference in geometric means between concurrent conditions. Level change is calculated by dividing the larger value by the smaller and assigning the sign of the change (e.g., \times for a level increase or \div for a level decrease). *Celeration multiplier* refers to the change in celerations between conditions (Datchuk & Kubina, 2011). Celeration multipliers are calculated as: (a) if celeration signs match, divide the larger value by the smaller and assign the sign of the change; or (b) if celeration signs are different, multiply the values and assign the sign of the change (Pennypacker et al., 2003). *Bounce change* quantifies variability differences between conditions and is calculated by dividing the larger bounce value by the smaller value and assigning the sign of the change. *II change*

compares two η^2 measures between consecutive conditions (Kubina, 2019), and follow the acceleration multiplier and η^2 formulas.

3.6.4.2 t -tests

Within-subjects t -test compare the means of one group at two periods of time. The researcher compared Adam's mean use of PVI and NVI behaviors during baseline and intervention. The two-tailed test had a null hypothesis of no differences in Adam's behavior between conditions ($H_0: \mu_{DIFF} = 0$). The alternative hypothesis was Adam's mean behavior differed between conditions ($H_a: \mu_{DIFF} \neq 0$). Within-subjects t -tests calculate the mean difference between conditions as well as the standard error (variability in means). t statistics above their critical value (absolute values greater than 1.734 for $df = 18$) reject the null hypothesis, indicating significant mean differences between conditions.

For example, t -tests compared Adam's mean use of GP during baseline and intervention. Comparing the difference in means and standard deviations between conditions results in a t statistic. With $df = 18$ (20 observations minus two for each condition), t statistics above a critical absolute value of 1.734 indicates significant differences in Adam's GP between conditions. Within-subject t -tests are appropriate when data fit several assumptions: (1) all participants appear in both conditions/groups; (2) continuous dependent variables; and (3) normal distribution of dependent variable. The case-study data met each of these assumptions.

3.7 Independent Variable

One teacher met intervention entrance criteria (described in Procedures). Adam's baseline and intervention data constitute the data set for Experiment 2. The intervention package included an in-person training on PVI and NVI behaviors, performance feedback emails following each observation, and teachers' reflective goal-setting.

3.7.1 Training

The researcher met with the Adam individually for an after-school training following the last day of baseline. The prerecorded training (slides provided in Figure B8) included: (a) background information about the role and influence of teacher-student interactions; (b) descriptions and examples of positive and negative interactions and the ratio between these behaviors; (c) strategies for increasing positive interactions and decreasing negative interactions (described below); (d) opportunities to practice and discuss the strategies with researcher feedback; (e) presentation of baseline performance to-date on SCCs; and (f) review of the material and description of intervention procedures (i.e., upload data, receive and review feedback, etc.). The opportunity to ask questions and presentation of an informational handout ended the training.

The training provided Adam with three strategies to utilize in his classroom. First, the training discussed maximizing positive statements. Based on decades of behavior analytic research, Adam was encouraged to provide positive praise statements: (a) within three to five seconds; (b) with a sincere and authentic tone of voice and body language; (c) and including explicit mention of appropriate or desired behavior (Conroy et al., 2009). The training also encouraged Adam to vary the targets (e.g., to individual students, small groups, or the class as a

whole), mode (i.e., private versus public, including gesture or signal, etc.), and word choice of his positives. Strategy two taught Adam to withhold verbal/nonverbal attention (or ‘pivot,’ Whitman & Jackson, 2006) from distracting or disruptive student behaviors. Teachers first ignore the behavior while providing positives to other students or engaging in another activity (e.g., writing on lesson plan, checking other student work). Once the disruptive behavior ceases, teachers reengage with the student provide positive statement(s) focused on an appropriate behavior. Finally, Adam was asked to minimize his negative or corrective statements. Adam learned about the predictable side-effects of corrective and coercive statements (i.e., avoidance of and/or escape from coercive people or settings, retaliatory behavior, etc.; Latham, 1998, Sidman, 1989) to highlight the importance of reducing these statements. If negative statements were necessary, the training encouraged Adam to address the behavior, restate his expectations, and move on to another activity/student rather than provide extra attention via long conversations about the undesired behavior. Adam was encouraged to deliver corrective statements in steady tone of voice and look for subsequent opportunities to provide extra positives to the student and classroom. To assess learning, Adam role-played four classroom scenarios focused around challenging behaviors with the researcher. He identified and role-played an appropriate strategy for each situation. The researcher ensured Adam demonstrated each strategy component before moving on to the next scenario.

3.7.2 Performance Feedback Emails

The researcher provided performance feedback emails following each data collection session. The scripted emails included: (a) an introductory greeting; (b) the frequency count of PVIs and NVIs; (c) a specific instance of success (i.e., providing more positives, pivoting from

disruptive behavior, etc.); (d) a specific instance of missed implementation or opportunity for refinement (i.e., attending to disruptive student behavior, missed opportunity to provide positive, etc.); and (e) behavior performance graphs with discussion. Figure B9 presents a performance feedback email template.

3.7.3 Reflective Goal-Setting

After reviewing each performance feedback email, Adam was asked to self-reflects on that days' instruction. He sent an email reply, identifying a reflective statement about something that went well during that data collection session (e.g., *Today, I felt like I had an appropriate rate of positives*) and a 'work on' goal for the next session (e.g., *Tomorrow, I'd like to work on purposefully reducing the frequency of negatives*). A list of the reflective statements was provided within the email (see Figure B9) and included the option Adam to write his own statements (e.g., 'Other').

3.8 Procedures

3.8.1 Pre-Study Observation

One 30-minute pre-study observation within each classroom allowed the researcher to observe the participants' teaching behaviors, collect in-person data on all teacher DVs, and identify the classroom disruptive behaviors (student behavior DV). To increase the social validity of the study, each teacher identified two behaviors they struggled with addressing in the targeted class

(e.g., calls outs, talking to peers, etc.). A list of behaviors was compiled across all teacher responses, with duplicates eliminated. After completing the pre-study observations within each classroom, the behavioral frequencies of each student behavior were summed, and the two most frequent behaviors (vocalizations without permission and out-of-seat without permission) served as the student dependent measures across all classrooms.

3.8.2 Observational Data/Baseline

After completing the pre-study observation, all participants entered baseline. Each participant was taught to turn on the video camera located in the corner of their classroom and a portable audio recorder before students arrived. The teacher recorded their entire lesson (teaching as they normally would) and turned off the equipment after dismissing the class. The teachers uploaded daily video and audio files to a shared Pitt Box account only accessible by the teacher and researcher. During baseline/observation, the participants received no training, performance feedback email, nor graphic data display. Participants remained in baseline until a teacher demonstrated stable, consistent responding for at least five data collection session. Adam was the first teacher to meet entrance criteria after 13 days of baseline.

3.8.3 Training

The day following the last day of baseline, the researcher provided a training (described above) to Adam.

3.8.4 Performance Feedback Emails

Once entering intervention, Adam received performance feedback emails every afternoon following data collection sessions. He reviewed the emails and data displays and responded with two self-reflection statements. Experiment 2 concluded after seven days due to COVID-19.

3.9 Reliability, Believability, and Procedural Integrity

Interobserver agreement (IOA) acted as a measure of believability. A secondary observer (a graduate student in special education) independently scored 20 randomly selected observations, or 28.57% of all classroom recordings. The independent observer received training on each DV definition and data collection procedure. Together, the researcher and secondary coder watched a sample video, discussed each instance of PVI and NVI, and practiced entering data into the coding documents. The researcher and secondary coder engaged in ongoing conversation around coding questions or concerns. The researcher also recoded 15 observations (21.43%) selected through a random integer generator for reliability.

Reliability and IOA were calculated through point-by-point agreement for total frequency of PVIs and NVIs across each minute of data collection. Point-by-point agreement is calculated as $\frac{\text{agreements}}{\text{agreement} + \text{disagreements}} \times 100$ (Kazdin, 2012; Kennedy, 2005). The overall average agreement between observers was 74.33% for PVIs (range, 58.25 - 93.75%) and 46.31% for NVIs (range, 22.57 - 71.42%). The overall average reliability was 85.21% for PVIs (range, 64.58 - 100%) and 77.47% for NVIs (range, 40.74 - 100%).

The researcher completed a treatment integrity checklist following training (Figure B6) and each performance feedback email (Figure B7) to ensure procedural fidelity. Adam's training included all necessary components (100%) and each of the performance feedback emails included all required elements (100%).

3.10 Social Validity

At the end of the study, the researcher provided Adam with a social validity measure containing both closed- and open-ended questions (Figure B10). Adam answered 20 questions about the acceptability and effectiveness of PVIs/NVIs and intervention components. The Likert-style survey was adapted from the Behavior Intervention Rating Scale (BIRS; Von Brock & Elliott, 1987) with a scale of 1 – *strongly disagree* to 6 – *strongly agree*. The BIRS is a well-used treatment acceptability measure, with higher scores indicating greater levels of acceptability (out of 120 possible points). The open-ended questions invited Adam to share his opinion on the intervention components, recommendation of intervention for other teachers, and satisfaction with achieved intervention outcomes.

4.0 Results

The results are organized into five sections: (a) rates and characteristics teacher's baseline use of positive and negative vocal interactions in response to student behavior; (b) rates of classroom disruptive behavior and the relationship between teacher behavior(s) and classroom disruptive behavior; (c) qualitative analysis of teachers' vocal statements; (d) the effectiveness of a novel treatment package within a brief case-study; and (e) one teacher's social validity survey results. Descriptive statistics (i.e., mean, standard deviation, etc.) describe teachers' use of specific PVI and NVI behaviors, as well as classroom disruptive behavior during Experiment 1 (observational data). Multiple linear regressions estimated the relationships between classroom disruptive behavior and specific correlated teacher behaviors. The researcher qualitatively analyzed teachers' positive and negative statements through initial holistic coding, followed by secondary concept and thematic coding (Saldaña, 2016). Finally, SCCs, behavioral metrics, statistical tests, and a social validity survey evaluated the effectiveness of novel intervention package (Experiment 2).

4.1 Rates and Characteristics of Teachers' Positive and Negative Vocal Interactions

The researcher used Stata/SE 16.1 to calculate the descriptive statistics (mean, range, etc.) for each teacher behavior (frequency and number of words) within the observational dataset. The researcher performed ANOVAs in Stata to assess for statistically significant differences in

teacher's use of each DV. ANOVA results add context to the descriptive statistics by highlighting congruence and divergence across teachers.

Tables 8-11 presents descriptive statistics for each teacher DV behavior across all participants' observational data. On average, participating teachers engaged in substantially more negative interactions ($M = 27.23$ per observation) than positive ($M = 5.81$ per observation), representing a positive-negative ratio of 1:4.67. Individual teacher averages ranged from 2.40 (Katie) to 8.14 (Chris) positives per observation and between 20.20 (Evelyn) and 34.08 (Adam) negatives per observation. On top of using negatives more frequently, the participating teachers also provided substantially longer negative statements ($M = 7.47$ WPI) than positive statements ($M = 3.59$ WPI).

4.1.1 Positive Interactions

Teachers engaged in low rates of positive interactions with their students ($M = 5.81$ PVIs per 16-minute observation). While teachers provided significantly different frequencies of positive statements [$F(3, 66) = 4.36, p = 0.0073$], teachers provided one positive nearly every three minutes on average. The positives statements averaged less than four words each ($M = 3.59$ WPI). Teachers provided general praise at higher rates than behavior specific praise with no significant difference across teachers. Individual BSP statements were nearly twice as long as general praise statements – this is not surprising given the extra criteria inherent for BSP. *Other positives* accounted for a majority (82.61%) of teacher's positive interactions with students. These short statements ($M = 3.09$ WPI) typically included acknowledging correct answers or responses and encouragement to keep working.

Table 8. Frequency of Teachers' Positive Vocal Interaction Behaviors per 16-min. Observation

Behavior	Across All	Teacher			
	Teachers (<i>n</i> = 70)	Adam (<i>n</i> = 13)	Chris (<i>n</i> = 22)	Evelyn (<i>n</i> = 15)	Katie (<i>n</i> = 20)
General Praise	0.67 (1.28)	0.38 (0.65)	1.00 (1.57)	0.93 (1.75)	0.35 (0.59)
Behavior-Specific Praise	0.20 (0.47)	0.46 (0.66)	0.00 (0.00)	0.20 (0.41)	0.20 (0.52)
Other Positive	4.80** (5.19)	5.08 (5.85)	7.14 (6.70)	5.07 (3.43)	1.85 (1.57)
Total	5.81** (5.65)	6.69 (6.10)	8.14 (7.27)	6.20 (4.06)	2.40 (1.67)

Note. Cell entries represent mean and (standard deviation) per 16-min. observation.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (resulting from ANOVAs of difference across teachers)

Table 9. Number of Words Associated with Teachers' Positive Vocal Interaction Behaviors

Behavior	Across All	Teacher			
	Teachers (<i>n</i> = 70)	Adam (<i>n</i> = 13)	Chris (<i>n</i> = 22)	Evelyn (<i>n</i> = 15)	Katie (<i>n</i> = 20)
General Praise	9.83 (8.07)	6.25 (2.95)	13.89 (8.81)	11.00 (9.03)	5.17 (3.72)
Behavior-Specific Praise	10.09 (4.08)	7.80 (2.40)	0.00 (0.00)	15.00 (2.16)	9.00 (3.56)
Other Positive	16.05** (15.64)	22.09 (19.53)	22.17 (17.05)	14.69 (10.58)	5.47 (4.30)
Total	20.98* (16.96)	25.45 (21.67)	26.11 (15.07)	24.46 (16.35)	9.31 (7.61)

Note. Cell entries represent mean and (standard deviation) per 16-min. observation.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (resulting from ANOVAs of difference across teachers)

There were no statistically significant differences in teachers use GP and BSP statements. Adam and Katie provided equal, albeit low, levels of GP and BSP. Chris and Evelyn heavily favored GP to BSP. However, teachers demonstrated significant difference in their use of other positives [$F(3, 66) = 4.16, p = 0.0092$]. Adam, Chris, and Evelyn provided similar rates of other positives ($M = 5.50, SD = 5.90; M = 6.35, SD = 6.22; M = 5.07, SD = 3.43$, respectively), more

Table 10. Frequency of Teachers' Negative Vocal Interaction Behaviors per 16-min. Observation

Behavior	Across All	Teacher			
	Teachers (<i>n</i> = 70)	Adam (<i>n</i> = 13)	Chris (<i>n</i> = 22)	Evelyn (<i>n</i> = 15)	Katie (<i>n</i> = 20)
Criticism	4.14*** (4.17)	7.85 (5.03)	2.00 (2.64)	3.93 (3.08)	5.20 (4.10)
Arguing	0.14* (0.52)	0.00 (0.00)	0.05 (0.21)	0.00 (0.00)	0.45 (0.89)
Logic/Lecture	0.94 (1.50)	1.69 (1.60)	0.59 (1.53)	0.47 (0.92)	0.87 (1.54)
Questioning	0.87 (1.54)	0.77 (0.93)	0.68 (1.32)	0.67 (0.98)	1.30 (2.27)
Sarcasm	1.26*** (1.77)	0.77 (0.83)	0.64 (1.14)	3.27 (2.40)	0.75 (1.12)
Pleading	0.09* (0.33)	0.31 (0.63)	0.00 (0.00)	0.00 (0.00)	0.10 (0.31)
Threats	0.24 (0.65)	0.23 (0.60)	0.05 (0.21)	0.20 (0.41)	0.50 (1.00)
Force	0.20* (0.67)	0.23 (0.44)	0.00 (0.00)	0.00 (0.00)	0.55 (1.15)
Attention to Junk	14.13** (9.01)	15.46 (6.70)	16.77 (9.73)	6.73 (5.56)	15.90 (9.12)
Other Negative	4.94 (5.22)	6.77 (7.25)	2.95 (3.29)	4.93 (3.51)	5.95 (6.06)
Total	27.23** (12.98)	34.08 (14.37)	23.73 (10.74)	20.20 (9.88)	31.90 (13.18)

Note. Cell entries represent mean and (standard deviation) per 16-min. observation.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (resulting from ANOVAs of difference across teachers)

frequently than Katie ($M = 1.74$, $SD = 1.52$). Taken together, these results highlight teachers' overall low rate of positive vocal interactions with students.

4.1.2 Negative Interactions

Participating teachers engaged in a high level of negative interactions in response to student behavior ($M = 27.23$, $SD = 12.98$). Teachers in this study provided significantly different rates of

Table 11. Number of Words Associated with Teachers' Negative Vocal Interaction Behaviors

Behavior	Across All	Teacher			
	Teachers (<i>n</i> = 70)	Adam (<i>n</i> = 13)	Chris (<i>n</i> = 22)	Evelyn (<i>n</i> = 15)	Katie (<i>n</i> = 20)
Criticism	39.09** (37.92)	61.01 (47.03)	33.31 (43.24)	30.41 (18.75)	34.18 (30.94)
Arguing	15.17* (7.29)	0.00 (0.00)	8.00 (0.00)	0.00 (0.00)	16.60 (7.17)
Logic/Lecture	39.47* (38.07)	63.33 (51.00)	35.33 (36.03)	28.75 (14.53)	26.09 (18.54)
Questioning	14.13 (15.55)	9.14 (5.14)	21.67 (19.44)	12.33 (8.50)	14.18 (18.66)
Sarcasm	14.45*** (11.64)	12.83 (3.53)	6.14 (2.90)	20.25 (16.42)	14.25 (4.99)
Pleading	12.60 (10.52)	13.33 (11.09)	0.00 (0.00)	0.00 (0.00)	11.50 (9.50)
Threats	17.07 (8.06)	18.50 (8.96)	13.00 (0.00)	13.33 (4.19)	18.67 (8.69)
Force	7.56* (4.90)	5.75 (2.17)	0.00 (0.00)	0.00 (0.00)	9.00 (5.90)
Attention to Junk	61.16*** (52.14)	87.41 (71.84)	63.86 (43.60)	30.50 (21.57)	62.58 (49.92)
Other Negative	46.74* (42.92)	76.35 (66.66)	38.82 (33.42)	38.00 (23.86)	42.18 (33.88)
Total	160.02*** (119.28)	250.28 (162.04)	116.74 (103.78)	130.36 (52.89)	169.74 (100.74)

Note. Cell entries represent mean and (standard deviation) per 16-min. observation.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (resulting from ANOVAs of difference across teachers)

negative interactions with students [$F(3, 66) = 4.73, p = 0.0048$], averaging of one negative statement every 35 seconds. The average negative statement was 7.47 words, with teachers average WPI ranging from 6.50 (Chris) to 10.23 (Adam). Attention to junk behavior (or positive or negative attention directed at a disruptive behavior) accounted for a majority (51.89%) of teacher's negative interactions ($M = 14.13, SD = 9.01$). Other common negative teacher interactions included other negatives ($M = 4.94, SD = 5.22$), criticisms ($M = 4.14, SD = 4.17$), and sarcasm statements ($M =$

1.26, $SD = 1.77$). Teacher's negative statements contained an average of twice as many words ($M = 7.47$) than their positive statements ($M = 3.59$). Teachers demonstrated particularly verbose logic/lecture statements ($M = 20.68$ WPI) about once per observation. In contrast, teachers provided frequent ($M = 14.13$ per observation) but brief ($M = 5.47$ WPI) attention to junk behavior statements.

ANOVA results show unequal means and standard deviations across teachers, indicating teachers use NVI behaviors idiosyncratically. Adam provided more criticisms ($M = 7.85$, $SD = 5.03$) and logic/lecture statements ($M = 1.69$, $SD = 1.60$) than his colleagues. Additionally, Adam's negative statements typically included more words ($M = 10.23$ WPI) than Chris ($M = 6.05$), Evelyn ($M = 7.00$), and Katie ($M = 7.08$). Although Chris provided the second lowest overall number of negative statements ($M = 23.73$, $SD = 10.74$), he engaged in the highest rate of attention to junk behavior ($M = 16.77$, $SD = 9.73$). Chris also had the shortest overall negative statements ($M = 6.50$ WPI). Evelyn used sarcasm ($M = 3.27$, $SD = 2.40$) more often than Adam ($M = 0.83$, $SD = 0.83$), Chris ($M = 0.50$, $SD = 0.89$), and Katie ($M = 0.74$, $SD = 1.15$). She also provided the least attention to junk behavior ($M = 6.73$, $SD = 5.56$) compared to her colleagues. Katie provided comparatively high levels of criticism ($M = 5.20$, $SD = 4.10$) and questioning ($M = 1.30$, $SD = 2.27$) behaviors.

4.2 Classroom Disruptive Behavior

The researcher also calculated descriptive statistics for classroom disruptive behavior, a nominal variable combining the frequencies of vocalizations without permission and out-of-seat without permission. Students in the participating classrooms demonstrated high levels of disruptive behavior (see Table 12). On average, the students engaged in 213.28 disruptive

Table 12. Frequency of Classroom Disruptive Behaviors per 16-min. Observation

Behavior	Across All	Teacher			
	Teachers (n = 60)	Adam (n = 12)	Chris (n = 22)	Evelyn (n = 10)	Katie (n = 16)
Vocalization w/o Permission	205.38*** (102.60)	275.42 (93.81)	207.23 (101.69)	97.90 (52.02)	217.50 (84.06)
Out-of-seat w/o Permission	8.07 (13.13)	18.33 (21.27)	4.18 (4.49)	5.30 (7.94)	7.44 (12.96)
Total	213.28*** (107.68)	293.75 (100.72)	210.95 (103.57)	103.20 (55.76)	224.94 (88.73)

Note. Cell entries represent mean and (standard deviation) per 16-min. observation.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (resulting from ANOVAs of difference across teachers)

behaviors per observation, or over 13 behaviors per minute. Students' vocalizations without permission ($M = 206.98$, $SD = 104.28$) accounted for most (96.30%) classroom disruptive behaviors. ANOVA results show student disruptive behavior varied significantly across classrooms [$F(3, 56) = 7.78$, $p = 0.0002$]. Students in Evelyn's classroom demonstrated the lowest average disruptive behaviors ($M = 103.20$, $SD = 55.76$) when compared to students in Adam's, Chris', and Katie's classrooms ($M = 298.64$, $SD = 104.14$; $M = 210.70$, $SD = 105.20$; $M = 232.00$, $SD = 87.07$; respectively).

4.2.1 Multiple Linear Regression

Multiple linear regressions were then conducted to explore the influence of specific teacher behaviors (i.e., praise, criticism, attention to junk, etc.) on classroom disruptive behavior. The iterative steps which follow explain how the researcher ensured the data met the five assumptions of multiple linear regressions (normality, linearity, homoscedasticity, absence of significant outliers, and absence of multicollinearity).

Figure 1 presents a kernel density plot of the frequency of classroom disruptive behaviors in each classroom. Kernel density plots fit a smooth curve to the density distribution of continuous data. Across teachers, the distributions of disruptive behavior are roughly normally distributed, despite varying distribution peaks and spread. Visual analysis and nonsignificant skewness nor kurtosis results show the outcome values are normally distributed. Pairwise correlations identified the teacher behaviors significantly correlated with classroom disruptive behavior (see Table 13). Visual analysis of scatterplots (Figure 2) showed positive linear relationships between classroom disruptive behavior and most correlated teacher behaviors. The absence of linear relationships with classroom disruptive behavior led to the removal of pleading (frequency and words) and force

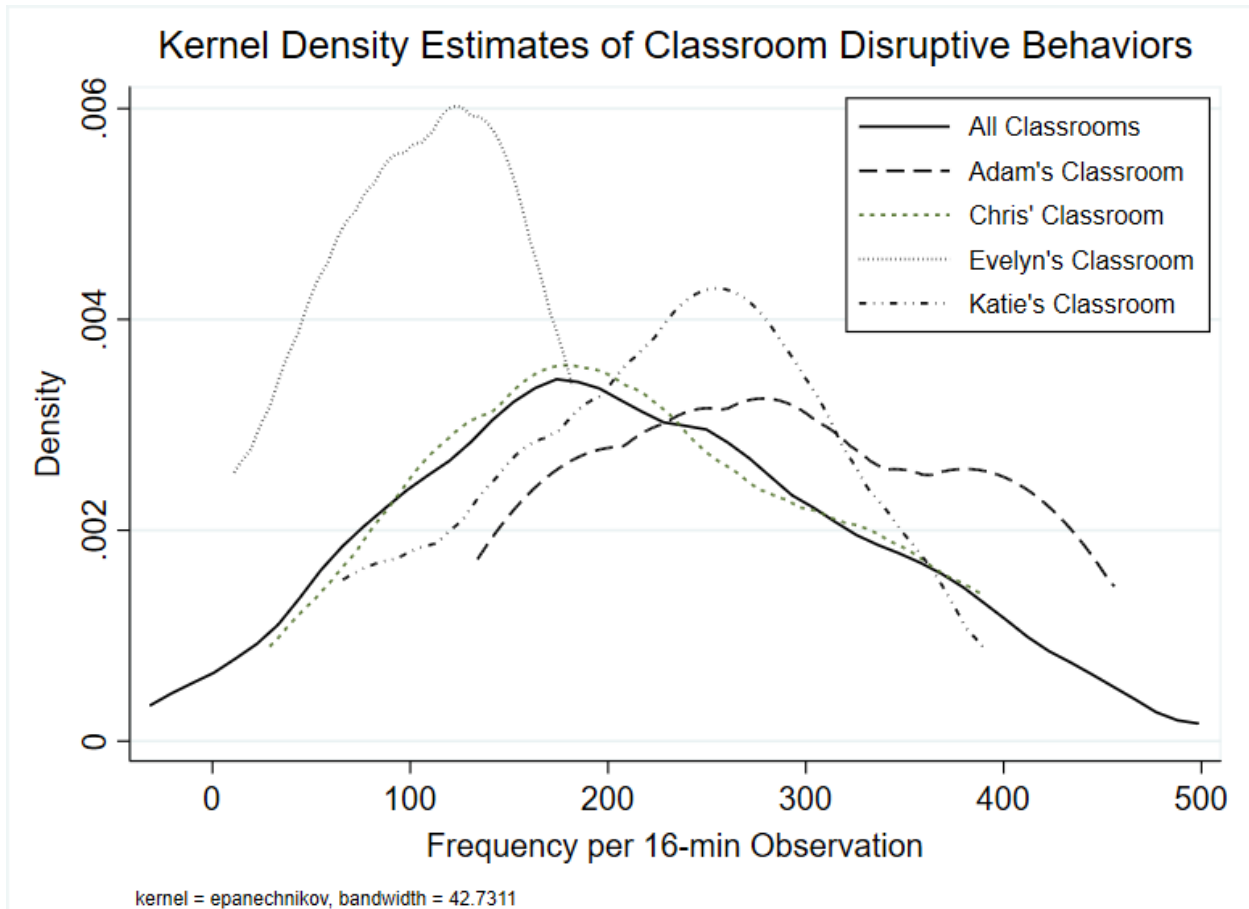


Figure 1. Kernel Density Estimations of Classroom Disruptive Behaviors

Table 13. Pairwise Correlations Between Classroom Disruptive Behavior and Teacher Behaviors

Teacher Behavior	Classroom Disruptive Behavior	
General praise – freq.	0.0754	
General praise – words	0.0382	
Behavior specific praise – freq.	0.1635	
Behavior specific praise – words	0.0336	
Other positive – freq.	0.2423	
Other positive – words	0.2933	*
Total positives – freq.	0.2595	*
Total positives – words	0.2547	*
Criticism – freq.	0.3840	**
Criticism – words	0.3556	**
Arguing – freq.	0.1796	
Arguing – words	0.2015	
Logic/lecture – freq.	0.0993	
Logic/lecture – words	0.0524	
Questioning – freq.	0.1637	
Questioning – words	0.1702	
Sarcasm – freq.	-0.0169	
Sarcasm – words	-0.0391	
Pleading – freq.	0.3650	**
Pleading – words	0.3113	*
Threats – freq.	0.1103	
Threats – words	0.1052	
Force – freq.	0.3749	**
Force – words	0.3378	**
Attention to junk – freq.	0.5023	***
Attention to junk – words	0.5691	***
Other negative – freq.	0.3256	*
Other negative – words	0.3088	*
Total negatives – freq.	0.6460	***
Total negatives – words	0.5510	***

Note. Freq. = frequency; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

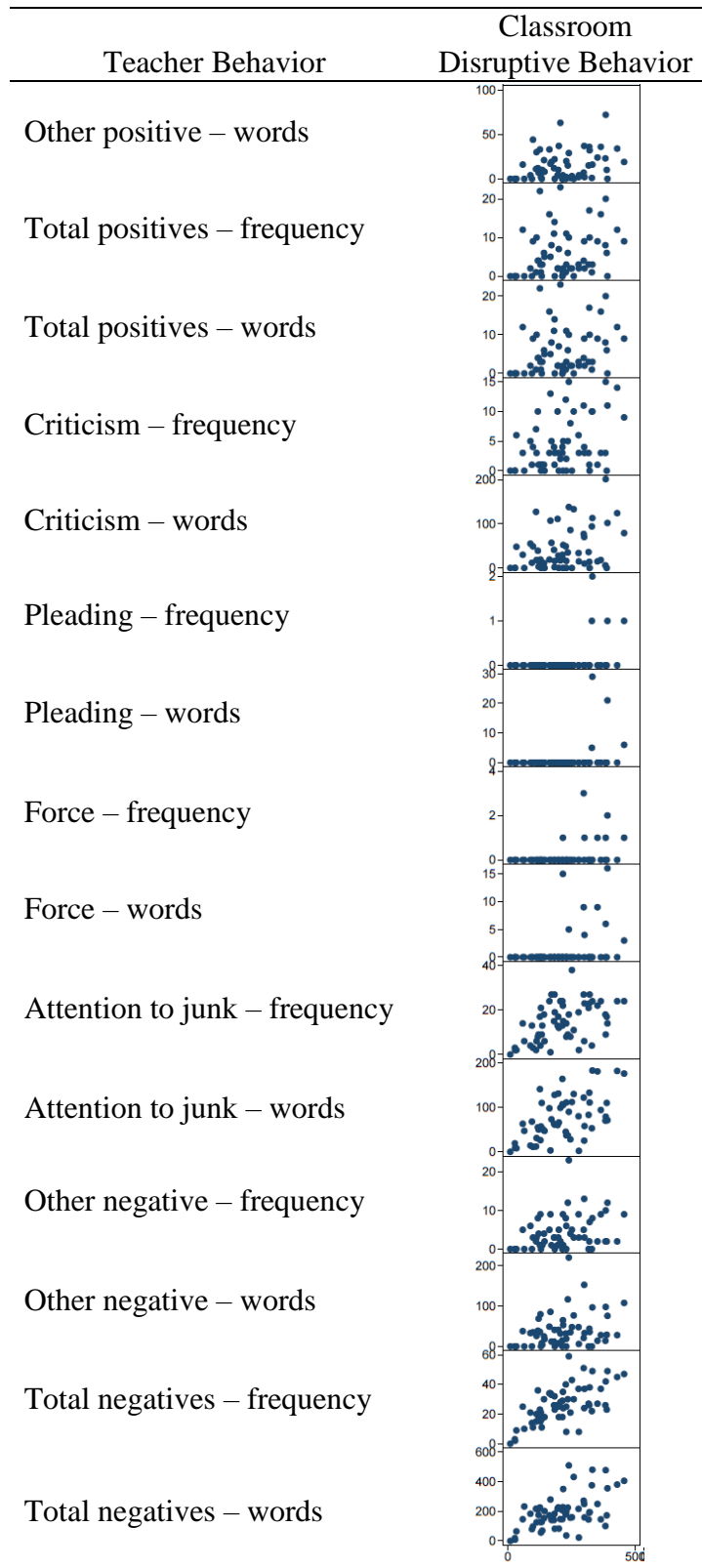


Figure 2. Scatterplots of Classroom Disruptive Behavior and Correlated Teacher Behaviors

(frequency and words) from the working model. Strong correlations between paired frequency and word values (e.g., frequency of criticism and number of criticism words) led to the development of two working models. The working *frequency model* included nominal variables for frequency of *all positive interactions*, *criticism*, *attention to junk*, *other negatives*, and *all negative interactions*. The working *word model* included nominal variables for the total number of words associated with *other positives*, *all positive interactions*, *criticism*, *attention to junk*, *other negatives*, and *all negative interactions*.

Assessing homoscedasticity involved estimating an outcome value (\hat{Y}) for each observation based on the predictors included in the model. Residuals then quantify the distance between the estimated and observed outcome values. Histograms showed relatively normal distributions of residuals, confirming the assumption of homoscedasticity. Cook's distance (Cook's d) was then calculated for each observation in both models to identify observations which too strongly influence the regression model (values greater than $\frac{4}{n}$). Three observations had Cook's d values greater than $\frac{4}{n} = \frac{4}{60} = 0.0\overline{66}$, and were removed from the dataset. Finally, a variance inflation factor (VIF) was computed for each model to assess the presence of multicollinearity (correlations between predictor variables). The absence of multicollinearity indicates each predictor variable accounts for a unique percentage of the model variance. Through iterative analyses, the two final models had mean VIF of 1.19 indicating no multicollinearity within the models. The final *frequency model* included nominal variables for frequency of *all positive interactions*, *criticism*, *attention to junk*, and *other negatives*. The final *word model* included nominal variables for the total number of words associated with *all positive interactions*, *criticism*, *attention to junk*, and *other negatives*.

Tables 14 and 15 present the results of two regression models assessing the influence of specific frequency and word variables on classroom disruptive behaviors within the observational dataset. The average base prediction of classroom disruptive behaviors (e.g., rate when all predictors equal zero) ranged between 50.95 and 73.28 disruptive behaviors per observation. This translates to a rate of 3.18 to 4.58 disruptive behaviors per minute.

Table 14. Frequency Regression Model Results

	β	<i>SE</i>	95% CI	
			LL	UL
Intercept	50.96*	25.22	0.29	101.62
<i>All positive interactions</i>	2.21	2.03	-1.86	6.29
<i>Criticisms</i>	8.13**	2.87	2.36	13.89
<i>Attention to junk</i>	6.48***	1.30	3.87	9.09
<i>Other negatives</i>	5.38	3.54	-1.73	12.49
<i>n</i>	55			
<i>R</i> ²	0.4635			

Note. β cell entries are estimated unstandardized regression coefficients; SE: standard error; LL: lower limit, UL: upper limit.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15. Word Regression Model Results

	β	<i>SE</i>	95% CI	
			LL	UL
Intercept	73.28**	22.72	27.66	118.91
<i>All positive interactions</i>	0.77	0.54	-0.31	1.86
<i>Criticisms</i>	0.42	0.28	-0.13	0.98
<i>Attention to junk</i>	1.24***	0.23	0.78	1.69
<i>Other negatives</i>	0.45	0.41	-0.37	1.28
<i>n</i>	55			
<i>R</i> ²	0.5022			

Note. β cell entries are estimated unstandardized regression coefficients; SE: standard error; LL: lower limit, UL: upper limit.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The set of frequency predictors significantly effect classroom disruptive behavior [$F(4, 50) = 12.66, p < 0.001$] and explain 46.35% of the variance in classroom disruptive behavior. Attention

to junk behavior statements ($p < 0.001$) and criticisms ($p < 0.01$) significantly predicted the number of classroom disruptive behaviors. For every additional attention to junk statement, the students demonstrated an average of six more disruptive behaviors. The students also demonstrated an average of eight more disruptive behaviors for every additional criticism.

Within the second model, the set of word predictors also significantly effect classroom disruptive behavior [$F(4, 50) = 12.61, p < 0.001$] and explain 50.22% of the variance in classroom disruptive behavior. Only the number of words associated with attention to junk behaviors significantly affected the number of classroom disruptive behaviors ($\beta = 1.24, p < 0.001$) – for every additional word in response to junk behavior lead to an average of one additional disruptive behavior.

4.3 Qualitative Analysis of Teachers' Vocal Statements

The researcher also analyzed the common features and themes of teachers' vocal interactions following student behavior. Since substantive empirical work surrounds teacher positives including GP and BSP, the researcher focused analysis on the prevalent themes and characteristics of teacher's NVI behaviors. Table 16 presents the resulting themes and supportive illustrative quotes.

4.3.1 Criticisms

Teachers provided criticisms to a single student, specific students, or the entire class. Criticisms typically included the student's name, often in a raised tone with another indicator of

Table 16. Themes and Illustrative Quotes from Qualitative Analysis of Teacher Behaviors

Teacher Behavior/Theme	Illustrative Quotes
Criticism	
<i>Included student name</i>	<ul style="list-style-type: none"> • <i>NAME, don't be rude.</i> • <i>Which means NAME please put the drawing aside. You should have- NAME, you should have these problems out in front of you.</i>
<i>Identify students not meeting expectations</i>	<ul style="list-style-type: none"> • <i>I've said this multiple times, let's get started please. NAME, please put the book away now. Put away the other distractions. NAME [different student], find this paper. NAME [different student], find this paper. NAME [different student] find this paper.</i>
<i>Broad or imprecise class-wide statements</i>	<ul style="list-style-type: none"> • <i>I will wait.</i> • <i>I'm moving on here. I need you to stick with me because we're wasting too much time today.</i>
<i>Direct mention of undesired behavior</i>	<ul style="list-style-type: none"> • <i>Look at the board instead of your hands.</i> • <i>You're not listening to me.</i> • <i>You can't run in here.</i> • <i>You're taking the easy way out, you just copied it.</i>
<i>Direct mention of desired behavior</i>	<ul style="list-style-type: none"> • <i>Get writing.</i> • <i>I'm noticing people are not writing the equation that they are using.</i> • <i>You need to be following along with every word.</i>
<i>Included personal emotion</i>	<ul style="list-style-type: none"> • <i>I really don't appreciate the attitude you're showing me right now.</i>
<i>Attempts to soften statements</i>	<ul style="list-style-type: none"> • <i>You're a clever guy, this is not helping.</i> • <i>I know you're kidding but, NAME and NAME please stop.</i>
Arguing	
<i>Convince students, get last word</i>	<ul style="list-style-type: none"> • <i>This doesn't require you to talk about anything, so stop touching-</i> • <i>Yes you did, "to be or not to be, that is the question." That's exactly what you wrote.</i> • <i>Yes you did.</i>

Table 16. Themes and Illustrative Quotes from Qualitative Analysis of Teacher Behaviors (continued)

Additional words after stating expectation

- *No, separate. I've already told both of you to stop.*
- *Well you were before I came over here.*

Logic/Lecture

Appeals to reason

- *Because it's a skill that you need. The four main skills we learn here are reading, writing, listening, and speaking. And you have to be able to speak and use that as a skill.*
- *We are working together. You are going to have to do this on your own in just a moment. And then your work is going to be graded. I'm trying to show you how to do this so you can do it on your own.*
- *But what I'm actually telling you is this is very useful. Like, I would bet a good bit that you will use this again. Almost 100% in high school chemistry. Definitely this will be useful in 8th grade physics, and high school physics, in converting anything, for the rest of time. So, if you learn this method now, this will serve you well.*
- *Because it's a requirement*
- *Make sure you have the right answers so this is something that's good quality and correct and something you could study from in the future, please.*
- *I will say this. You are all going to be tasked with a very similar challenge, and it wouldn't be very kind if someone said these things about you. It takes a lot of courage to get up on stage and perform. And when it comes to scenes like that, which are very dramatic and take a lot of effort to go through.*

Compassion statements followed by request

- *As beautiful as our star field was last time, I would really appreciate it if you would just do what I did. And, I'm serious about this. This is not the time to fill the board with your lovely creative artwork. I just want the answer under the problem so we can all see it clearly without all the visual clutter please.*
- *Okay, I like that we're sharing ideas. I do want to take your interests into account so this is a project you will like. Please, let's stop talking over each other.*
- *I said we would do nothing until I get out all of the instructions then I will say go. And I have not said that yet. So I will wait until we are all focused. I want to make sure you have all the instructions, and then you can get started.*

Table 16. Themes and Illustrative Quotes from Qualitative Analysis of Teacher Behaviors (continued)

Questioning

Rhetorical questions

- *Are you following along?*
- *I'm talking, right?*
- *Where's your book? Where's NAME's book?*
- *Why are you drawing NAME? Is that what you're supposed to be doing? Are you supposed to be drawing NAME?*
- *Why do I have to keep asking you?*
- *You just saw that I was yelling at them, so why are you talking?*

Convey expectations

- *Can we get it together please?*
- *Can we settle?*
- *NAME, why is that in your folder and not out and not working on it?*
- *You don't have a pencil? Do you know where the pencils are? Did you check all of them? Did you check by the door? Did you ask someone for a pencil?*

Sarcasm

Convey expectations

- *Third trimester, eighth grade. You need a pencil.*
- *I could answer your question if I had a hand to know who was asking it.*
- *So those of you who are choosing to take notes on this, I would write this down.*
- *You need to look at the math problem I'm doing if this is going to make any sense.*

Tease/mock students

- *That's usually on the title page in the front. That's like primary library class, just saying. Oh okay, that's a good excuse, uh huh.*
- *Judgey Judgerson.*
- *You're so high maintenance, jeez.*
- *You're the seat police today.*

Pleading/Despair

Explicit mention of teacher emotion

- *I'm feeling pretty disrespected today.*
- *I'm tired of having to yell. I'm trying to give some blocking here!*
- *Sixth grade, I'm having a hard time focusing.*

Table 16. Themes and Illustrative Quotes from Qualitative Analysis of Teacher Behaviors (continued)

'If, Then' statement • *If we can get through these, I can start playing a little video at the end.*

Threats

Focused on terminating undesired behavior

- *I'm about to start switching seats.*
- *NAME and NAME, this your last warning before you are all separated and you do not get to sit next to each other again.*
- *So this table is the next place I'm going to move someone, if needed.*
- *So you get one change to show me you can or I move you so you can work better.*

Pressure students to initiate behavior

- *Everyone who is sitting in the perimeter, if you are not looking at your lines right now, that tells me that you have everything memorized, everything is perfect, we can open the show tomorrow.*
- *I'm going around, I'm going around making sure people have their five sentences.*
- *I'm going to cold-call people if you don't answer.*
- *Sixth grade, no one is leaving until everything is cleaned up.*

Force

Demand immediate behavior change

- *NAME, stop talking!*
- *Sixth grade, there will be no talking, because you are abusing this time.*

Used to terminate series of negative interactions

- *Okay, get your paper, get your cards, you're going out with Mr. X.*
- *Write this, now.*
- *When you come back in here for tomorrows class, do not sit beside each other.*

Attention to Junk Behavior

Class-wide

- *I'm waiting for everyone to be sitting in a seat. Waiting for you to be quiet.*
- *Alright- Sixth grade. I need quiet so I can help NAME and NAME here.*
- *I'm giving 30 seconds right now to get back to your seats so we can move on.*
- *Raising hands, not shouting out. So other people get the chance to answer too.*

Table 16. Themes and Illustrative Quotes from Qualitative Analysis of Teacher Behaviors (continued)

Calling out student(s) by name

- *NAME, please sit down. No one is leaving yet.*
- *NAME. NAME, shh. Shh, yes. NAME, you're muted.*
- *Shh, you are not talking when the quiz is out. NAME. NAME. NAME.*

Other Negatives

Convey expectations

- *NAME, NAME, NAME, NAME, put the desk down.*
- *Please ask to use my materials.*
- *So I'm going to give you a choice. If you actually do what I'm asking you to do and stop fighting with me about it, you might get to participate more.*
- *Wait, you guys need to fix these desks. NAME, scrap paper needs cleaned up.*

Influence student behavior(s)

- *NAME, I think you should go grab a drink and then come back. Go ahead.*
- *You gave me the same exact answer on your redo as on your original. Come on dude.*

frustration or anger. Teachers also provided criticisms in which they called out each student not meeting their expectations. Criticisms directed at the entire class were typically broad or imprecise. Some criticisms included teacher's direct mention of the undesired behavior. Such responses provide teacher (and likely peer) attention to a behavior they wish would terminate. Other criticisms included direct mention of a desired alternative behavior. These statements seem to serve as a reminder of behavior expectations or instructional in nature. Teachers occasionally include personal feelings or emotions into criticisms or attempted to soften their criticisms by complimenting the student or undermining their request.

4.3.2 Arguing

Statements coded as arguing prolonged negative teacher-student interactions to convince students or get the last word. Arguing statements also added more negative words into the environment after stating the expectation or request.

4.3.3 Logic/Lecture

Teachers used verbose logic/lecture statements as appeals to reason around skill development and utility, class requirement, or emotions. Logic/lecture statements occasionally included compassionate statements followed by a request or refer back to prior events.

4.3.4 Questioning

Teachers using questioning statements typically asked questions they already knew the answer to and/or a rhetorical question in which an answer was not needed. Additionally, participating teachers conveyed expectations for behavior change phrased as questions.

4.3.5 Sarcasm

Teachers used sarcastic statements to convey teachers' expectations. Teachers also teased or mocked students and behaviors.

4.3.6 Pleading/Despair

Most pleading/despair statements included explicit mention of teachers' emotions to guilt or persuade student compliance. Anecdotally, teachers delivered pleading/despair statements in an emotional way (e.g., raised voice, etc.). The pleading statements also sometimes resembled 'if, then' statements.

4.3.7 Threats

Teachers seemed to use threat statements to compel change in student behavior. Threat statements commonly focused on terminating undesired behaviors. Teachers also used threats to negatively pressure students to initiate specific behavior(s).

4.3.8 Force

Teachers used force statements to demand immediate behavior change. Sometimes force statements terminated a series of negative interactions. Verbal force statements were challenging to identify given the role of teacher's intent in using the statement.

4.3.9 Attention to Junk Behavior

Any teacher comments in response to vocalizations without permission or out-of-seat without permission qualified as attention to junk behavior. Teachers directed responses to the class a whole. Teachers commonly named one or more students explicitly in their attention to junk statements. Teacher's attention to junk statements also included acknowledging, responding to, or accepting called out responses from students.

4.3.10 Other Negatives

Teachers used other negative or corrective comments to convey classroom expectations or influence student(s) behavior.

4.4 Effectiveness of Intervention Package via Brief Case-Study

Figure 3 presents Adam's use of positive and negative vocal interactions with his students during 13 days of baseline and seven days of intervention. In the graphs of teacher behavior (3a and 3b), dots represent PVIs and Xs represent NVIs. Celeration and bounce envelopes (solid for

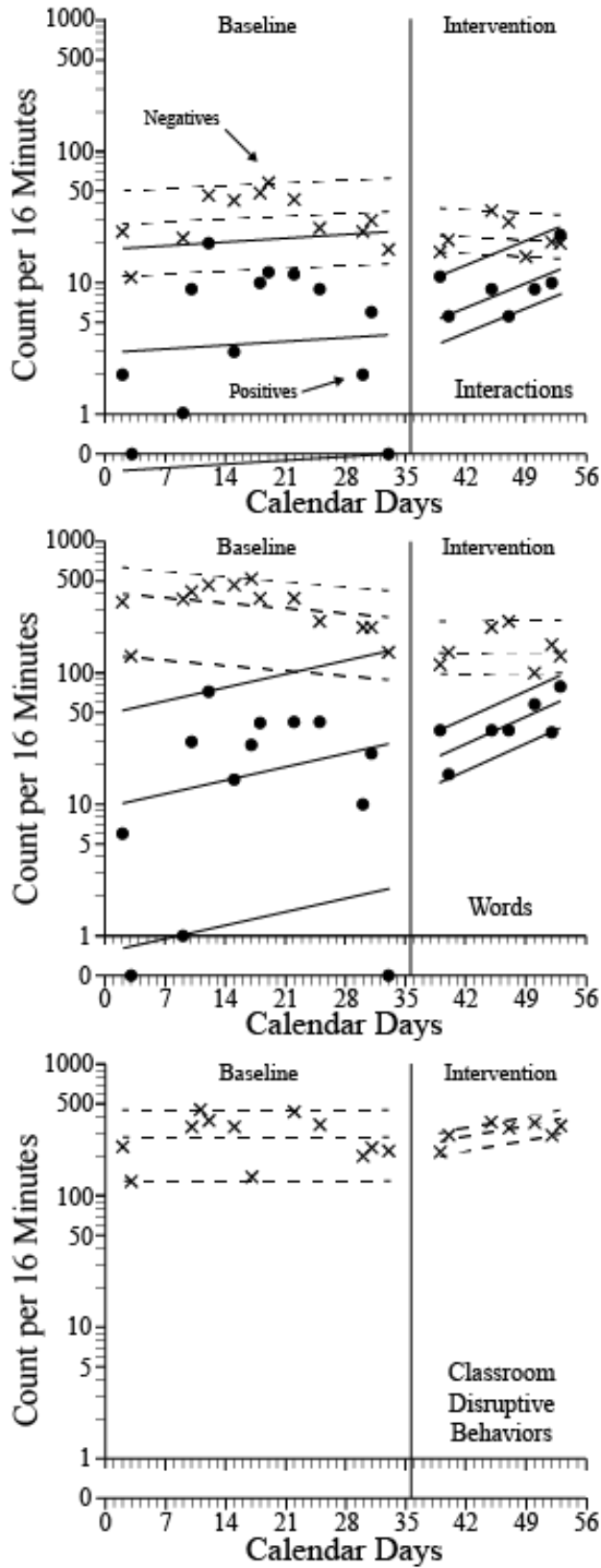


Figure 3. SCCs for Adam's Positive and Negative Verbal Interactions (Frequency and Words) and Classroom Disruptive Behavior

Table 17. Within-Condition and Between-Condition Measures of Adam’s Positive and Negative Vocal Interaction Frequencies

	Baseline			Intervention			Baseline to Intervention					
	Level*	Celeration	Bounce	Improv. Index	Level*	Celeration	Bounce	Improv. Index	Level Change	Celeration Multiplier	Bounce Change	II Change
PVI	5.64	×1.08	×50.00	÷1.05	8.43	×1.35	×3.20	×1.40	×1.49	×1.25	÷15.63	×1.47
NVI	30.95	×1.03	×4.50		19.80	÷1.04	×2.20		÷1.56	÷1.07	÷2.05	

Note. *: frequency per 16-minute observation (geometric mean); Improv. Index: Improvement Index; II Change: Improvement Index Change

Table 18. Within-Condition and Between-Condition Measures of Words Associated with Adam’s Positive and Negative Vocal Interactions

	Baseline			Intervention			Baseline to Intervention					
	Level*	Celeration	Bounce	Improv. Index	Level*	Celeration	Bounce	Improv. Index	Level Change	Celeration Multiplier	Bounce Change	II Change
PVI	19.16	×1.19	×65.00	×1.31	39.79	×1.53	×2.50	×1.51	×2.08	×1.82	÷26.00	×1.15
NVI	307.47	÷1.10	×5.00		162.80	×1.01	×2.50		÷1.89	÷1.09	÷2.00	

Note. *: frequency per 16-minute observation (geometric mean); Improv. Index: Improvement Index; II Change: Improvement Index Change

Table 19. Within-Condition and Between-Condition Measures of Classroom Disruptive Behavior

	Baseline			Intervention			Baseline to Intervention		
	Level*	Celeration	Bounce	Level*	Celeration	Bounce	Level Change	Celeration Multiplier	Bounce Change
	277.12	×1.00	×3.30	305.82	×1.17	×1.70	×1.10	×1.17	÷1.94

Note. *: frequency per 16-minute observation (geometric mean)

PVIs and dashed for NVIs) appear within each phase, and solid vertical lines signify transitions between phases. Tables 16 – 18 show the within-condition and between-condition measures of Adam’s PVIs, NVIs, and disruptive classroom behavior. The researcher analyzed the effectiveness of this intervention through visual analysis and within- and between-condition metrics afforded by the SCC as well as statistical tests (e.g., within-subject *t*-test).

4.4.1 Baseline

Adam used substantially fewer positive statements than negative (a positive-negative ratio of 1:5.49) in baseline, consistent with the sample as a whole. Considerable baseline bounce envelopes highlight large variability in Adam’s use of positives and words associated with said positive interactions. A baseline improvement index of $\div 1.05$ indicates a slight worsening of behavior over time (using less positives and more negatives).

4.4.2 Intervention

Adam demonstrated meaningful improvements in his interactions with students during a short intervention. Accelerating positives, decelerating negatives, and smaller intervention bounce envelopes show Adam improved his use of PVIs. Adam provided significantly more general praise statements following the addition of training, performance feedback emails, and reflective goal-setting [$t(18) = -6.02, p < .001$]. Adam significantly reduced his overall use of NVIs from baseline to intervention [$t(18) = 2.34, p < .05$] and total negative words [$t(18) = 3.22, p < .01$]. While Adam demonstrated less of each NVI behavior during intervention, only his use of logic/lecture statements reduced significantly ($t(18) = 2.49, p < .05$). Bounce envelopes shrank considerably

during intervention, reaching levels demonstrating strong stimulus control and preliminary experimental effect. An improvement index change of $\times 1.47$ indicates Adam improved his performance by 47% from baseline to intervention. Interestingly, Adam demonstrated a slower deceleration of negatives ($\div 1.07$), indicating it may be harder for teachers to reduce NVIs than increase PVIs. A slight increase in classroom disruptive behavior occurred from baseline to intervention. However, this increase was non-significant.

4.4.3 Social Validity

Adam completed a Likert-style survey and responded in writing to several open-ended questions following the termination of the study to describe his perceptions of intervention acceptability and effectiveness (Figure 4). Adam's total social validity score was 113 out of a possible 120 points (94.17%). Adam found the procedures (training, performance feedback emails, and reflective goal-setting) enjoyable, easy, and a fair way to address classroom disruptive behaviors. He expanded:

The training was brief and direct enough to not be an overload of information, and the process of reflecting and getting consistent feedback kept the strategies fresh and helped me notice opportunities to use more of them or helped to remind me of strategies that would be appropriate for different situations that arose or remind me of strategies I was using less than others.

Despite only receiving seven days of intervention, Adam felt he “*was making small but measurable improvements*” in his PVI/NVI behaviors which “*will be on [his] mind in [his] future teaching.*”

	Strongly Disagree	1	2	3	4	5	Strongly Agree
1. This was an acceptable intervention for the children's problem behavior.	1	2	3	4	5	6	6
2. Most teachers would find this intervention appropriate for challenging behaviors.	1	2	3	4	5	6	5
3. The intervention proved effective in changing children's problem behavior.	1	2	3	4	5	6	5
4. I would suggest the use of this intervention to other teachers.	1	2	3	4	5	6	6
5. The children's behavior problem was severe enough to warrant the use of this intervention.	1	2	3	4	5	6	5
6. Most teachers would find this intervention suitable for the behavior problem described.	1	2	3	4	5	6	6
7. I would be willing to use this intervention in other classroom settings.	1	2	3	4	5	6	6
8. The intervention did not result in negative side effects for the child.	1	2	3	4	5	6	6
9. The intervention would be an appropriate intervention for a variety of children/classrooms.	1	2	3	4	5	6	6
10. The intervention is consistent with those I have used in classroom settings.	1	2	3	4	5	6	6
11. The intervention was a fair way to address the children's problem behavior.	1	2	3	4	5	6	6
12. The intervention is reasonable for the behavior problem described.	1	2	3	4	5	6	6
13. I liked the procedures used in the intervention.	1	2	3	4	5	6	5
14. This intervention was a good way to handle the children's challenging behaviors.	1	2	3	4	5	6	6
15. Overall, the intervention was beneficial for the children.	1	2	3	4	5	6	5
16. Overall, the intervention was beneficial for myself as a teacher.	1	2	3	4	5	6	6
17. The intervention would produce a lasting improvement in my behavior.	1	2	3	4	5	6	6
18. Soon after using the intervention, I noticed a positive change in the children's problem behavior.	1	2	3	4	5	6	4*
19. I will continue to use this intervention in my future teaching.	1	2	3	4	5	6	6
20. The procedures were easy and other teachers should also respond well to them.	1	2	3	4	5	6	6

Adapted from Behavior Intervention Rating Scale (BIRS) – Von Brock, M., & Elliott, S. (1987). Influence of treatment effectiveness information on the acceptability of classroom interventions. Journal of School Psychology, 25, 131-144.

*I would rate this higher, but we only got a few days into the intervention after baseline. I'm confident with more time I would have continued to see improved behavioral results. I think the behavioral science informing the intervention is sound and I was becoming better at implementing the intervention with continued practice.

Figure 4. Adam's Completed Social Validity Measure

This study included: (a) trainings on positive and negative verbal interactions with students, (b) performance feedback emails, and (c) reflective goal-setting about instructional strength and 'work on' strategy for the next day. Do you think these components were effective at changing your instructional behaviors?

I think these components were effective. The training was brief and direct enough to not be an overload of information, and the process of reflecting and getting consistent feedback kept the strategies fresh and helped me notice opportunities to use more of them or helped to remind me of strategies that would be appropriate for different situations that arose or remind me of strategies I was using less than others.

Would you recommend this type of intervention/procedures with other teachers? Why or why not?

I would recommend it. It was not difficult to learn or implement, and I think it would be hard to argue that giving more positive feedback and fostering a more positive class climate has any significant drawbacks for instruction.

What components of the intervention/procedures did you like the most? The least? What aspects would you recommend changing before using this with other teachers?

I liked the feedback from the researcher because, as I mentioned above, it kept the strategies fresh in my mind and encouraged me to continue trying strategies and finding the best places to use them. It was also good to hear if my reflections on what went well and what could be improved on matched the thoughts of an outside observer. The most challenging part was in my efforts to reduce my negative feedback, I felt I was sometimes losing time or ceding more control of the classroom to students than I was comfortable with. I think with more practice and with implementing this type of climate from the beginning of a school year these issues would be negligible, as these techniques would become a normal part of my teaching practice. I don't have much to recommend changing.

Are you satisfied with the outcomes of the study? How has this study affected your teaching style/work, if any?

Unfortunately the study was cut short by the COVID-19 crisis, but at the time we ended I felt that I was making small but measurable improvements in the metric for myself being tracked (positive to negative feedback ratio). It was an agreeable enough idea and easy enough to implement and self-monitor that it will continue to be on my mind in my future teaching. Even with distance learning, I have been trying to give more quick positive and specific praise on things I like in student work, though it is not quite the same as managing behaviors in a classroom environment.

Is there anything else you'd like the researcher or research team to know about?

I have nothing substantive to add. I will be curious to know if any data were found that have a bearing on the suggested 4:1 Positive:Negative ratio from other psychological research that was referenced in the training for this study.

Figure 4. Adam's Completed Social Validity Measure, con't

5.0 Discussion

Teachers' interactions with students contribute considerably to an ever-evolving classroom environment (e.g., Hughes, 2011; McCormic et al., 2013). Historically, researchers largely focused on improving teacher positives as a method for altering classroom environments (Jenkins et al., 2015). Other descriptive and experimental work grouped the majority of teachers' interaction behavior with students into two overarching categories: positives and negatives (e.g., Caldarella et al., 2020; Sarno Owens et al., 2018). Consolidating topographically independent behaviors into categories makes fine-grained analyses of teachers' *specific* behaviors difficult. The ability to understand the frequency, role, and effect of specific teacher interactions also suffers. A greater empirical understanding of the nature, qualities, and types of teachers positive and negative vocal interactions with students appears critical. The present study sought to: (a) describe the natural rates and characteristics associated with teachers' positive and negative vocal interactions towards student behavior (Experiment 1), and (b) examine the preliminary effectiveness of a low-effort intervention package to alter teachers' interactions with students (Experiment 2).

Participating teachers delivered low rates of positive statements along with frequent and verbose negative statements regarding student behavior. Higher rates of teacher negatives support previous research (e.g., Apter et al., 2010; Sarno Owens et al., 2018). However, teachers in other descriptive studies provided more positives and less negatives than most of the participating teachers (e.g., Caldarella et al., 2020; Floress et al., 2018; Reddy et al., 2013). Divergent results in comparison to previous studies suggests teachers vary considerably in their use of positive vocal interaction (PVI) and negative vocal interaction (NVI) behaviors. Teachers' variability likely stems from multiple places, including pre-service training, school environment/ethos, and teaching

history. However, preservice teachers are less likely to use force or controlling statements when addressing off-task behavior or peer interactions (Black et al., 2016). Teachers apparently *learn* that NVIs effectively stop behavior throughout their time in the classroom. As a result, teachers continue to use these statements without training in the predictable side-effects (e.g., escape, avoidance, learned helplessness; Sutherland & Singh, 2004; Sutherland & Wehby, 2001).

Descriptive and inferential statistics show teachers used distinct collections of statements towards student behavior. Participating teachers provided a collection of positive statements, including general praise (GP) and behavior-specific praise (BSP). However, correlation and regression results found no significant connections between the individual PVI behaviors and classroom disruptive behavior. This may be due to the relative scarcity of ‘positive’ datapoints across teachers. Similar to Sarno Owens et al. (2018), the relationships diverge from previous research that suggests teacher ‘positive’ behavior and student outcomes are positively related (e.g., Apter et al., 2010; Caldarella et al., 2020; Floress et al., 2018). Differing empirical results suggests the distinction between GP and BSP may be less important for teachers providing low rates of PVIs. In classrooms with lower rates of positive reinforcement it may be more powerful to increase the rate in totality, rather than specific PVI topographies. Whereas teachers using praise at higher rates may benefit from increasing BSP to foster the development of specific desired behaviors through positive reinforcement (as in Floress et al., 2018).

Participating teachers used multiple types of negative statements. Contrary to PVIs, multiple NVI behaviors significantly correlated with classroom disruptive behavior. Regression results (Experiment 1) indicate criticisms (frequency) and attention to junk (frequency and number of words) statements significantly influence the frequency of classroom disruptive behaviors, mirroring previous work (e.g., Caldarella et al., 2020). Teachers NVIs appear more influential on

student behavior. Exploring the impact of teachers' negative interactions across 10 behavioral topographies provides a fine-tuned analysis contributing to the research base on teacher behavior. NVI use appears ubiquitous – teachers failed to assess the function of student behavior. Many vocalizations without permission, a challenging student behavior identified by all participating teachers, may function as attention-seeking behavior. By responding to such call outs, teachers match the function of the undesired behavior and thus increase the future probability of that behavior.

The presence of variably-influential teacher negatives highlights the complicated impact of NVIs, and calls into question the utility and feasibility of proposed one-size-fits-all positive-to-negative ratios (PNRs). Participating teachers' ratios fall well below PNR recommendations of four positives for every negative or five positives to one negative (Cook et al., 2017; Sabey et al., 2019), consistent with previous research (Cook et al., 2017). Collapsing teacher behavior into PNRs presents multiple methodological and conceptual concerns.

First, such reductions to dimensionless behavior (e.g., four-to-one) fails to account for important contextual attributes of time and function. Research has yet to what effect a 4:1 ratio has at different time periods (across 10 minutes, 20, 30, etc.) or at different magnitudes (4:1, 40:10, 400:100, etc.). Second, given students spend a considerable amount of time in classrooms, PNRs likely have cumulative effects on the classroom environment. For example, a short albeit forceful negative statement likely carries longer-term implications than four inauthentic positives. Alternatively, one high quality BSP statement may carry more weight than four minor negative interactions. Third, PNRs fail to account for the wide-ranging topographies and characteristics of teacher negatives. Between-teacher differences in NVI behaviors highlights the necessity for a finer-grained classification of teachers' negative behaviors. Differences in teachers' body

language, delivery style, and other characteristics may also influence the impact of their vocal statements. Exploring NVIs as independent behavioral topographies, rather than condensable equivalent behaviors, could help educational researchers develop a deeper understanding of teachers' minute-by-interactions within their classrooms.

5.1 Coercion in the Classroom

A vast majority of the empirical record surrounding teacher's instructional behavior focuses on understanding and altering the way teachers respond to student behavior. Most research, however, focuses on praise or other 'positives' in isolation of negative behaviors. Previous studies summed teacher negative vocal behavior into larger categories, such as "corrective feedback" (Reddy et al., 2013), "reprimands" (Caldarella et al., 2020), or "social or behavioral negative" (Apter et al., 2010). Exploring teachers' negatives in terms of coercives (e.g., Latham, 1997; Sidman, 1989) provides a more nuanced understanding of the specific behavioral topographies teachers use in relation to student behavior. Classifying negatives across 10 categories shows teachers rely on individual coercive/negative 'crutches' while teaching. For example, while all participating teachers provided a high level of attention to junk statements and criticisms, Evelyn provided significantly more sarcastic statements than her teammates. Given regression results show all coercives are not equal (e.g., only attention to junk statements and criticisms influenced classroom disruptive behavior), this classification also allows researchers to target specific negative DVs and provides teachers with more concrete direction for behavior change. Collaborating with teachers to reduce specific NVI crutches would likely contribute to more positive classrooms.

Coercion plays a critical lasting role in students' educational experience (Mainhard et al., 2011). Exploring the qualitative aspects of teachers' coercives in their classrooms provides an important starting point to understand their complex verbal repertoire of NVIs. First, teachers used various coercives (including criticism, questioning, logic, and sarcasm) in an apparent attempt to convey classroom or behavioral expectations. Teachers conveyed expectations either by explicitly requesting desired alternative behavior or by calling attention to undesired behavior. Additionally, teachers' coercives also sought to terminate or otherwise influence undesired student behaviors. Subjectively "harsher" NVIs (such as criticism, threats, force, etc.) typically terminates the undesired behavior quickly through positive punishment – adding an aversive stimulus into the environment to reduce/stop the behaviors (Cooper et al., 2020). As a result, teachers likely value this momentary relief and continue using the coercive crutches because they work (e.g., the statements elicit immediate, or near immediate, cessation of the targeted behavior). However, since punishment ultimately reinforces the punisher, it is likely teachers continue to use their coercive crutches because they work (e.g., the statements elicit immediate, or near immediate, cessation of the targeted behavior). Finally, teachers often called students out by name when responding to student behavior. While this likely served to get the student's attention, educators should be mindful of the compounding effects of a student consistently called out or only hearing their name within NVIs. Taken alongside descriptive and inferential results, reducing more frequent, corrective coercives may have greater payoff for teachers and researchers. Shifting focus in the face of disruptive behaviors (e.g., pivoting) would allow teachers to convey expectations through positive interactions with students.

The inclusion of coercives as teacher dependent variables allows for an empirical reexamination of coercives in the classroom. Coercives as theoretical constructs resulted from

studies of relational power dynamics, such as marriages or families. Researchers saw the obvious connections to the power relationships inherent within classrooms and offered educational coercives (Latham, 1997, 1998; Sidman, 1989). Coercives are presented as a collection of topographically independent behaviors which teachers should avoid. However, a dearth of empirical data on teachers' coercives tendencies exists. Results from the current experiments shows coercives may be a more staple component of instruction than previously believed. Rather than present a collective "don't do these" mentality based on theoretical definitions and training, researchers could be better served by relying on data to identify teachers' coercives behaviors of concern. Descriptive and inferential results also challenge the flat-organized aggregate of equally influential coercives presented in foundation work. Instead, a hierarchy may exist in which certain coercive behaviors are empirically more impactful on student behaviors. These hierarchies may also vary across individuals, adding additional support for the need for teacher-level assessments regarding their own NVI behaviors. Coercives also fails to include attention to junk statements, which are often dismissed in empirical work. Results show teachers routinely use attention to junk statements during instruction and have significant impact on classroom behavior. Therefore, attention to junk appears to be a staple, conditioned aspect of classroom instruction warranting significant empirical investigation.

5.2 Intervention Effectiveness

Training, performance feedback (PF) emails, and reflective goal-setting helped Adam adjust his instructional behavior in just seven days of intervention. Results support findings across the substantial PF literature base (e.g., Cavanaugh, 2013; Cornelius & Nagro, 2014; Fallon et al.,

2015). Reflective goal-setting, however, represents a novel intervention approach. Taking time to reflect critically on his instructional strengths and opportunities for improvement promoted changes to Adam's vocal interactions in response to student behavior. It is possible, however, that teachers may need training or coaching on how to reflect on their own instruction (Gün, 2011) or require support through the reflection process (Barton et al., 2011). Reflective journaling (see Rathel et al., 2014) provides participants the opportunity to consider their reactions to, and incorporation of, researcher-provided feedback. The production of a permanent product (e.g., reflective journal) allows researchers to shape future interventions components in a way that better supports the participants' professional development.

5.3 Limitations

Some limitations of the present study exist. First, coding teacher's negative statements into ten categories exposed some challenges for future research to negotiate. Negative behaviors are not mutually exclusive, meaning a teachers' statement could be a question and attention to junk behavior statement or sarcastic criticism. Additionally, teachers 'tempered' their negatives statements with embedded positives. For example, "*I appreciate your enthusiasm, but let's hold that until it can be a whole group discussion.*" Researchers should consider how to best account for these complex negatives through improved definitions and future empirical investigation. Second, these data collection challenges likely influenced the lower-than-expected reliability and IOA results. Third, the effectiveness of the intervention package (Experiment 2) was explored via a brief case-study design. Results demonstrated an initial experimental effect – the implementation of a one-time training, daily performance feedback emails, and teachers' reflective goal-setting

helped Adam improve his interactions with students. The inclusion of additional teachers in a multiple baseline across participants design might have allowed for the identification of a broader experiment effect.

5.4 Directions for Future Research

Considering the aforementioned limitations, multiple directions for future research exist. A lack of observational modeling studies on PVI and NVI behaviors limits our current empirical understanding of how teachers use these behaviors within their classrooms. Replications of the observational study (Experiment 1) would provide important and timely data regarding teachers' instructional behavior. Given the large variability in teacher behaviors within this study and compared to existing literature, researchers should observe teachers of varying education levels and years of experience from diverse schools and student populations. This data could drive further understanding of teacher coercives and aide in the development of tiered models for their impact on student behavior. Larger-scale observational studies would also allow for the use of higher-level statistical models (e.g., hierarchical linear modeling) to account for the role of classroom/school in understanding teacher and student behaviors. Individual teachers have unique and divergent learning histories, affecting their use of coercives and other NVIs. Future studies should also focus on understanding the qualitative features and uses of coercives within educational environments.

5.5 Implications for Practice

A review of available literature and results from Experiments 1 and 2 also have implications for practice. Teachers and other education professionals should critically self-assess their own use of PVIs and NVIs with students. Considering the results from Experiment 1, teachers should consider their overall PNR, frequencies of individual behaviors, and statement characteristics (e.g., WPI). In light of the considerable time required for this detailed introspection, teachers could rely on other classroom staff or team members to collaboratively collect data and determine their own coercive crutches. On-going PF or other reflective processes would likely help teachers achieve and maintain resulting behavior change. Educators should also consider adopting a collection of vocal (i.e., reminders, redirections, etc.) and non-vocal (i.e., body position, proximity, gesture or other visual cue, etc.) responses to inappropriate behavior.

Appendix A Literature Review Descriptive Study Characteristics

Participant Demographics
Teacher – <i>age, gender, race/ethnicity, certification(s), teaching experience (in years)</i>
Unit of Analysis – <i>dyad, triad, small group, class-wide</i>
Specific Students – <i>grade(s), age(s), gender(s), race/ethnicity</i>
Class-Wide – <i>grade(s), ages, genders, races/ethnicities, number of students</i>
Setting – <i>region, urbanicity, SES, performance data, school statistics reported</i>
Experimental Methodology
Research Design
Independent Variable(s)
Dependent Variable(s) – <i>teacher behavior(s), student behavior(s)</i>
Data Collection
Methods – <i>frequency, time of day, observation length, observer training</i>
Teacher Behavior – <i>data collected, data reported</i>
Student Behavior – <i>data collected, data reported</i>
Accuracy, Reliability, & Believability Measures
Procedural Fidelity Measures
Intervention Results
Teacher Outcome(s)
Student Outcome(s)
Social Validity

Appendix Figure A1. Literature Review Descriptive Study Characteristics

Appendix B Study Materials

<p>University of Pittsburgh <i>Institutional Review Board</i></p>	<p>Human Research Protection Office 3500 Fifth Avenue, Suite 106 Pittsburgh, PA 15213 Tel (412) 383-1480 www.hrpo.pitt.edu</p>
<p>APPROVAL OF SUBMISSION (Exempt)</p>	
Date:	November 2, 2019
IRB:	STUDY19100231
PI:	Justin Coy
Title:	Adjusting Teachers Instructional Behavior Through a Low-Impact Intervention Package Including Performance Feedback and Reflective Goal-Setting
Funding:	None
<p>The Institutional Review Board reviewed and approved the above referenced study. The study may begin as outlined in the University of Pittsburgh approved application and documents.</p>	
<p>Approval Documentation</p>	
Review type:	Initial Study
Approval Date:	11/2/2019
Exempt Category:	(1) Educational settings
Approved Documents:	<ul style="list-style-type: none"> • Data Collection - Landscape.docx, Category: Data Collection; • Social Validity Measure.docx, Category: Data Collection; • STUDY19100231 - Exemption_Educational Strategies.docx, Category: IRB Protocol; • STUDY19100231 - Parent Notification Letter.pdf, Category: Recruitment Materials; • STUDY19100231 - Teacher Consent.pdf, Category: Recruitment Materials; • STUDY19100231 - Teacher Consent.pdf, Category: Consent Form; • STUDY19100231 - Teacher Recruitment Letter.pdf, Category: Recruitment Materials;
<p>As the Principal Investigator, you are responsible for the conduct of the research and to ensure accurate documentation, protocol compliance, reporting of possibly study-related adverse events and unanticipated problems involving risk to participants or others. The HRPO Reportable Events policy, Chapter 17, is available at http://www.hrpo.pitt.edu/.</p>	
<p>Clinical research being conducted in an UPMC facility cannot begin until fiscal approval is received from the UPMC Office of Sponsored Programs and Research Support (OSPARS).</p>	
<p>If you have any questions, please contact the University of Pittsburgh IRB Coordinator, Larry Ivanco.</p>	
<p>Please take a moment to complete our Satisfaction Survey as we appreciate your feedback.</p>	

Appendix Figure B2. University of Pittsburgh Institutional Review Board Study Approval



University of
Pittsburgh | School of
Education

Dear Ms./Mr. [TEACHER'S NAME],

You are being invited to participate in a research study to determine the effectiveness of performance feedback and reflective goal-setting on adjusting teachers' instructional behavior. The duration of the study will be approximately six weeks and will include an initial investment of no more than 30 minutes outside of your normal time working with students for training. Following that initial training, you will be asked to record your regular teaching each day. The researcher will review a small time-frame of the daily video and provide you with feedback on your performance. You are asked to take 10 minutes to review the emailed feedback and respond. Additionally, you will be asked to complete a brief anonymous social validity measure on your experiences with the training, performance feedback, and instructional changes.

There are no foreseeable risks associated with this project, and no identifiable information from you, other than your email address and the capturing of your likeness on video, will be collected. All of the collected data is confidential and will be kept secure. Each participant will receive a \$50 debit card following completion of the study. Your participation is completely voluntary, and you may withdraw from the study at any time.

Thank you very much for your consideration in participating in this research study. This study is being conducted by Justin Coy, a doctoral student at the University of Pittsburgh. He can be reached at (703) 963-8425 or at JNC42@pitt.edu should you have any questions.

Sincerely,

A handwritten signature in black ink that reads "Justin Coy".

Justin N. Coy, M.Ed., BCBA
Graduate Student Researcher
University of Pittsburgh School of Education
Department of Instruction and Learning
5153 Wesley W. Posvar Hall
230 South Bouquet Street
Pittsburgh, PA 15260
[\(703\) 963-8425](tel:(703)963-8425)
JNC42@pitt.edu

Appendix Figure B3. Teacher Recruitment Letter



University of
Pittsburgh

School of
Education

CONSENT TO ACT AS A SUBJECT IN A RESEARCH STUDY

- TITLE:** Adjusting Teachers Instructional Behavior Through a Low-Impact Intervention Package Including Performance Feedback and Reflective Goal-Setting
- PRINCIPAL INVESTIGATOR:** Justin N. Coy, M.Ed., BCBA
Graduate Student Researcher
University of Pittsburgh School of Education
Department of Instruction and Learning
5153 Wesley W. Posvar Hall
230 South Bouquet Street
Pittsburgh, PA 15260
(703) 963-8425
JNC42@pitt.edu
- FACULTY MENTOR:** Douglas E. Kostewicz, Ph.D., BCBA-D
Assistant Professor of Special Education
University of Pittsburgh
5148 Wesley W. Posvar Hall
230 S. Bouquet St.
Pittsburgh, PA 15260
(412) 648-7018
dekost@pitt.edu

The purpose of this study is to evaluate the effectiveness of performance feedback and reflective goal-setting on teachers' instructional behavior (ratio of positive and negative interactions with students). Additionally, this research study evaluates the effects of this instructional ratio on the frequency of inappropriate student behaviors.

If you agree to participate, you will be observed teaching in your classroom for at least one 30-minute observations. Following this data collection, you will participate in a training consisting of instruction, rehearsal, practice, and feedback. The training should take no longer than 30 minutes. You will then practice the skills in the classroom and receive feedback on your performance.

For several weeks following the training you will receive feedback from the researcher by email. You will also be asked to respond to the email with specific things you feel like you did well that day and what changes you think will help you more successful the next day. While you are participating in the study, you will be asked to turn on a video camera

and audio-recorder prior to an identified challenging class/subject. Each day, the researcher will visit your classroom, download the data, and delete it from the camera/recorder. The researcher will select a random 15-minute section of video for analysis.

The video camera will be positioned in the back of the classroom to reduce distraction. No individual, private, or identifiable information will be collected about the students. It is expected over the period of the study you will participate in approximately one and a half hours of training, practice, and feedback. It is also expected that reading your performance feedback email and responding to the researcher should only take about 10 minutes each day.

There is little risk involved in this study. No invasive procedures or medications are included. **The major potential risk is a breach of confidentiality**, but we will do everything possible to protect your privacy. Another potential risk associated with your participation is the **nervousness** some people experience when they know they are being observed. This is not unusual, and if you like, we will discuss your feelings and concerns, and work to make the situation as stress-free as possible. There is no guaranteed benefit from this study; however, your management of the classroom may improve along with your interactions with students.

There are **no costs** to you for participating in this study, and you will get to keep the training materials for future reference. You will also receive a \$50 honorarium for completing the study. You are not obligated to participate in this study. If you choose not to participate or withdraw at any time from the study, there will be no penalty.

All records pertaining to your involvement in this study are kept **strictly confidential** and any data that includes your identity will be stored in locked files, and will be retained by us for a minimum of seven years. Your identity will not be revealed in any description or publications of this research. All video- and audio-recordings will be stored on a password-protected computer in password-protected files for a minimum of seven years, as stated above. Results will not be shared with your administrators at your school or district. It is possible that authorized representatives from the University of Pittsburgh Research Conduct and Compliance Office (including the University of Pittsburgh IRB) may review your data for the purpose of monitoring the conduct of this study. In unusual cases, your research records may be released in response to an order from a court of law. Also, if the investigators learn that you or someone with whom you are involved is in serious danger or potential harm, they will need to inform the appropriate agencies, as required by Pennsylvania law.

Your participation in this study is completely voluntary. You may refuse to take part in it, or you may stop participating at any time, even after signing this form. Your decision will not affect your relationship with the University of Pittsburgh, your employment, or standing with the school or district.

If you have questions about this research study, you may contact the investigators listed at the beginning of this consent form. If you have questions about your rights as a research subject, please contact the Human Subjects Protection Advocate at the University of Pittsburgh IRB Office, 1-866-212-2668.

SUBJECT'S CERTIFICATION

- I have read the consent form for this study and any questions I had, including explanation of all terminology, have been answered to my satisfaction. A copy of this consent form will be provided to me.
- I understand that I am encouraged to ask questions about any aspect of this research study during the course of this study, and that those questions will be answered by the researchers listed on the first page of this form.
- I understand that my participation in this study is voluntary and that I am free to refuse to participate or to withdraw my consent and discontinue my participation in this study at any time without affecting my future relationship with this institution.
- I agree to participate in this study.

Subject's Signature

Date

Subject's Printed Name

CERTIFICATION OF INFORMED CONSENT

I certify that I have explained the nature and purpose of this research study to the above-named individual, and I have discussed the potential benefits and possible risks of study participation. Any questions the individual has about this study have been answered, and we will always be available to address future questions as they arise.

Printed Name of Person Obtaining Consent

Role in Research Study

Signature of Person Obtaining Consent

Date

Appendix Figure B6. Full Teacher Consent (page 3)



University of
Pittsburgh

School of
Education

Dear Parent,

Hello! My name is Justin Coy, I am a doctoral student at the University of Pittsburgh School of Education and I am conducting a research study in [TEACHER'S NAME] classroom. I am interested in studying how to help teachers use effective, evidence-based teaching strategies in their classrooms. I am also interested in seeing how classroom behavior changes as a result of changing these teaching behaviors. The goal of my research is to identify how best to support current and future teachers, giving them the skills and training necessary to be successful.

I appreciate [TEACHER'S NAME] and your child's participation in this study. Confidentiality and security are an utmost concern when I conduct research. It is important to me that you fully understand that your child's identity will be kept as confidential as possible. A small video camera will be placed in the back of the classroom for the purposes of data collection. I will personally review a small segment of the video (about 15 minutes), and then the video file will be deleted from the camera. During data collection, I will summarize the total number of student behaviors *across the entire class*; this means that I will report a class-wide frequency, and not individual student behavior. The only risk is that your child's image may inadvertently be recorded during the video recording of the classroom. However, all files and data collection documents will be kept in a locked files and all recordings will be stored on a password-protected computer in password-protected files.

My research focus is on observing [TEACHER'S NAME] instructional methods and delivery. I will not record your child's name at any time during this research study. Your child will not be asked to do anything outside of his/her normal classroom activities and there are no foreseeable risks associated with this project. Your child's participation will not affect his/her grade. The only risk is that your child's image may inadvertently be captured during the video recording of the classroom (see confidentiality information above). If you decide that you do not want your child to participate, please contact me using the information below.

If you have any questions or concerns about the study, please feel free to contact me:

Justin Coy
(703) 963-8425
JNC42@pitt.edu

If you have questions about your rights as a research subject, please contact the Human Subjects Protection Advocate at the University of Pittsburgh IRB Office, 1-866-212-2668.

Appendix Figure B7. Parent Notification Letter

Data Collection Form – Performance Feedback for Teachers

Teacher: _____ Date: _____ Time: _____ Coding: L 1st R Condition: B IV M

Time	PTIs			NTIs										Student Behavior		
	GP	BSP	OTHER	CRIT	ARG	LOG	QUE	SAR	PLE	THR	FOR	ATJ	OTHER	BX 1	BX 2	BX 3
0																
1																
2																
3																
4																
5																
6																
7																
8																
9																
10																
11																
12																
13																
14																
15																

Video length _____
 Instruction starts at _____
 Start coding at minute _____
 Start coding at _____
 _____ students

Notes:

Appendix Figure B8. Data Collection Form

<u>Procedural Fidelity – Trainings</u>	
Teacher: _____	Date: _____
<u>Checklist</u>	
<input type="checkbox"/> All audio-recorded slides <input type="checkbox"/> Practice/discussion, feedback <input type="checkbox"/> Review of training <input type="checkbox"/> Present & review baseline P:N data <input type="checkbox"/> Invite questions, answer as necessary <input type="checkbox"/> Provide summative handout	
Notes:	
Total training time: _____	

Appendix Figure B9. Procedural Fidelity Checklist – Trainings

<u>Procedural Fidelity – PF Emails to Teachers</u>	
Teacher: _____	Date: _____ BL IV M
<u>Checklist</u>	
<input type="checkbox"/> Send PF email <ul style="list-style-type: none"> <input type="checkbox"/> Numbers for the day <input type="checkbox"/> At least 1 specific positive instance <input type="checkbox"/> At least 1 specific missed opportunity <input type="checkbox"/> Attach copy of graph <input type="checkbox"/> Discussion of graph (trend, improvement, decrease, ...) <input type="checkbox"/> Request reflective goal-setting	
Score/Notes:	
_____ /7	

Appendix Figure B10. Procedural Fidelity Checklist – Emails

*Researcher: start the stopwatch

Positive and Negative Interactions with Students

1

Background

Teacher-student interactions play an important role in classroom environments, relationships, experiences, and academic achievement

These minute-by-minute interactions influence student classroom experiences

Teachers using more positive interactions establish supportive classroom environments which promote student academic and behavioral improvements (Jenkins et al., 2015; Markels & Taylor, 2016; Noyce et al., 2018)

2

The Ratio – 4:1 PNR

'Best practice' recommendations vary: 4 positives for every negative, six praises per 15-min, 1 BSP every 2 min., etc. (Cook et al., 2017; Myers, Simmons, & Suggs, 2012; Sutherland, Wehby, & Cooperand, 2008)

No empirical work compares PNRs

However, PNR improvements have been shown to decrease student disruptive behavior and increase academic engagement (Cook et al., 2017)

3

Teacher Interactions

'POSITIVES'	'NEGATIVES'
Praise and related verbal statements intended to increase future behavior	Criticism, reprimands, and related verbal statements intended to decrease future behavior
<p>"I like how hard everyone is working."</p> <p>"Thank you Jody for volunteering and working the problem on the board."</p> <p>"I like that!"</p> <p>"That's a really good argument."</p> <p>"I can tell you took the time to check your answers and got them right!"</p>	<p>"Hey, I asked you to stop doing that already."</p> <p>"I'd call on you if you were raising your hand."</p> <p>"I'm about to start switching seats if you can't pay attention."</p> <p>"Are you following along?"</p> <p>"I'm talking, you're not."</p>

4

Attention/positives = \$\$\$

YOU'RE IN CONTROL OF YOUR ATTENTION

5

Strategies to Achieve 4:1

Goal: enrich classroom and student relationships by being mindful of your instructional moves and vocal statements

- Maximize positive praise statements
 - Within 3-5 seconds
 - Sincere/authentic: tone of voice, body language, facial expressions
 - Behavior-specific: explicitly mention appropriate/desired behavior
 - Vary the positives for meeting expectations
 - Individual students, groups, class as a whole
 - Social interaction, verbal praise, gestures/signals, etc.
 - Phrasing/word choice

6

Strategies to Achieve 4:1

Goal: enrich classroom and student relationships by being mindful of your instructional moves and vocal statements

- Pivot from junk behavior – "ignoring with a purpose"
 - Focus: distracting or disruptive behaviors (not dangerous to self, property, or others)
 - Ignore the disruptive behavior
 - Attend to another student/activity
 - Provide positives, praise, attention, etc.
 - Provide positive when behavior stops
 - Behavior you want to see, precursor behavior, etc.

7

Strategies to Achieve 4:1

Goal: enrich classroom and student relationships by being mindful of your instructional moves and vocal statements

- Minimize negatives/corrective statements
 - Avoid: arguing, questioning, sarcasm, threats, etc.
 - Predictable side-effects: avoid, escape, imitation, getting-even, ...
 - Address behavior, state expectations, and move on
 - Deliver corrective statements in steady tone of voice; direct and focused on the behavior of concern
 - Look for subsequent opportunities to provide extra positives

8

Appendix Figure B11. PowerPoint Training

Proficiency Check

Here are some scenarios

Identify and demonstrate a strategy we discussed in response

We will discuss & role play each scenario for about 15-20 seconds

9

Proficiency Check

There's a group of students chatting at a back table. You can tell none of them are paying attention as other students are reading aloud.

1. Maximize positive praise statement
2. Pivot
3. Minimize negative statements

10

Proficiency Check

You've asked for volunteers to come up to the board and answer questions. One student is calling out, waving his arms, etc. to get your attention.

1. Maximize positive praise statement
2. Pivot
3. Minimize negative statements

11

Proficiency Check

The students are sitting in small table groups discussing a section of a book they just read. Two students are yelling across the room to one another.

1. Maximize positive praise statement
2. Pivot
3. Minimize negative statements

12

Proficiency Check

Students are independently writing at their desks. A student yells out, "My pencil broke!" They begin to ask students around them for a pencil, begin to walk around to other students, etc.

1. Maximize positive praise statement
2. Pivot
3. Minimize negative statements

13

Your Behavior To-Date

You and the researcher will now discuss your baseline performance

Data is presented on **Standard Celeration Charts** – a technical, logarithmically-scaled chart

- Shows behavior change over time

14

Review

The collection of teachers' positive and negative interactions with students influences classroom environment and relationships

Strategies:

- Maximize positive praise statements
- Pivot from junk behavior
- Minimize negative statements

Remember, behavior change is hard! It takes time to change our behavior and see changes in students' – stay strong!

15

What Intervention Looks Like

PARTICIPANT	RESEARCHER
Record and upload files each day (ASAP)	View/code
Review performance feedback email, data displays	Compose and send performance feedback emails
Reply with reflective goal statements	Will reply to your emailed response

16

Appendix Figure B12. PowerPoint Training, con't

[Date]

Hi [Participant name],

[Introductory note, including description of coded instructional segment (whole group, etc.)]

Performance Feedback:

As you know, we're looking at your use of positive and negative interactions during instruction. Today, you provided [#] **positives** and [#] **negatives** during the coded observation. This is [up, down] from the last class.

You can see your progress to-date on the attached graph. [Performance comment(s); motivational comment]

Specific Moments:

- [Describe moment of successful skill implementation]
- [Describe moment of unsuccessful skill implementation, missed opportunity, etc.]

Reflective Goal-Setting:

Please take a few minutes to reflect on your instruction during [segment coded], your daily performance numbers, progress to date, and specific moments mentioned above. Then, please select a reflective statement and 'work on goal' for next class.

- Today, I felt like I
 - *Recognized opportunities to provide positives*
 - *Provided less negatives overall*
 - *Built in opportunities for to use positives*
 - *Ignored disruptive behavior; pivoted from junk behavior*
 - *Monitored and maintained good P:N ratio*
 - *Other:*
- Tomorrow, I'd like to work on
 - *Recognizing opportunities to provide positives*
 - *Using visual/audio cues as self-prompt*
 - *Ignoring disruptive behaviors; pivoting from junk behavior*
 - *Focusing on positive student behaviors*
 - *Monitoring and maintaining good P:N ratio*
 - *Intentionally focus on providing more positives*
 - *Other:*

As always, please make sure to **respond with your reflective statements tonight**. I will reply later tonight/tomorrow morning to acknowledge receipt and provide feedback. Please review my feedback and your 'work on' strategy tomorrow morning as a personal reminder. Don't forget to wear the microphone and turn on the camera 😊.

Feel free to email/call/text with any questions or concerns!

- Justin

Appendix Figure B13. Performance Feedback Email Template

Please complete this social validity survey about your experience during the study. Please keep the following definitions in mind when completing the measure:

- Intervention: increasing positive teacher interactions & reducing negative teacher interactions
- Procedures: training, performance feedback, and reflective goal-setting

	Strongly Disagree		Neutral			Strongly Agree
1. This was an acceptable intervention for the children's problem behavior.	1	2	3	4	5	6
2. Most teachers would find this intervention appropriate for challenging behaviors.	1	2	3	4	5	6
3. The intervention proved effective in changing children's problem behavior.	1	2	3	4	5	6
4. I would suggest the use of this intervention to other teachers.	1	2	3	4	5	6
5. The children's behavior problem was severe enough to warrant the use of this intervention.	1	2	3	4	5	6
6. Most teachers would find this intervention suitable for the behavior problem described.	1	2	3	4	5	6
7. I would be willing to use this intervention in other classroom settings.	1	2	3	4	5	6
8. The intervention did not result in negative side effects for the child.	1	2	3	4	5	6
9. The intervention would be an appropriate intervention for a variety of children/classrooms.	1	2	3	4	5	6
10. The intervention is consistent with those I have used in classroom settings.	1	2	3	4	5	6
11. The intervention was a fair way to address the children's problem behavior.	1	2	3	4	5	6
12. The intervention is reasonable for the behavior problem described.	1	2	3	4	5	6
13. I liked the procedures used in the intervention.	1	2	3	4	5	6
14. This intervention was a good way to handle the children's challenging behaviors.	1	2	3	4	5	6
15. Overall, the intervention was beneficial for the children.	1	2	3	4	5	6
16. Overall, the intervention was beneficial for myself as a teacher.	1	2	3	4	5	6
17. The intervention would produce a lasting improvement in my behavior.	1	2	3	4	5	6
18. Soon after using the intervention, I noticed a positive change in the children's problem behavior.	1	2	3	4	5	6
19. I will continue to use this intervention in my future teaching.	1	2	3	4	5	6
20. The procedures were easy and other teachers should also respond well to them.	1	2	3	4	5	6

Adapted from Behavior Intervention Rating Scale (BIRS) – Von Brock, M., & Elliott, S. (1987). Influence of treatment effectiveness information on the acceptability of classroom interventions. Journal of School Psychology, 25, 131-144.

Appendix Figure B14. Social Validity Measure (close-ended questions)

Please answer the following questions - feel free to type answer, extra paper, etc. as necessary.

This study included: (a) trainings on PTIs and NTIs, (b) performance feedback emails, and (c) reflective goal-setting about instructional strength and 'work on' strategy for the next day. Do you think these components were effective at changing your instructional behaviors?

Would you recommend this type of intervention/procedures with other teachers? Why or why not?

What components of the intervention/procedures did you like the most? The least? What aspects would you recommend changing before using this with other teachers?

Are you satisfied with the outcomes of the study? How has this study affected your teaching style/work, if any?

Is there anything else you'd like the researcher or research team to know about?

**Thank you again for participating in this research study – please reach out
Justin Coy (JNC42@pitt.edu) with questions or comments at any time!**

Appendix Figure B15. Social Validity Measure (open-ended questions)

Bibliography

- Apter, B., Arnold, C., & Swinson, J. (2010). A mass observation study of student and teacher behaviour in British primary classrooms. *Educational Psychology in Practice*, 26(2), 151-171. <https://doi.org/10.1080/02667361003768518>.
- *Allday, R. A., Hinkson-Lee, K., Hudson, T., Neilsen-Gatti, S., Kleinke, A., & Russel, C. S. (2012). Training general educators to increase behavior-specific praise: Effects on students with EBD. *Behavioral Disorders*, 37(2), 87-98. <https://doi.org/10.1177/019874291774087>.
- Barton, E. E., Kinder, K., Casey, A.M., & Artman, K. M. (2011). Finding your feedback fit: Strategies for designing and delivering performance feedback systems. *Young Exceptional Children*, 14(1), 29-46. <https://doi.org/10.1177/1096250610395459>.
- Black, J., Noltemeyer, A. L., Davis, D. R., & Schwartz, T. (2016). Pre-service teachers' responses to student behavior in a mixed-reality environment. *SAGE Open*, 6(1), 1-10. <https://doi.org/10.1177/2158244016633494>.
- Caldarella, P., Larsen, R. A., Williams, L., Downs, K. R., Wills, H. P., & Wehby, J. H. (2020). Effects of teachers' praise-to-reprimand ratios on elementary students' on-task behavior. *Educational Psychology*, 40, 1-17. <https://doi.org/10.1080/01443410.2020.1711872>.
- Cavanaugh, B. (2013). Performance feedback and teachers' use of praise and opportunities to respond: A review of the literature. *Education and Treatment of Children*, 36(1), 111-137. <https://doi.org/10.1353/etc.2013.0001>.
- Clark-Carter, D. (2005). Geometric mean. In B. Everitt and D. Howell (Eds.) *Encyclopedia of Statistics in Behavioral Science* (pp. 744-745). Chichester, West Sussex: John Wiley & Sons.
- Conroy, M., Sutherland, K., 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. Retrieved September 28, 2020, from <http://www.jstor.org/stable/24011740>.
- Cook, B. G., Buysse, V., Klinger, J., Landrum, T. J., McWilliam, R. A., Tankersley, M., & Test, D. W. (2015). CEC's standards for classifying the evidence base of practices in special education. *Remedial and Special Education*, 36(4), 220-234. <https://doi.org/10.1177/0741932514557271>.
- Cook, C. R., Grady, E. A., Long, A. C., Renshaw, T., Coddling, R. S., Fiat, A., & Larson, M. (2017). Evaluating the impact of increasing general education teachers' ratio of positive-to-negative interactions on students' classroom behavior. *Journal of Positive Behavior Interventions*, 19(2), 67-77. <https://doi.org/10.1177/1098300716679137>.

- Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis (3rd ed)*. Upper Saddle River, NJ: Pearson.
- *Cossairt, A., Hall, R. V., & Hopkins, B. L. (1973). The effects of experimenter's instructions, feedback, and praise on teacher praise and student attending behavior. *Journal of Applied Behavior Analysis*, 6(1), 89-100. <https://doi.org/10.1901/jaba.1973.6-89>.
- Cornelius, K. E. & Nagro, S. A. (2014). Evaluating the evidence base of performance feedback in preservice special education teacher training. *Teacher Education & Special Education*, 37(2), 133-146. <https://doi.org/10.1177/0888406414521837>.
- Datchuk, S. M. & Kubina, R. M. (2011). Communicating experimental findings in single case design research: How to use celeration values and celeration multiplier to measure direction, magnitude, and change of slope. *Journal of Precision Teaching and Celeration*, 27, 3-17.
- Elliott, S. N., & Von Brock Treuting, M. (1991). The behavior intervention rating scale: Development and validation of a pretreatment acceptability and effectiveness measure. *Journal of School Psychology*, 29(1), 43-51. [https://doi.org/10.1016/0022-4405\(91\)90014-i](https://doi.org/10.1016/0022-4405(91)90014-i).
- Fallon, L. M., Collier-Meek, M. A., Maggin, D. M., Sanetti, L. M., & Johnson, A. H. (2015). Is performance feedback for educators an evidence-based practice? A systematic review and evaluation based on single-case research. *Exceptional Children*, 81(2), 227-246. <https://doi.org/10.1177/0014402914551738>.
- Floress, M. T., Jenkins, L. N., Reinke, W. M., & McKown, L. (2018). General education teachers' natural rates of praise: A preliminary investigation. *Behavioral Disorders*, 43(4), 411-422. <https://doi.org/10.1177/0198742917709472>.
- Gelfuso, A. & Dennis, D. V. (2014). Getting reflection off the page: The challenges of developing support structures for pre-service teacher reflection. *Teaching and Teacher Education*, 38, 1-11. <https://doi.org/10.1016/j.tate.2013.10.012>.
- Gün, B. (2011). Quality self-reflection through reflection training. *ELT Journal*, 65(2), 126-135. <https://doi.org/10.1093/elt/ccq040>.
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>.
- *Horton, G. O. (1975). Generalization of teacher behavior as a function of subject matter specific discrimination training. *Journal of Applied Behavior Analysis*, 8(3), 311-319. <https://doi.org/10.1901/jaba.1975.8-311>.
- Hughes, J. N. (2011). The longitudinal effects of teacher and student perceptions of teacher-student relationship qualities on academic achievement. *The Elementary School Journal*, 112(1), 38-60. <https://doi.org/10.1086/660686>.

- Jenkins, L. N., Floress, M. T., & Reinke, W. (2015). Rates and types of teacher praise: A review and future directions. *Psychology in the Schools, 52*(5), 463-476. <https://doi.org/10.1002/pits.21835>.
- Johnston, J. M. & Pennypacker, H. S. (2009). *Strategies and tactics of behavioral research* (3rd ed.). New York, NY: Routledge.
- Kalis, T. M., Vannest, K. J., & Parker, R. (2007). Praise counts: Using self-monitoring to increase effective teaching practices. *Preventing School Failure, 51*(3), 20-27. <https://doi.org/10.3200/psfl.51.3.20-27>.
- Kazdin, A. E. (2012). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press.
- Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston, MA: Pearson Education, Inc.
- 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. <https://doi.org/10.1177/0741932512452>.
- Kubina, R. M. (2018, June 5). *What is variability and why do you as a behavior analyst care?* Retrieved from <http://bsci21.org>
- Kubina, R. M. (2019). *The Precision Teaching Implementation Manual*. Pittsburgh, PA: Greatness Achieved.
- Kubina, R. M., Kostewicz, D. E., Brennan, K. M., & King, S. A. (2017). A critical review of line graphs in behavioral analytic journals. *Educational Psychology Review, 29*(3), 583-598. <https://doi.org/10.1007/s10648-015-9339-x>.
- Latham, G. (1997). *Behind the schoolhouse door: Eight skills every teacher should have*. Logan, UT: Mountain Plain Regional Resource Center, Utah State University.
- Latham, G. (1998). *Keys to classroom management*. North Logan, UT: P & T Ink.
- Lindsley, O. R. (2005). Standard celeration chart system. In M. Hersen, J. Rosqvist, A. G. Gross, R. S. Drabman, G. Sugai, & R. Horner (Eds.), *Encyclopedia of behavior modification and cognitive behavior therapy* (pp. 1546-1548). Thousand Oaks, CA: Sage Publications, Inc.
- MacSuga-Gage, A. S., Simonsen, B., & Briere, D. E. (2012). Effective teaching practices that promote a positive classroom environment. *Beyond Behavior, 22*(1), 14-22. <https://doi.org/10.1177/107429561202200104>.
- Mainhard, M. T., Brekelmans, M., & Wubbels, T. (2011). Coercive and supportive teach behaviour: Within- and across-lesson associated with the classroom social climate. *Learning and Instruction, 21*, 3(345-354). <https://doi.org/10.1016/j.learninstruc.2010.03.003>.

- Markelz, A. M. & Taylor, J. C. (2016). Effects of teacher praise on attending behaviors and academic achievement of students with emotional and behavioral disorders. *The Journal of Special Education Apprenticeship*, 5(1). Retrieved from <http://josea.info/index.php?page=vol5no1>.
- Martens, B. K., Witt, J. C., Elliott, S. N., & Darveaux, D. X. (1985). Teacher judgments concerning the acceptability of school-based interventions. *Professional Psychology: Research and Practice*, 16(2), 191–198. <https://doi.org/10.1037/0735-7028.16.2.191>.
- McCormic, M. P., O'Connor, E. E., Capella, E., & McClowry, S. G. (2013). Teacher-child relationships and academic achievement: A multilevel propensity score model approach. *Journal of School Psychology*, 51(5), 611-624. <https://doi.org/10.1016/j.jsp.2013.05.001>.
- Moore, T. C., Maggin, D. M., Thompson, K. M., Gordon, J. R., Daniels, S., & Lang, L. E. (2019). Evidence review for teacher praise to improve students' classroom behavior. *Journal of Positive Behavior Interventions*, 21(1), 3-18. <https://doi.org/10.1177/1098300718766657>.
- *Mrachko, A. A., Kostewicz, D. E., & Martin, W. P. (2017). Increasing positive and decreasing negative teacher responses to student behavior through training and reinforcement. *Behavior Analysis: Research and Practice*, 17(3), 250-265. <https://doi.org/10.1037/bar0000082>.
- Myers, D. M., Simonsen, B., & Sugai, G. (2011). Increasing teachers' use of praise with a response-to-intervention approach. *Education and Treatment of Children*, 34, 35–59. <https://doi.org/10.1353/etc.2011.0004>.
- *O'Handley, R. D., Dufrene, B. A., & Whipple, H. (2018). Tactile prompting and weekly performance feedback for increasing teachers' behavior-specific praise. *Journal of Behavioral Education*, 27(3), 324-342. <https://doi.org/10.1007/s10864-017-9283-6>.
- Pennypacker, H. S., Gutierrez, A., & Lindsley, O. R. (2003). *Handbook of the standard celeration chart*. Concord, MA: Cambridge Center for Behavioral Studies.
- Rathel, J. M., Drasgow, E., Brown, W. H., & Marshall, K. J. (2014). Increasing induction-level teachers' positive-to-negative communication ratio and use of behavior-specific praise through e-mailed performance feedback and its effect on students' task engagement. *Journal of Positive Behavior Interventions*, 16(4), 219-233. <https://doi.org/10.1177/1098300713492856>.
- Reddy, L. A., Fabiano, G. A., Dudek, C. M., & Hsu, L. (2013). Instructional and behavior management practices implemented by elementary general education teachers. *Journal of School Psychology*, 51(6), 683-700. <https://doi.org/10.1016/j.jsp.2013.10.001>.
- Reinke, W. M., Herman, K. C., & Stormont, M. (2013). Classroom-level positive behavior supports in schools implementing SW-PBIS: Identifying areas for enhancement. *Journal of Positive Behavioral Interventions*, 15(1), 39-50. <https://doi.org/10.1177/1098300712459079>.

- *Reinke, W. M., Lewis-Palmer, T., & Merrell, K. (2008). The classroom check-up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review, 37*(3), 315-332.
- Reinke, W. M., Stormont, M., Herman, K. C., Puri, R., & Goel, N. (2011). Supporting children's mental health in schools: Teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly, 26*(1), 1-13. <https://doi.org/10.1037/a0022714>.
- Royer, D. J., Lane, K. L., Dunlap, K. D., & Ennis, R. P. (2018). A systematic review of teacher-delivered behavior-specific praise on k-12 student performance. *Remedial and Special Education, 40*(2), 112-128. <https://doi.org/10.1177/0741932517751054>.
- Sabey, C. V., Charlton, C., & Charlton, S. R. (2019). The “magic” positive-to-negative interaction ratio: Benefits, applications, cautions, and recommendations. *Journal of Emotional and Behavioral Disorders, 27*(3), 154-164. <https://doi.org/10.1177/1063426618763106>.
- Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd edition). London: SAGE Publications Ltd.
- Sarno Owens, J., Holdaway, A. S., Smith, J., Evans, S. W., Himawan, L. K., Coles, E. K., Giron-Herrera, E., Mixon, C. S., Egan, T. E., & Dawson, A. E. (2018). Rates of common classroom behavior management strategies and their associated with challenging student behavior in elementary school. *Journal of Emotional and Behavioral Disorders, 26*(3), 156-169. <https://doi.org/10.1177/1063426617712501>.
- Schles, R. A. & Robertson, R. E. (2019). The role of performance feedback and implementation of evidence-based practices for preservice special education teachers and student outcomes: A review of the literature. *Teacher Education and Special Education, 42*(1), 36-48. <https://doi.org/10.1177/0888406417736571>.
- Sidman, M. (1989). *Coercion and its fallout*. Boston, MA: Authors Cooperative.
- 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. <https://doi.org/10.1353/etc.0.0007>.
- *Simonsen, B., Freeman, J., Dooley, K., Maddock, E., Kern, L., & Myers, D. (2017). Effects of targeted professional development on teachers' specific praise rates. *Journal of Positive Behavior Interventions, 19*(1), 37-47. <https://doi.org/10.1177/1098300716637192>.
- Simonsen, B., Myers, D., & DeLuca, C. (2010). Teaching teachers to use prompts, opportunities to respond, and specific praise. *Teacher Education and Special Education, 33*(4), 300-318. <https://doi.org/10.1177/0888406409359905>.
- Snodgrass, M. R., Chung, M. Y., Meadan, H., & Halle, J. W. (2018). Social validity in single-case research: A systematic literature review of prevalence and application. *Research in Developmental Disabilities, 74*, 160-173. <https://doi.org/10.1016/j.ridd.2018.01.007>.

- State, T. M., Kern, L., Starosta, K. M., & Mukherjee, A. D. (2011). Elementary pre-service teacher preparation in the area of social, emotional, and behavioral problems. *School Mental Health, 3*, 13-23. <https://doi.org/10.1007/s12310-010-9044-3>.
- Sutherland, K. S. & Singh, N. N. (2004). Learned helplessness and students with emotional or behavioral disorders: Deprivation in the classroom. *Behavioral Disorders, 29*(2), 169-181. <https://doi.org/10.1177/019874290402900208>.
- Sutherland, K.S. & Wehby J. H. (2001). Exploring the relationship between increased opportunities to respond to academic and behavioral outcomes of students with emotional and behavioral disorders: A review. *Remedial and Special Education, 22*(2), 113–121. <https://doi.org/10.1177/074193250102200205>.
- Sutherland, K. S., Wehby, J. H., & Copeland, S. R. (2000). Effect of varying rates of behavior-specific praise on the on-task behavior of students with EBD. *Journal of Emotional and Behavioral Disorders, 8*(1), 2-8. <https://doi.org/10.1177/106342660000800101>.
- *Thompson, M. T., Marchant, M., Anderson, D., Prater, M. A., & Gibb, G. (2012). Effects of tiered training on general educators' use of specific praise. *Education and Treatment of Children, 35*(4), 521-546. <https://doi.org/10.1353/etc.2012.0032>.
- Von Brock, M. B. & Elliott, S. (1987). Influence of treatment effectiveness information on the acceptability of classroom interventions. *Journal of School Psychology, 25*(2), 131-144. [https://doi.org/10.1016/0022-4405\(87\)90022-7](https://doi.org/10.1016/0022-4405(87)90022-7).
- Walter, H. J., Gouze, K., & Lim, K. G. (2006). Teachers' beliefs about the mental health needs in inner city elementary schools. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*(1), 61-68. <https://doi.org/10.1097/01.chi.0000187243.17824.6c>.
- Watts, M. & Lawson, M. (2009). Using a meta-analysis activity to make critical reflection explicit in teacher education. *Teaching and Teacher Education, 25*, 609-616. <https://doi.org/10.1016/j.tate.2008.11.019>.
- White, M. A. (1975). Natural rates of teacher approval and disapproval in the classroom. *Journal of Applied Behavior Analysis, 8*(4), 367-372. <https://doi.org/10.1901/jaba.1975.8-367>.
- Whitman, B. Y., & Jackson, K. (2006). Tools for psychological and behavioral management. In M. G. Butler, P. D. K. Lee, & B. Y. Whitman (Eds.), *Management of Prader-Willi syndrome* (3rd ed., pp. 317-343). New York, NY: Springer.
- Yoon, K. S., Duncan, T., Lee, S., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from <http://ies.ed.gov/ncee/edlabs>.

*References marked with an asterisk indicate studies included in the literature review.