5-2020

Publication Bias in the Performance Diagnostic Checklist and its Variations

Ashley Dwire
St. Cloud State University, xt9196ao@go.minnstate.edu

Publication Bias in the Performance Diagnostic Checklist and its Variations

by

Ashley R. Dwire

A Thesis

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in Applied Behavior Analysis

May, 2020

Thesis Committee:
Benjamin N. Witts, Chairperson
Odessa Luna
Kimberly Schulze

Abstract

Publication bias—selectively publishing studies with positive outcomes—poses a problem to science as it can lead to inaccurate reports of intervention effects. Sham and Smith (2014) found that the published and unpublished pivotal response treatment literature differed, calling for more investigation into behavior-analytic research for publication bias. In this study, comparisons between the published and unpublished literature on the Performance Diagnostic Checklist, Performance Diagnostic Checklist-Human Services, and the Performance Diagnostic Checklist-Safety were conducted across three effect size measures: percentage of non-overlapping data, improvement rate difference, and percentage of data exceeding the median. Generally published literature outperformed the unpublished literature, providing further evidence of an overselling of results in the field.

Acknowledgements

I would like to express my deepest appreciation to my committee chair, Dr. Benjamin N. Witts, whose intense, blunt feedback has taught me attention to detail and precision in terminology. Without his knowledge of the field of applied behavior analysis, its principles, and his respect for research this paper would not have been possible. I would like to thank my committee members, Dr. Odessa Luna, Dr. Kimberly Schulze, and Dr. Benjamin Witts for your excellent feedback and contributions to my project. A special thanks to my supervisor Melyssa McDonough for helping me navigate company policies in my initial project, and for giving me time off as necessary to very quickly begin and complete a whole new research project when COVID-19 occurred.

I am extremely grateful to my grandparents, Bradley and Julie Nelson, who have never stopped believing in me and have supported me through each endeavor. Without your continued love and support it would have been much more convenient to sit on my goals rather than pursue them.

Finally, and most importantly, thank you to my husband, Nathan Dwire, and my five children—Preslee, Cohen, Payton, Blakely, and Macrae. Nathan, your time and commitment to late nights with the children, late nights listening to me read my paper out loud, and your ability to counsel me through hard times contributed significantly to the completion of this project. It would not have been possible without you. To my children, it is all for you.

Table of Contents

List of Tables

List of Figures

Figure Page

Chapter I: Introduction

Within the psychological sciences, there is a tendency for research to be published if it is positive (Chambers 2017; Ferguson, 2007; Hilgard, Sala, Boot, & Simons, 2019; Van Aert, Wicherts, & Van Assen, 2019). By positive, it is simply meant that the research produces results that suggest the intervention (or equivalent) was effective. In most psychological research, positive results are those that are statistically significant, typically with a *p*-value equal to or smaller than 0.05. However, while inferential statistics are useful for hypothesis-driven fields, behavior analysis focuses on inductive logic, and the prediction and control of behavior is our measure of success. The prediction and control of behavior is demonstrated through functional relations, in which some variable, when introduced, removed, or altered, demonstrates a predictable change in an organism's behavior (Schlinger & Normand, 2013). Thus, in behavior analysis, visual inspection of single-case behavior change data serves as the metric by which "positive" outcomes are derived.

Sham and Smith (2014) investigated the extent to which non-overlapping data metrics yielded different findings from the published and unpublished (e.g., dissertations) literature for pivotal response training. They found that differences existed, with the published literature producing less overlap between baseline and intervention data. Though including non-published studies did not influence the interpretation that pivotal response training was effective, their findings spurred concerns over the representational nature of the published literature. In their conclusions, Sham and Smith called for other technologies in behavior analysis to undergo similar investigations (Shadish, Hedges, & Pustejovsky, 2014; Shadish, Zelinsky, Vevea, & Kratochwill, 2016). One emerging area of intervention in behavior analysis that has yet to

receive a critical analysis of published and unpublished studies is with the Performance Diagnostic Checklist and its variants.

Chapter II: Literature Review

Performance Diagnostic Checklist

The Performance Diagnostic Checklist (PDC) is an assessment that is used to identify variables that might be influencing at-risk performance in an organizational setting (Austin, Weatherly, & Gravina, 2005). The PDC consists of four domains for employee performance: (a) training; (b) task clarification and prompting; (c) resources, materials, and processes; and (d) performance consequences, effort, and competition. The PDC has been used in restaurants (Amigo, Smith, & Ludwig, 2008; Austin et al., 2005; Rodriguez et al., 2006), university settings (Gravina, VanWagner, & Austin, 2008; Lebbon, Austin, Rost, & Stanley, 2011), department stores (Doll, Livesey, McHaffie, & Ludwig, 2007; Eikenhout, & Austin, 2005; Loughrey, Marshall, Bellizzi, & Wilder, 2013; Shier, Rae, & Austin, 2008), school districts (Berc, Doucette, DiGennaro Reed, Neidert, & Henley, 2014), and a coffee shop (Pampino, Heering, Wilder, Barton, & Burson, 2003) to determine the variables contributing to at risk behavior by the staff working in those settings.

In PDC research, a researcher interviews a supervisor, manager, or the employees with questions guided by the PDC. The researcher may also complete direct observation to answer the questions on the PDC. For example, Lebbon et al. (2011) used the PDC to determine the variables contributing to staff members' unsafe patient transfers (i.e., moving a client from one location to another). The researcher interviewed the employees and identified that the antecedents, equipment and processes, knowledge and skills, and consequences as reasons to why staff were engaging in at-risk behaviors. Based on the PDC outcomes, the researcher implemented an indicated intervention (i.e., task clarification to inform employees of expected

behaviors and social praise following correct greetings and closings) and found that staff's safe performance with patient transfers increased.

Performance Diagnostic Checklist-Human Services

Carr, Wilder, Majdalany, Mathisen, and Strain (2013) revised the PDC to assess the performance of employees in the human-service settings who are responsible for caring for others. The Performance Diagnostic Checklist-Human Services (PDC-HS) has been used in university-based settings working with individuals who have been diagnosed with autism (Carr et al., 2013; Ditzian, Wilder, King, & Tanz, 2015; Wilder, Lipschultz, & Gehrman, 2018), in a public school (Bowe & Sellers, 2017), and at a nonprofit school serving individuals diagnosed with autism (Merritt, DiGennaro Reed, & Martinez, 2019). Most studies reported researchers or BCBAs using the PDC-HS to identify variables contributing to staff performance and implementing interventions to increase staff integrity of protocols. However, Bowe and Sellers (2017) recruited teachers in the public-school setting to complete the PDC-HS to assess variables contributing to para-professionals inaccurate implementation of error-correction procedures. The researchers then used the PDC-HS completed by the teachers to implement an indicated intervention to increase accurate implementation of such procedures. In all of the studies, indicated interventions proved successful while non-indicated interventions did not.

Performance Diagnostic Checklist-Safety

Martinez-Onstott, Wilder, and Sigurdsson (2016) adapted the PDC to meet the needs of organizations where safe staff performance is instrumental. The adapted version, the Performance Diagnostic Checklist-Safety (PDC-Safety), focuses specifically on safe and at-risk behaviors of staff members in these organizations. To date, Martinez-Onstott et al. (2016) and

Cruz et al. (2019) have evaluated the PDC-Safety in a private university setting and center-based treatment facility for individuals who have been diagnosed with autism and other intellectual disabilities. In both studies, researchers and BCBAs completed the PDC-Safety and implemented the needed interventions to increase safe staff behavior. In general, indicated intervention based off of PDC-Safety results increased safe staff performance. When researchers used non-indicated staff treatment packages, minimal changes in safe staff performance were seen.

Purpose

Published investigations indicate the PDC, PDC-HS, and PDC-Safety are useful assessments to identify interventions to improve staff performance in a range of industries. However, if positive results are published more often than unsuccessful investigations with the PDC (or its variations), it may lead to exaggerated claims of its effectiveness. By comparing the published and unpublished outcomes of the PDC (and its adaptions), researchers and practitioners will have better understanding of its effectiveness. Thus, the current systematic review replicates procedures outlined in Sham and Smith (2014) described results of the PDC (and its adapted versions) among published and unpublished investigations to identify similarities or differences in (a) staff performance issues, (b) indicated interventions used to address staff performance, and (c) overall outcomes of the assessments effectiveness using percentage of nonoverlapping data (PND), improvement rate difference (IRD), and percentage of data exceeding a median (PEM).

Chapter III: Method

Search Strategy and Selection Criteria

Published and unpublished articles on the PDC, PDC-HS, and PDC-Safety were identified through a search of Google Scholar on the St. Cloud State University library database with the key term "*Performance Diagnostic Checklist."* Table 1 summarizes the focus of the review and Figure 1 displays the PRISMA guide flowchart (Moher, Liberati, Tetzlaff, & Altman, 2009).

Figure 1

Literature Search Flow Chart (Moher et al., 2009)

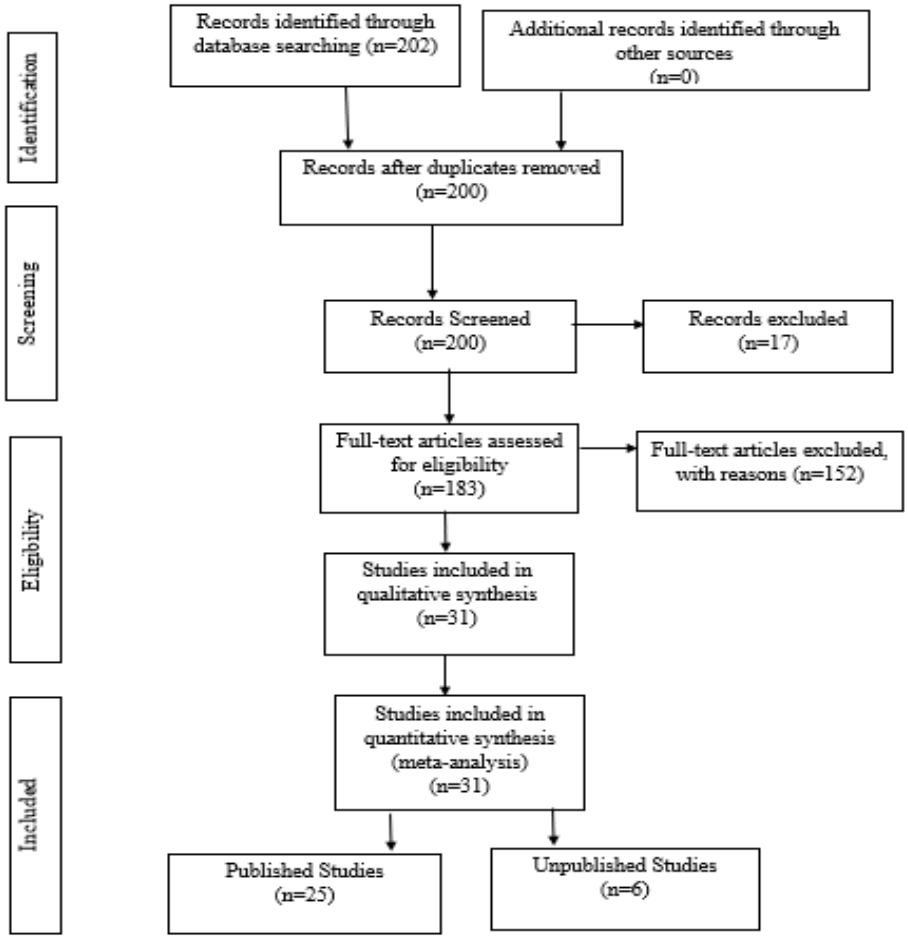


Table 1

Focus of the Review

Human participants between the ages of birth and 65 years
Included the use of the Performance Diagnostic Checklist, Performance Diagnostic Checklist-Human Services, or the Performance Diagnostic Checklist-Safety
Single-subject research designs (e.g., reversal, multiple baseline, and multi-element)
Indicated by the Performance Diagnostic Checklist, Performance Diagnostic Checklist-Human Services, or the Performance Diagnostic Checklist-Safety
No Treatment, Alternative Treatment, No Comparison Treatment
Increased Staff Performance

All studies identified through the literature search were screened for eligibility. Studies were included if they met the following criteria: (a) the study enrolled human participants; (b) the study implemented the PDC, PDC-HS, or the PDC-Safety; (c) functional control was demonstrated using a single-subject research design (e.g., reversal, multiple baseline); and (d) an intervention was implemented that was indicated by the PDC, PDC-HS, or the PDC-Safety. Studies were excluded if they referenced the PDC, PDC-HS, and/or the PDC-Safety but did not use them to identify an intervention.

Analyses of Effect Size

Both individual and group data were included. Effect sizes of group studies were not analyzed and graphed separately because there was only one unpublished article included for comparison. In this review, three different nonoverlap methods were used to calculate effect size: (a) percentage of nonoverlapping data (PND); (b) improvement rate difference (IRD); and (c) percentage exceeding the median (PEM). To calculate each of these effect sizes, lines were drawn with a pencil and a ruler following the steps as outlined by Rakep (2015). In some cases,

where it was harder to determine each datum point value across phases by hand, the graph was copied into a word document where straight lines could be drawn from the y-axis across the x-axis.

PND is the percentage of intervention phase data points that exceed the highest, or lowest for behavior reduction studies, data point in the baseline phase (Scruggs, Mastropieri, & Casto, 1987). All PND scores were calculated by hand, using the graphs in each article, with a ruler and a pencil. The highest baseline data point was identified and a line was drawn from that data point across the intervention phase. The data points in the intervention phase above the line were counted and this count was divided by the total number of data points in the intervention phase. The quotient was then multiplied by 100 and the resulting number equaled the PND score. For studies using multiple baseline designs the PND was calculated for each baseline-intervention contrast and then an average score was determined for the effect size of the full design. If a study used a withdrawal design, the first baseline data phase to determine a nonoverlap line, in line with Scruggs et al. (1987). If an alternating treatment design was implemented, the last intervention phase was used to calculate the effect size for the study (Rakep, 2015). The most notable disadvantage to this non-overlap method is that outlier baseline data points can misconstrue the overall effect size of the study.

IRD is the difference in improvement rate between Phase A and Phase B (Parker, Vannest, & Brown, 2009). When calculating IRD, the fewest overlapping data points are subtracted from the total data points across baseline and intervention phases and then divided by the total data points across baseline and intervention phases (Parker, Vannest, & Davis, 2011). The quotient was then multiplied by 100 and the resulting number equaled the IRD score. Like

PND, in a multiple baseline design, IRD is computed for each baseline-intervention contrast and then an average score of each contrast combined is computed for an IRD score for the full design (Rakep, 2015). One disadvantage of IRD is that it can be harder to calculate by hand for more complex designs.

Finally, PEM is calculated by extending a median line from the baseline to the intervention data and calculating the percentage of data points in the intervention phase that are above the median line (Olive & Franco, 2008). The total number of data points above the median line are divided by the total number of data points in the intervention phased. The quotient was then multiplied by 100 and the resulting number equaled the PEM score. For behavior reduction studies the percentage of data points below the median line were calculated. Similar to PND, in a multiple baseline design, PEM is computed for each baseline-intervention contrast and then an average score of each contrast is computed for a PEM full design score (Rakep, 2015). When calculating PEM for an alternating treatment design the overall effect size for the design should be calculated using the last intervention phase (Wolery, Busick, Reichow, & Barton, 2010).

Methodological Quality

Sham and Smith (2014) also rated the quality of the studies that they reviewed using a scale developed by Maggin, Briesch, and Chafouleas (2012) from the What Works Clearinghouse criteria for evaluating SSED studies (Kratochwill et al., 2010). The scale coded the presence or absence of six quality indicators: (a) the independent variable was systematically manipulated; (b) the dependent variable was repeatedly measured by more than one assessor; (c) interobserver agreement was assess for at least 20% of sessions for each eligible dependent variable; (d) interobserver agreement that is 80% or higher; (e) at least three attempts included to

demonstrate intervention effect; and (f) there were three or more data points in each phase. All studies in this systematic review (n = 31) were rated using these six quality indicators and were give a score from 0 (least rigorous methodology) to 6 (most rigorous methodology).

Chapter IV: Results

Tables 2, 3, 4, and 5 give a summary of characteristics of published and unpublished studies that implemented an intervention indicated by the PDC and adapted versions. Of the 25 published articles, 68% of them implemented the PDC, 24% implemented the PDC-HS, and 8% of them implemented the PDC-Safety. Of the six unpublished articles, 17% implemented the PDC, 50% implemented the PDC-HS, and 33% implemented the PDC-Safety. In all 25 published studies and six unpublished studies, researchers or BCBAs implement the checklist and intervention. Published studies were published in the *Journal of Applied Behavior Analysis* (20%), *Journal of Organizational Management* (40%), *Behavior Analysis in Practice* (28%), *Performance Improvement Quarterly* (4%), *Journal of Foodservice Business Research* (4%), and *Journal of Behavior Analysis in Health, Sports, Fitness, and Medicine* (4%). There were six or more participants in 44% of the published studies. Unpublished studies had three participants (33.3%), four participants (33.3), or six or more participants (33.4%). For both published (64%) and unpublished studies (66.6%), multiple domains were targeted with an intervention package as opposed to targeting one domain. Social validity was measured in 52% of published studies and 83% of the unpublished articles.

Published studies produced an average quality index of 5.24 (*SD* = 0.93, median = 5, range = 2-6) and unpublished studies produces an average quality index of 5.00 (*SD* = 1.26, median = 5.5, range = 3-6; see Tables 4 and 5 for specifics. Of the studies reviewed (n = 31), two of the published studies and one of the unpublished studies did not systematically manipulate the independent variable because they implemented quasi-experimental designs such as the nonconcurrent multiple baseline design. In two of the published studies, the dependent variable

was not measured repeatedly over time by multiple assessors and in one study there were six data points across phases. One published study did not establish adequate levels of interobserver agreement. Two unpublished studies and 11 published studies did not sufficiently demonstrate three attempts at an intervention effect because of the design implemented or the lack of baseline intervention contrasts. Three published studies and one unpublished study did not have at least three data points per phase.

Table 2

Published Articles Characteristics

Characteristic	*n*	%
Journal		
Journal of Applied Behavior Analysis	5	20
Journal of Organizational Behavior Management	10	40
Behavior Analysis in Practice	7	28
Performance Improvement Quarterly	1	4
Journal of Foodservice Business Research	1	4
Journal of Behavior Analysis in Health, Sports, Fitness, & Medicine	1	4
Participants		
1	0	0
2	1	4
3	5	20
4	6	24
5	2	8
6 or More	11	44
Year Published		
2000-2005	4	16
2006-2010	7	28
2011-2016	7	28
2017-2020	7	28
Setting		
Department Store	4	16
Restaurant	5	20
Clinic/Center Providing Supports to Individuals	5	20
School/Classroom	5	20
Grocery Store	2	8
Day Treatment Center/Residential Setting	2	8
Sport Practice Field	1	4
Library	0	0
University	1	4
Social Validity		
Yes	13	52
No	12	48
Follow-up		
Yes	4	16
No	21	84
Assessment		
Performance Diagnostic Checklist	17	68
Performance Diagnostic Checklist-Human Services	6	24
Performance Diagnostic Checklist-Safety	2	8
Administered PDC & Indicated Training		
Researcher	25	100
Participant Supervisor	0	0
Domain(s) Targeted		
Antecedents & Information	1	4
Equipment & Processes	0	0
Training	3	12
Consequences	5	20
More than One	16	64

Note. Domain names are listed for the Performance Diagnostic Checklist but the table incorporates information from all published studies (n=25). Information for adapted checklist (PDC-HS, PDC-Safety) was entered into the corresponding domain.

Table 3

Unpublished Article Characteristics

Characteristic	*n*	%
Participants		
1	0	0
2	0	0
3	2	33.3
4	2	33.3
5	0	0
6 or More	2	33.3
Year Published		
2000-2005	0	0
2006-2010	0	0
2011-2016	1	16.7
2017-2020	5	83.3
Setting		
Department Store	1	16.7
Restaurant	0	0
Clinic/Center Providing Supports to Individuals	2	33.3
School/Classroom	1	16.7
Grocery Store	0	0
Day Treatment Center/Residential Setting	0	0
Sport Practice Field	1	16.7
Library	1	16.7
University	0	0
Social Validity		
Yes	5	83.3
No	1	16.7
Follow-up		
Yes	2	33.3
No	4	66.7
Assessment		
Performance Diagnostic Checklist	1	16.7
Performance Diagnostic Checklist-Human Services	3	50
Performance Diagnostic Checklist-Safety	2	33.3
Administered PDC & Indicated Training		
Researcher	6	100
Participant Supervisor	0	0
Domain(s) Targeted		
Antecedents & Information	1	16.7
Equipment & Processes	0	0
Training	1	16.7
Consequences	0	0
More than One	4	66.6

Note. Domain names are listed for the Performance Diagnostic Checklist but the table incorporates information from all published studies (n=25). Information for adapted checklist (PDC-HS, PDC-Safety) was entered into the corresponding domain.

Table 4

Published Studies Characteristics Summary

Author	Participants	Intervention	Design	Outcome	PND[a]	IRD[b]	PEM[c]	Quality
Amigo, Smith, and Ludwig (2008)	Four Female Lunchtime Servers	Task Clarification Memo and Goal Setting Individual Verbal and Graphic Feedback Group Feedback	ABC	Intervention Effectiveness in Decreasing the Amount of Time it Took a Server to Correctly Bus a Table	50	83	100	4
Austin, Weatherly, and Gravina (2005)	7 Dishwashers and 11 Servers	Intervention Package Consisting of a Posted Checklist, Verbal Feedback from Management, and Posted Graphic Feedback	Multiple Baseline Design with a Limited Component Analysis Across Groups of Employees	Indicated Intervention Package Increased Task Completion for both Groups of Employees	100	100	100	5
Berc, Doucette, DiGennaro Reed, Neidert, and Henley (2014)	Seven Teachers who were Undergraduate Female Students	Prompts, and Feedback Task Clarification Video Based Training Feedback	Concurrent Multiple Baseline Design Across Shifts	Video Based Training Effective in Increasing % Accuracy Feedback Training Effective in Increasing % Accuracy	100	100	95	5
Bowe and Sellers (2018)	Four Female Paraprofessionals	Non-Indicated Intervention: Task Clarification and Prompting Indicated Intervention: Behavior Skills Training	Concurrent Multiple Baseline Design Across Participants	Non-Indicated Intervention Ineffective Indicated Intervention Resulted in Performance Improvement for all Paraprofessionals	100	100	100	6

Table 4 (*continued)*

Author	Participants	Intervention	Design	Outcome	PND[a]	IRD[b]	PEM[c]	Quality
Carr, Wilder, Majdalany, Mathisen, and Strain (2013)	15 Graduate Student Employees	Indicated Intervention Individual Training and Graphic Feedback Posted Prior to Each Shift (Package) Non-indicated Intervention: Task Clarification and Materials	Concurrent Multiple Baseline Design Across Treatment Rooms	Indicated Intervention Improved Performance Non-Indicated Intervention Did Not Improve Performance	100	100	100	6
Cruz, Wilder, Phillabaum, Thomas, Cusick, and Gravina (2019)	Three Behavior Therapists	Non-indicated: Access to a Metal Clip with Hand Sanitizer on Participants Belt Throughout Treatment Session Indicated Intervention: Email Reminder Evening Before Requiring Response	Concurrent Multiple Baseline Design Across Participants	Indicated Intervention Increased Staff Performance for One Participant. One Participant Needed an Added Booster Email. On Participant Needed Booster E-mail, Job Aid, Feedback, and a Printed Copy of Response to Feedback. Non Indicated Intervention Ineffective for All Three Participants	68	91	97	6
Dagen and Austin (2008)	Three Female Swimmers	Intervention: Graphic and Verbal Feedback, Self-Talk, and Self-Monitoring	ABAC	Graphic and Verbal Feedback Alone Did Not Produce Measurable Improvement Self-Monitoring, Self-Talk, Graphic Feedback, and Verbal Feedback Increased Swim Performance by as Much as .7 Seconds	93	96	96	5

Table 4 (*continued)*

Author	Participants	Intervention	Design	Outcome	PND[a]	IRD[b]	PEM[c]	Quality
DePaolo, Gravina, and Harvey (2019)	12 Female Players	Intervention Package: Prompts and Sprint Contingency	ABAB Reversal Design	Intervention Effective in Improving Names on Passes (Group) Individual Data Not Collected Due to Fast Pace of the Sport	100	100	100	4
Ditzian, Wilder, King, and Tanz (2015)	Four Female Staff Therapists	Indicated Intervention: Individual Verbal/Graphed Feedback by Supervisor Non-Indicated Intervention: Written Prompt Outside Door	Concurrent Multiple Baseline Design Across Participants	Indicated Intervention Improved Performance for all Therapists	98	99	98	6
Doll, Livesey, McHaffie, and Ludwig (2007)	7 Employees	Intervention Package Based on PIC/NIC Analysis: Task Clarification, Behavior Checklist, Consequences added (i.e. Graphic and Written Feedback)	ABC	Initial Intervention Produced Improvements Across All 5 Targeted Behaviors. Written Feedback Produced Further Improvement	100	100	100	6
Eikenhout and Austin (2005)	115 Employees	Intervention 1: Graphic Feedback Intervention 2: Weekly Goals, Weekly Feedback, and Social Reinforcement in the Form of Verbal Praise (Package)	ABAC and Multiple Baseline Design	Indicated Intervention Increased Customer Service-Related Behaviors	93	99.5	100	6

Table 4 (*continued)*

Author	Participants	Intervention	Design	Outcome	PND[a]	IRD[b]	PEM[c]	Quality
Fante, Shier, and Austin (2006)	16 Full-time Kitchen Employees	Task Clarification Phase and Self-Monitoring Phase in Which Managers Prompted Self-Monitoring	Within-Group Reversal Design	In the Task Clarification Phase There was Minimal Change in Performance In the Self-Monitoring Phased There was Some Improvement in the Mean Percentage of Appropriate Food Temperature Checks	14	96	100	5
Gravina, VanWagner, and Austin (2008)	Two Full-Time and Four Part-Time Workers	Package Intervention: Task Clarification, Equipment Manipulations, and Graphic Feedback	ABC Multiple Baseline Design Across Behaviors	Package Intervention Successful at Increasing the Completion of Preparation Tasks	93	97	93	5
Lebbon, Austin, Rost, and Stanley (2011)	3 Employees and Their Supervisor	Intervention Package: Employee Training (Task Analysis), Safe Lifting Procedures Training, Supervised Consumer Lifting Self-Report, and Feedback System	Reversal Design	Increased Staff Performance with Indicated Intervention Package Note: Based off of Condition Means	70	92	97	5
Loughrey, Marshall, Bellizzi, and Wilder (2013)	Two Female Part-time Sales Associates	Intervention Package: Video Modeling, Role Play, Visual Prompts, and Feedback	Nonconcurrent Multiple Baseline Design Across Participants	Package Intervention Increased Credit Card Promotion for Both Participants	100	100	100	2

Table 4 (*continued)*

Author	Participants	Intervention	Design	Outcome	PND[a]	IRD[b]	PEM[c]	Quality
Martinez-Onstott, Wilder, and Sigurdsson (2016)	Three Employees	Graphic Feedback Delivered by the Experimenter	Nonconcurrent Multiple Baseline Design Across Participants	Intervention Increased Safe Performance for All Participants	92	95	92	5
Miller, Carlson, and Sigurdsson (2014	Three Female Participants	Feedback Verbal Following Baseline and Then Written Following Days	Concurrent Multiple Baseline Design Across Participants	Treatment Integrity Improved Consistently for All Participants During Intervention	89	98	94	6
Merritt, and DiGennaro Reed (2019)	Four Female Employees	Intervention A: Reviewed Expectations, Individualized Problem-Solving Discussion about the Variables Influencing Tardiness Modified Intervention B: Corrective Feedback Discontinued; Token Economy Modified Intervention C for One Participant: Praise and Token for Arriving to Work by 8:20 am	Multiple Baseline Design Across Participants with an Embedded Withdrawal	One Participant Experienced Intervention A, B, and C. None of the Interventions were Effective in Producing Stable Responding Improvements for Three Other Participants did not Maintain in Intervention A Responding was variable for all Three Participants when Intervention B was Implemented	100	100	100	6
Pampino, Heering, Wilder, Barton, and Burson (2003)	Four Employees	Intervention: Task Clarification, Training on the Use of a 95 Checklist, and Public Posting of Lottery Tickets Earned by Each Employee	Multiple Baseline Design Across Task Groups	Intervention Effective in Increasing Staff Performance of Completing Closing Tasks	100	100	100	5

Table 4 (*continued)*

Author	Participants	Intervention	Design	Outcome	PND[a]	IRD[b]	PEM[c]	Quality
Rice, Austin, and Gravina (2009)	12 Workers	Intervention Package: Task Clarification to Inform Employees of Expected Behaviors and Social Praise Following Correct Greetings and Closings	Multiple Baseline Design Across Behaviors	Increased Staff Performance with Indicated Intervention Package	98	97	100	5
Rodriguez, Wilder, Therrien, Wine, Miranti, Daratany, Salume, Baranovsky, and Rodriguez (2006)	18 Employees	Intervention Package: Task Clarification, Self-Monitoring, Equipment Modification, Goal Setting, and Graphic Feedback	Multiple Baseline Design Across Stores	Intervention Package Increased Percentage in Which a Promotional Stamp was Offered	10	88	100	5
Sellers, Clay, Hoffmann, and Collins (2018)	Initially Five BCBAs and Two BCBA-Ds One BCBA Discontinued as He Did Not Work with Children who Engaged in Significant Problem Behaviors	Intervention Package: Behavior Skills Training, Breakout Sessions Focusing on Training of Designing, Directing, and Analyzing Results	Multiple Baseline Design Across Participants	Completion Increased for 5 of 6 Participants Following Implementation of Intervention Package	70	92	87	6

Table 4 *(continued)*

Author	Participants	Intervention	Design	Outcome	PND[a]	IRD[b]	PEM[c]	Quality
Shier, Rae, and Austin (2003)	25 Employees	Package Intervention: Job Aids, Task Clarification, and Graphed Performance Feedback	Multiple Baseline Design Across Departments	Treatment Package Appeared to Produce an Increase in Cleaning Behaviors Across Five Departments	75	95	99	6
Smith and Wilder (2018)	Four Employees Who had Been Diagnosed with a Disability One Manager	Price Training Intervention that Included Three Steps for the Supervisors to Deliver to Their Supervisees. (i.e. Inform, Model, and Deliver Performance-Based Feedback.	Concurrent Multiple Baseline Design Across Participants	Indicated Intervention Improved Performance	100	100	100	5

[a]Percentage of nonoverlapping data.
[b]Improvement rate difference.
[c]Percentage of data exceeding median.

Table 5

Unpublished Studies Characteristics Summary

Author	Participants	Intervention	Design	Outcomes	PND[a]	IRD[b]	PEM[c]	Quality
Bowe (2017)	Four Special Education Para-educators and their three supervising special education teachers.	Non-indicated Intervention: Task Clarification Indicated Intervention One: Training Indicated Intervention Two: Not Needed	Concurrent Multiple Baseline Design Across Subjects	Non-indicated: Ineffective Indicated: Effectively increased percentage of correctly completed steps.	100	100	100	6
Cruz (2018)	Four Behavior Therapists	Indicated: Prompting and Task Clarification Non-Indicated: Access to Hand Sanitizer	Concurrent Multiple Baseline Design Across Participants	Indicated Intervention was Effective; Two Participants Needed Consequence Based Intervention Added Non-Indicated Intervention Ineffective	54	82	84	6
DePaolo (2018)	11 Female Athletes	Negative Reinforcement Only Phase Negative Reinforcement Plus Signal Phase Signal Only	A-B-BC-C-A-BC Withdrawal Design	Intervention Effective in Improving Names on Passes (Group) Individual Data non Collected Due to Fast Pace of the Sport	100	100	100	5
Smith (2016)	Six Employees who have Been Diagnosed with a Disability	Behavior Skills Training Several Task Clarification and Prompting Interventions Probed for Participant in Dyad 3	Concurrent Multiple Baseline Design Across Participants	Effective for Two of the Participants Participant in Dyad 3 did not Show Improvement in Performance	67	67	67	3

Table 5 *(continued)*

Author	Participants	Intervention	Design	Outcomes	PND[a]	IRD[b]	PEM[c]	Quality
Villacorta (2011)	Three Parent-Child Dyads	Experimenter Defined and Thorough Explained Each Step of the Mand Training Protocol Visual Flow Chart Parents were told they could Refer to. Experimenter answered all questions	Nonconcurrent Multiple Baseline Design Across Parent-Child Dyads Pre- and Post-Intervention Probes Used	Increased Performance by	65	91	97	4

[a]Percentage of nonoverlapping data.
[b]Improvement rate difference.
[c]Percentage of data exceeding median

Figures 2 and 3 display the results of the PND calculations in the form of a box plot. As shown in Figures 2 and 3, PND was higher in the 25 published studies, M=93%, 95% CI (68%, 100%), than in the six unpublished articles, M=77%, 95% CI (54%, 100%).

Figure 2

Box Plot Depicting Effect Size Percentage Nonoverlapping Data for Group Studies Combined with Aggregate Calculations from Single-Subject Studies

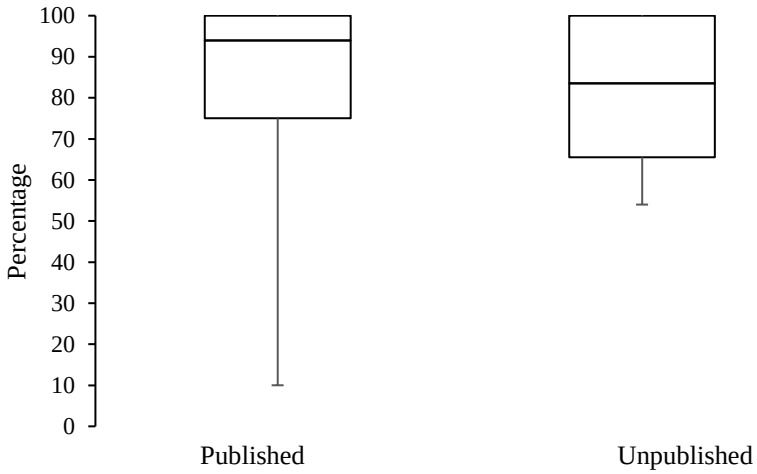
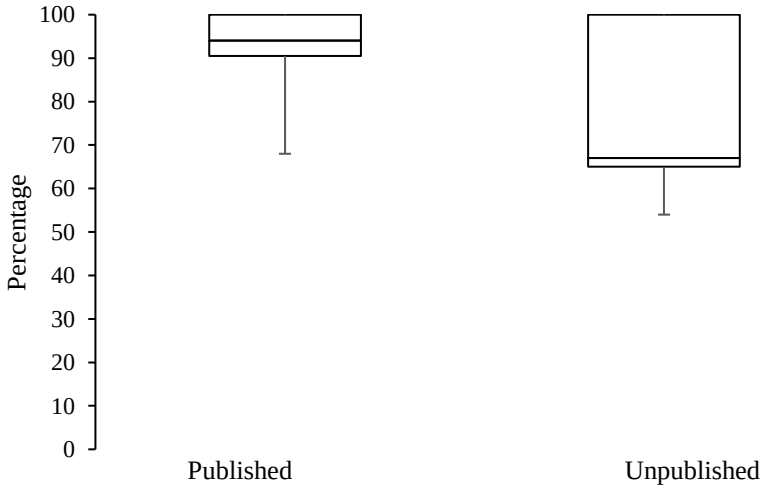


Figure 3

Box Plot Depicting Effect Size Percentage Nonoverlapping Data for Single-Subject Studies



Figures 4, 5, and 6 display the results of the PND calculations in the form of a boxplot for the PDC and each of its variants individually. One variant's calculations; the PDC-HS has a higher median effect for unpublished studies. It should be noted that this variant of the checklist only included data from three unpublished studies. In Figures 5 and 6, box plots for the unpublished studies show a dash with no boxes as there is no differences in the quartiles. The length of the boxplot shows the range of PND values from lowest (bottom) to highest (top); the entire box (shown with split median line) shows the density of the PND scores for each type of article, published and unpublished. The density of the PND scores for the published studies show less variation than the unpublished studies that show a larger range of scores.

Figure 4

Box Plot Depicting Effect Size Percentage Nonoverlapping Data for Single-Subject Studies on the Performance Diagnostic Checklist

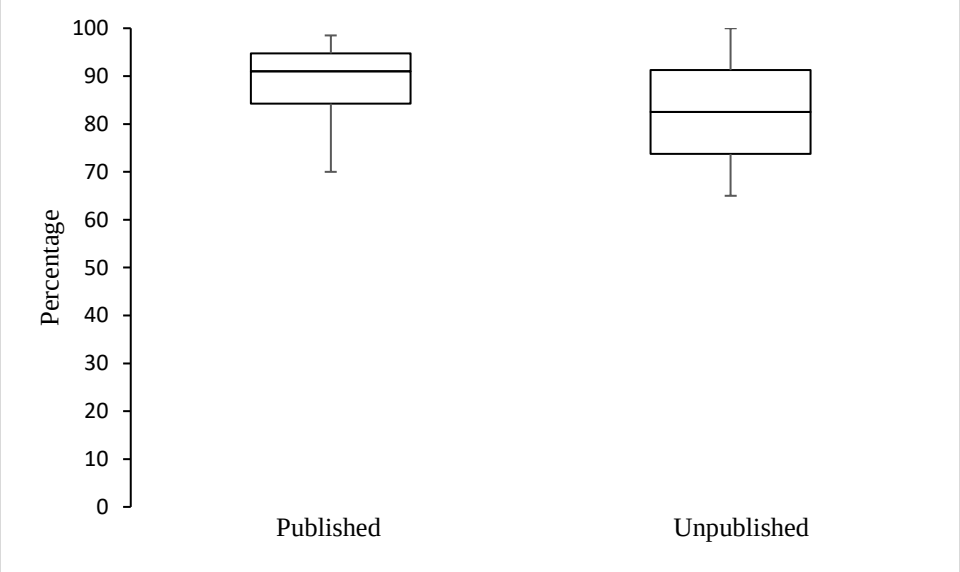


Figure 5

Box Plot Depicting Effect Size Percentage Nonoverlapping Data for Single-Subject Studies on the Performance Diagnostic Checklist-Human Services

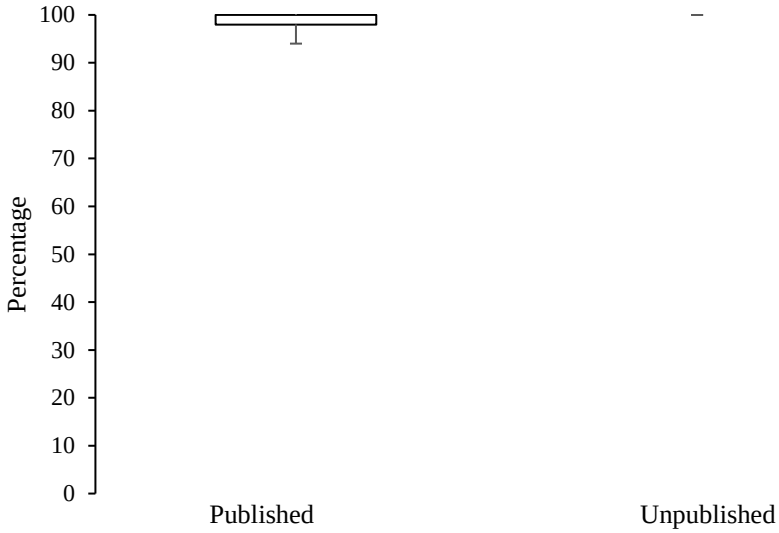
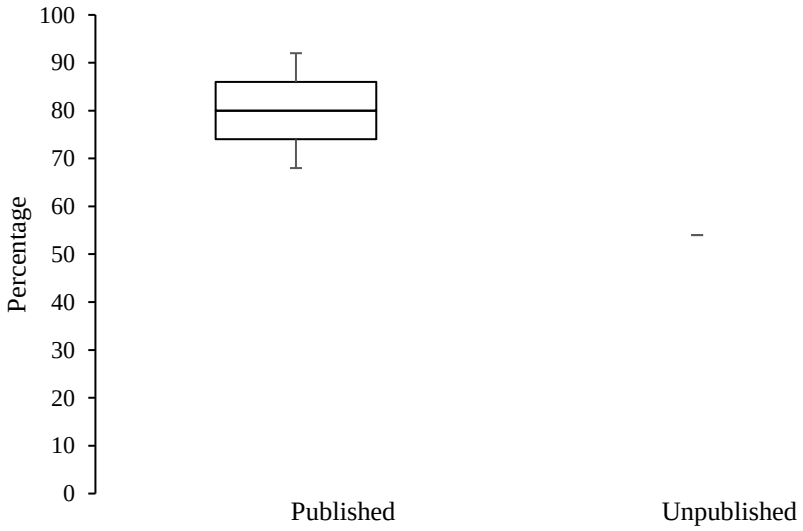


Figure 6

Box Plot Depicting Effect Size Percentage Nonoverlapping Data for Single-Subject Studies on the Performance Diagnostic Checklist-Safety



Figures 7 and 8 display the results of the IRD calculations in the form of a box plot. As show in each figure the published studies have a higher median in effect size when calculated using this nonoverlap method and the density of scores for the published articles show less variation than the unpublished articles that show a larger range of scores.

Figure 7

Box Plot Depicting Effect Size Improvement Rate Difference for Group Studies Combined with Aggregate Calculations from Single-Subject Studies

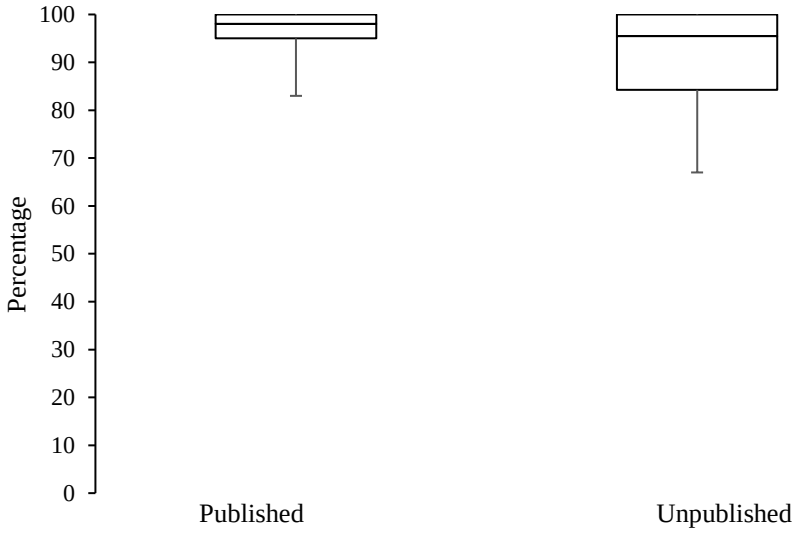
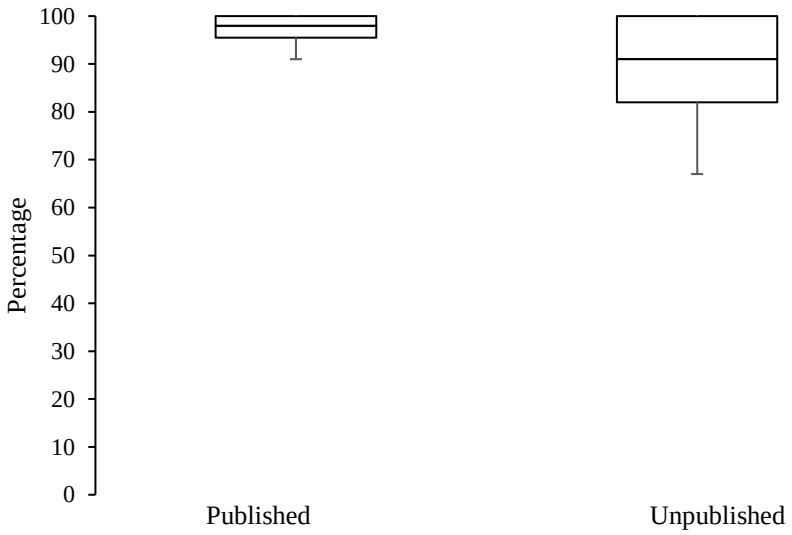


Figure 8

Box Plot Depicting Effect Size Improvement Rate Difference for Single-Subject Studies



Figures 9 and 10 display the results of the PEM calculations in the form of a box plot. As shown in each figure the published studies have a higher median in effect size when calculated using this nonoverlap method and the density of scores for the published articles show less variation than the unpublished articles that show a larger range of scores.

Figure 9

Box Plot Depicting Effect Size Percentage of Data Exceeding Median for Group Studies Combined with Aggregate Calculations from Single-Subject Studies

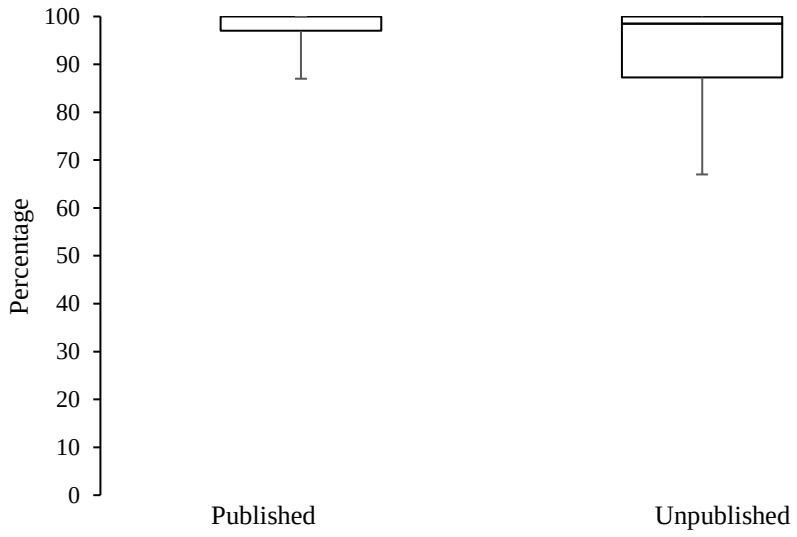
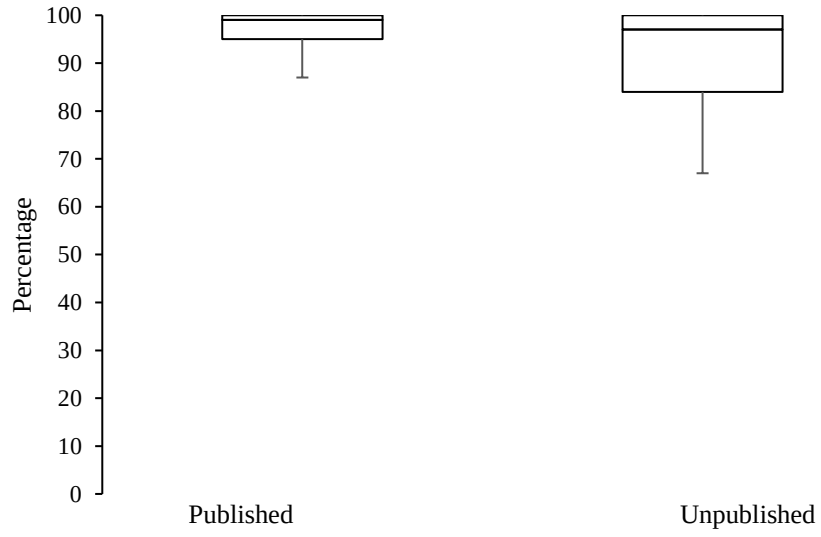


Figure 10

Box Plot Depicting Effect Sizes Percentage of Data Exceeding Median for Single-Subject Studies



Chapter V: Discussion

This systematic review shows that median effect sizes in published studies were higher than unpublished articles when implementing interventions indicated by the PDC, PDC-HS, and PDC-Safety. With the published studies having the higher median PND, IRD, and PEM scores one may conclude that publication bias is occurring where studies with positive outcomes are more likely to be published than studies with negative outcomes. Publication bias in the literature could have implications for the PDC, PDC-HS, and PDC-Safety. If studies with positive outcomes are more likely to be published an opportunity to discuss any feedback on the checklists and negative outcomes from interventions indicated by the checklist are lost. It should be noted that even though the median effect sizes in published studies are higher, unpublished studies for the PDC and one of its variations have effect sizes indicating effective and very effective intervention (Rakep, 2015). One unpublished study has an effect size indicating an intervention with a questionable effect. If we combine the effect size results for both the published and unpublished studies the PDC is still an effective assessment tool.

These results are consistent with Sham and Smith (2014) in that the effect size calculations show a disparity between published and unpublished studies. Like Sham and Smith, the published and unpublished studies had the same methodological quality rating. The effect size calculations showed that the data effect size was lower for unpublished studies suggesting that treatments with less effective results are often not published. This review extends the literature as it further assesses publication bias in the applied behavior analysis (ABA) literature. In addition, effect size was also calculated using IRD and PEM nonoverlap methods. The use of additional nonoverlap methods to calculate effect size adds to the estimates on the magnitude of intervention effect.

There are limitations that should be considered in this literature review. There were more published studies (n = 25) analyzed in this review than unpublished studies (n = 6). There were two additional unpublished articles that were not included in this review as the full article was not accessible online and the authors were not able to be contacted. This systematic literature review lacked an independent observer and there was no interrater reliability collected for the effect size calculations. A final limitation that should be considered is that although the effect sizes were calculated there are conditions that were not manipulated (e.g., one unpublished study did not report any data on one of its participants but reports that multiple interventions were attempted with no success).

This systematic literature review suggests that publication bias does exist in the literature in studies implementing interventions indicated by the PDC or variations of the checklist to increase staff performance. Future research should continue to assess for publication bias in published literature on other interventions or assessment tools. Specific to the PDC (and its adaptions), future research should assess whether or not the checklist can be used, by people who are not a BCBA or who have minimal training in applied behavior analysis, to determine an indicated intervention. In each of the studies, published and unpublished, researchers or BCBAs implement the checklist and indicated intervention. Future research should conduct a component analysis as they are important in identifying the active components of treatment packages that produce behavior change (Baer, Wolf, & Risley, 1968).

References

Amigo, S., Smith, A., & Ludwig, T. (2008). Using task clarification, goal setting, and feedback to decrease table busing times in a franchise pizza restaurant. *Journal of Organizational Behavior Management, 28*(3), 176-187. Retrieved from https://doi.org/10.1080/01608060802251106

Austin, J., Weatherly, N. L., & Gravina, N. E. (2005). Using task clarification, graphic feedback, and verbal feedback to increase closing-task completion in a privately owned restaurant. *Journal of Applied Behavior Analysis, 38*(1), 117-120. doi:10.1901/jaba.2005.159-03

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*, 91-97. doi:10.1901/jaba.1968.1-91

Berc, H., Doucette, J. L., DiGennaro Reed, F. D., Neidert, P. L., & Henley, A. J. (2014). Improving the quality of parent-teacher interactions in an early childhood classroom. *Behavior Analysis in Practice, 7*(2), 47-50. Retrieved from https://www.ncbi.nlm.nih.go v/p mc/articles /PMC4711759/

Bowe, M., & Sellers, T. P. (2017). Evaluating the Performance Diagnostic Checklist-Human Services to assess incorrect error-correction procedures by preschool paraprofessionals. *Journal of Applied Behavior Analysis, 51*(1), 166-176. doi:10.1002/jaba.428

Bowe, M. (2017). *The use of the Performance Diagnostic Checklist for human services to increase paraeducators' effective implementation of error-correction procedures during discrete trial training.* (Unpublished master's thesis). Utah State University, Utah, United States. Retrieved from https://digitalcommons.usu.edu/etd/5781/

Carr, J. E., Wilder, D. A., Majdalany, L., Mathisen, D., & Strain, L. A. (2013). An assessment-based solution to a human-service employee performance problem an initial evaluation of the Performance Diagnostic Checklist-Human Services. *Behavior Analysis in Practice, 6*(1), 16-32. Retrieved from Retrievedhttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3680147/

Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice.* Princeton, NJ: Princeton University Press.

Cruz, N., J. (2018). *Are we safe yet? Re-evaluation of the Performance Diagnostic Checklist-Safety (PDC-Safety).* (Unpublished master's thesis). University of Melbourne, Melbourne, Florida. Retrieved from https://repository.lib.fit.edu/bitstream/handle/11141/2774/CRUZ-THESIS-2018.pdf?sequence=1&isAllowed=y

Cruz, N. J., Wilder, D. A., Phillabaum, C., Thomas, R., Cusick, M., & Gravina, N. (2019). Further evaluation of the Performance Diagnostic Checklist-Safety (PDC-Safety). *Journal of Organizational Behavior Management, 39,* 266-279. doi:10.1080/01608061.2019.1666777

Dagen, J., & Austin, J. (2008). Using graphic feedback with self-talk and self-monitoring to improve high-school swim performance. *Journal of Behavior Analysis in Health, Sports, Fitness, & Medicine, 1*(2), 96-116. Retrieved from http://eitaneldar.com/wp-content/uploads/2011/08/Eldar-Hirschmann-Eldarn-hod-jbahsfm.pdf

DePaolo, J. (2018). “Get on the line”: Improving pass-naming in collegiate women’s lacrosse using negative reinforcement and signaling. (Unpublished master’s thesis). University of Melbourne, Melbourne, Florida. Retrieved from https://repository.lib.fit.edu/bitstream/handle/11141/2612/DEPAOLO-THESIS-2018.pdf?sequence=1&isAllowed=y

DePaolo, J., Gravina, N., & Harvey, C. (2018). Using behavioral intervention to improve performance of a women’s college lacrosse team. *Behavior Analysis in Practice, 12*(2), 407-411. doi:10.1007/s40617-018-0272-6

Ditzian, K., Wilder, D. A., King, A., & Tanz, J. (2015). An evaluation of the Performance Diagnostic Checklist-Human Services to assess an employee performance problem in a center-based autism treatment facility. *Journal of Applied Behavior Analysis, 48*(1), 199-203. doi:10.1002/jaba.17

Doll, J., Livesey, J., McHaffie, E., & Ludwig, T. D. (2007). Keeping an uphill edge: managing cleaning behaviors at a ski shop. *Journal of Organization Behavior Management, 27*(3), 41-60. doi:10.1300/j075v27n03_04

Eikenhout, N., & Austin, J. (2005). Using goals, feedback, reinforcement, and a performance matrix to improve customer service in a large department store. *Journal of Organizational Behavior Management, 24*(3), 27-62. doi:10.1300/J075v24n03_02

Fante, R., Shier, L., & Austin, J. (2006). Utilizing task clarification and self-monitoring to increase food temperature checks among restaurant staff. *Journal of Foodservice Business Research, 9,* 67-88. Retrieved from https://doi.org/101300/j369v09n02_05

Ferguson, C. J. (2007). Evidence for publication bias in video game violence effects literature: a meta-analytic review. *Aggression and Violent Behavior, 12*(4), 470-482. Retrieved from https://doi.org/10.1016/j.avb.2007.01.001

Gravina, N., VanWagner, M., & Austin, J. (2008). Increasing physical therapy equipment preparation using task clarification, feedback, and environmental manipulations. *Journal of Organizational Behavior Management, 28*(2), 110-122. doi:10.1080/01608060802100931

Hilgard, J., Sala, G., Boot, W. R., & Simons, D. J. (2019). Overestimation of action-game training effects: Publication bias and salami slicing. *Collabra: Psychology, 5*(1), 1-10. Retrieved from https://doi.org/10.1525/collabra.231

Kratochwill, T. R., Hitchcock, J., Jorner, R. H., Levin, J. R., Odom, S. L., Rindsckopf, D. M., & Shadish, W. R. (2010). Single-case design technical documentation. Retrieved from http:/ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Lebbon, A., Austin, J., Rost, K., & Stanley, L. (2011). Improving safe consumer transfers in a day treatment stetting using training and feedback. *Behavior Analysis in Practice, 4*(2), 35-43. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3357098/

Loughrey, T. O., Marshall, G. K., Bellizzi, A., & Wilder, D. A. (2013). The use of video modeling, prompting, and feedback to increase credit card promotion in a retail setting. *Journal of Organizational Behavior Management, 33*(3), 200-208. Retrieved from https://doi.org/10.1080/01608061 .2013.815097

Maggin, D. M., Briesch, A. M., & Chafouleas, S. M. (2012). An application of the What Works Clearinghouse Standards for evaluating single-subject research: synthesis of the self-management literature base. *Remedial and Special Education,* 115. doi:10.1177/0741932511435176

Martinez-Onstott, B., Wilder, D., & Sigurdsson, S. (2016). Identifying the variables contributing to at-risk performance: Initial evaluation of the Performance Diagnostic Checklist-Safety (PDC-Safety). *Journal of Organizational Behavior Management, 36,* 80-93. doi:10.1080/01608061.2016.1152209

Merritt, T. A., DiGennaro Reed, F. D., & Martinez, C. E. (2019). Using the Performance Diagnostic Checklist-Human Services to identify an indicated intervention to decrease employee tardiness. *Journal of Applied Behavior Analysis, 52*(4), 1034-1048. doi:10.1002/jaba.643

Miller, M. V., Carlson, J., & Sigurdsson, S. (2014). Improving treatment integrity in a human service setting using lottery-based incentives. *Journal of Organizational Behavior Management, 34*(1), 29-38. Retrieved from https://doi.org/10/1080/01608061.2013.873381

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), e1000097. Retrieved from https://doi.org/10.1371/journal.pmed.1000097

Olive, M. L., & Franco, J. H. (2008). Effect size matters: And so does the calculation. *The Behavior Analyst Today, 9,* 5-10. Retrieved from http://dx.doi.org/10.1037/h0100642

Pampino, R. N., Heering, P.W., Wilder, D.A., Barton, C.G., & Burson, L.M. (2003). The use of the Performance Diagnostic Checklist to guide intervention selection in an independently owned coffee shop. *Journal of Organizational Behavior Management, 23,* 5-19. Retrieved from https://doi.org/10.1300/J075v23n02_02

Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children, 75,* 135-150. Retrieved from https://doi.org/10.1177/001440290907500201

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine overlap techniques. *Behavior Modification, 35*(4), 303-322. Retrieved from https://doi.org/10.11770/0145445511399147

Rakep, S. (2015). Effect sizes as a result interpretation aids in single-subject experimental research: description and application of four nonoverlap methods. *British Journal of Special Education, 42,* 11-33. doi:10.1111/1467-8578.12091

Rice, A., Austin, J., & Gravina, N. (2009). Increasing customer service behaviors using manager-delivered task clarification and social praise. *Journal of Applied Behavior Analysis, 42*(3), 665-9. doi:10.1901/jaba.2009.42-665

Rodriguez, M., Wilder, D. A., Therrien, K., Wine, B., Miranti, R. Daratany, K., . . . Rodruguez, M. A. (2006). Use of the Performance Diagnostic Checklist to select an intervention designed to increase the offering of promotional stamps at two sites of a restaurant franchise. *Journal of Organizational Behavior Management, 25*(3), 17-35. doi:10.1300/J075v25n03_02

Schlinger, H. D., & Normand, M. P. (2013). On the origin and functions of the term *Functional Analysis*. *Journal of Applied Behavior Analysis, 46,* 285-288. Retrieved from https://doi.org/10.1002/jaba.6

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research methodology and validation. *Remedial and Special Education, 8,* 24-33. Retrieved from https://doi.org/10.1177/074193258700800206

Sellers, T. P., Clay, C. J., Hoffmann, A. N., & Collings, S. D. (2018). Evaluations of a performance management intervention to increase use of trial-based functional analyses by clinicians in a residential setting for adults with intellectual disabilities. *Behavior Analysis in Practice, 12,* 412-417. doi:10.1007/s40617-018-00276-8

Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with standardized mean difference statistic: A primer and applications. *Journal of School Psychology, 52,* 123-147. Retrieved from https://doi.org/10/1016/j.jsp.2013.11.005

Shadish, W. R., Zelinsky, N. A., Vevea, J. L., & Kratochwill, T. R. (2016). A survey of publication practices of single-case design researchers when treatments have small or large effects. *Journal of Applied Behavior Analysis, 49*(3), 656-673. Retrieved from https://doi.org/10.1002/jaba.308

Sham, E., & Smith, T. (2014). Publication bias in studies of an applied behavior-analytic intervention: an initial analysis. *Journal of Applied Behavior Analysis, 47,* 663-678. doi:10.1002/jaba.146

Shier, L., Rae, C., & Austin, J. (2008). Using task clarification, checklist, and performance feedback to improve the appearance of a grocery store. *Performance Improvement Quarterly, 16*(2), 26-40. Retrieved from https://doi.org/10.1111/j.1937-8327.2003.tb00277.x

Smith, M. (2016). *The use of the Performance Diagnostic Checklist-Human Services to assess and improve the job performance of individuals with disabilities.* (Unpublished master's thesis). University of Melbourne, Melbourne, Florida. Retrieved from https://repository.lib.fit.edu/bi tstream/handle/11141/1136/SMITH-THESIS.pdf?sequence=1&isAllowed=y

Smith, M., & Wilder, D. A. (2018). The use of the Performance Diagnostic Checklist-Human Services to assess and improve the job performance of individuals with intellectual disabilities. *Behavior Analysis in Practice, 11*(2), 148-153. doi:10.1007/s40617-018-0213-4

Van Aert, R. C., Wicherts, J. M., Van Assen, M. A. (2019). Publication bias examined in meta-analyses from psychology and medicine: A meta-meta-analysis. *PLos One, 14*(4), 1-31. Retrieved from https://doi.org/10.1371/journal.pone.0215052

Villacorta, J. (2011). *Evaluation of a tool to identify the variables responsible for poor parent treatment implementation: the Performance Diagnostic Checklist-Parent.* (Unpublished Master's Thesis). University of Melbourne, Melbourne, Florida. Retrieved from https://repository.lib.fit.edu/bitstream/handle/11141/2308/VILLACORTA-THESIS-2017.pdf?sequence=1&isAllowed=y

Wilder, D. A., Lipschultz, J., & Gehrman, C. (2018). An evaluation of the Performance Diagnostic Checklist-Human Services (PDC-HS) across domains. *Behavior Analysis in Practice, 11*(2), 129-138. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5959821/

Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *The Journal of Special Education, 44,* 18-28. doi:10.1177/0022466908328009