META-ANALYSIS OF FAMILY-CENTERED POSITIVE BEHAVIOUR SUPPORT WITH FAMILIES OF CHILDREN WITH DEVELOPMENTAL DISABILITIES AND PROBLEM BEHAVIOUR

by

Xin Gao

B.Ed., Beijing Normal University, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Special Education)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

September 2020

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis entitled:

Meta-Analysis of Family-Centered Positive Behaviour Support with Families of Children with Developmental Disabilities and Problem Behaviour

submitted by Xin Gao in partial fulfillment of the requirements for

the degree of Master of Arts

in Special Education

Examining Committee:

Dr. Joseph Lucyshyn, Associate Professor, Department of Educational and Counselling Psychology, and Special Education, UBC
Supervisor

Dr. Edward Kroc, Assistant Professor, Department of Educational and Counselling Psychology, and Special Education, UBC
Supervisory Committee Member

Dr. Nicholas Gage, Associate Professor, School of Special Education, School Psychology, & Early Childhood Studies, University of Florida
Supervisory Committee Member

Abstract

Children with developmental disabilities often engage in problem behaviour. Problem behaviour has a pervasive and disruptive effect on children's development and family quality of life. Family-centered Positive Behaviour Support (FCPBS) integrates core features of a positive behaviour support approach (e.g., functional assessment, multicomponent behaviour support plan, contextual fit) and family centered service delivery (e.g., family as unit of attention, collaborative partnership, family empowerment) into a behavioural support service delivery model that aims to improve child behaviour and child and family quality of life. The primary purpose of this study was to conduct a meta-analysis to examine the overall effect of FCPBS across 30 included studies published between 1997 to 2019. The secondary purpose was to evaluate the rigor of the 30 included studies based on What Works Clearinghouse (WWC) design and evidence standards for single case research, and to determine if FCPBS meets criteria necessary to be categorized as an empirically supported treatment (EST). In addition, a moderator analysis was conducted to determine whether there were effect size differences across the moderating variables of age, gender, disability category and WWC rigor evaluation. Result showed that FCPBS is effective at reducing problem behaviour of children with developmental disabilities and increasing their appropriate behavior and/or positive engagement. Result also showed that 12 of the 30 included FCPBS studies (40%) met WWC design and evidence standards. The WWC 5-3-20 criteria for an EST indicated that the first two criterion were met (i.e., 12 studies across 6 research groups and 7 locations) but the third criterion was approached but not met, indicating that FCPBS is an emerging rather than established EST. Results are discussed in terms of relation to the literature, limitations and future directions.

Lay Summary

Family Centered Positive Behaviour Support (FCPBS) aims to improve the behaviour and quality of life of children with developmental disabilities and their families. To date, there has been no meta-analysis of FCPBS. The goals to the study were to: (a) conduct a meta-analysis to synthesize the overall effect of FCPBS; (b) assess the quality of FCPBS studies; (c) determine if FCPBS is an empirically supported treatment (EST); and (d) explore whether age, gender, disability and research study rigor have impact on effects of FCPBS. Results showed FCPBS is effective at decreasing problem behaviour and/or increasing positive behaviour of children with developmental disabilities. Results also showed that although FCPBS studies approached requirements of EST, additional studies are needed for FCPBS to become an EST. These findings lend support to the use of FCPBS as an effective practice and encourage researchers to continue conducting FCPBS studies to establish FCPBS as an EST.

Preface

I participated in the identification and design of the focus of my thesis research, which is a meta-analysis of single case research studies of family-centered positive behaviour support (FCPBS). I performed all parts of the research methods, including the analysis of the research data. As this thesis was a meta-analysis of published data, there was no need for an ethics application. My research supervisor, Dr. Joseph Lucyshyn, provided assistance in the write-up of each section of the thesis including the Introduction, Method, Results and Discussion. In each section, I wrote the first draft and Dr. Lucyshyn provided editorial assistance with a second and final draft. In addition, Dr. Edward Kroc and Dr. Nicholas Gage, committee members, contributed to decisions about the procedures that were employed in the meta-analysis and described in the Method section. Dr. Edward Kroc also contributed to minor editorial support for each section of the manuscript. Chloe Wang, a doctoral student in the Special Education Area in the ECPS Department, assisted with data extraction and data coding. Dr. Lucyshyn assisted with the selection of the FCPBS studies that were included in the meta-analysis, and with the evaluation of What Works Clearinghouse design and evidence standards for the single case research studies that were included in the meta-analysis.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANOVA	analysis of variance
ASD	autism spectrum disorder
BC-SMD	between-case standardized mean difference
CEC	Council for Exceptional Children
CI	confidence interval
CICO	Check-In, Check-Out
EST	empirically supported treatment
FBA	functional behaviour assessment
FCPBS	family-centered positive behaviour support
FSS	First Steps to Success
HLM	hierarchical linear modeling
IES	U.S. Institute of Educational Science
IOA	interobserver agreement
IRA	interrater agreement
LRR	log response ratio
PBIS	positive behaviour interventions and supports
PBS	positive behaviour support
PND	percentage of non-overlapping data
SWPBS	school-wide positive behaviour support
WWC	What Works Clearinghouse

Acknowledgements

To begin, I offer my enduring gratitude to my supervisor, Dr. Joseph Lucyshyn, who offered me much hands-on support in the development of the thesis. I have been deeply impressed by your diligence, patience and professionalism. In the past four years, you have helped me in my application to UBC, transition to master's study in Canada, course selection and professional development in addition to the thesis. Especially during the pandemic, you were always available for help when I encountered difficulties, which made me feel so safe and empowered. I am so happy to have you as my graduate advisor and research supervisor.

I would like to express my gratitude to committee members, Dr. Edward Kroc and Dr. Nicholas Gage, who provided me with assistance with the methodological issues of the study. I would like to bring my additional thanks to Dr. Kroc. Through your courses, I became familiar with meta-analysis and other important content in statistics, which determined my mind to conduct the study.

I also want to express my gratitude to Chloe Wang who helped me with data collection and the rehearsal of the proposal and final defense. Thank you for devoting your time to my thesis.

Special thanks are owed to my father, who has supported me throughout my years of education, both morally and financially. Without your support, I would not have been able to concentrate on my study.

Chapter 1: Introduction

Children with developmental disabilities, such as Autism Spectrum Disorder (ASD) or intellectual disability, often engage in problem behaviour (Poppes, Van der Putten, & Vlaskamp, 2010; Wang & Singer, 2016), including noncompliance, disruptive behaviour, physical aggression, self-injury, and property destruction (Meadan, Ayvazo, & Ostrosky, 2016). Problem behaviour typically begins in early childhood and continues into adolescence and even adulthood, creating significant challenges to families (Lucyshyn et al., 2015). Parents of children with developmental disabilities and problem behaviour experience high levels of parental stress, lower levels of psychological well-being, financial strain, and worry about their child's future (Carroll, 2013; Ha, Hong, Seltzer, & Greenberg, 2008; Muir & Strnadová, 2014). Child problem behaviour clearly has a pervasive and disruptive effect on family quality of life. Therefore, there is a great need for behavioural support services for families with children with developmental disabilities.

1.1 Positive Behaviour Support

Over the past 25 years, PBS has emerged as an acceptable and effective empirically supported approach for addressing the problem behaviour of children with developmental disabilities and problem behaviour in family, school and community contexts (Brown, Anderson, & De Pry, 2015; Dunlap, Sailor, Horner & Sugai, 2009). With a strong conceptual foundation in applied behaviour analysis (Baer, Wolf, & Risley, 1968), and an equally strong commitment to human values such as respect, dignity and humanity, practitioners of PBS seek to improve an individual's behaviour and quality of life through primarily preventive, teaching and positive reinforcements strategies (Lucyshyn, Dunlap, & Freeman, 2015). Kincaid and Dunlap (2015) recently offered a comprehensive definition of PBS, as follows:

> PBS is an approach to behaviour support that includes an ongoing process of research-based assessment, intervention, and data-based decision making focused on building social and other functional competencies, creating supportive contexts, and preventing the occurrence of problem behaviours. PBS relies on strategies that are respectful of a person's dignity and overall well-being and that are drawn primarily from behavioural, educational, and social sciences, although other evidence-based procedures may be incorporated. PBS may be applied within a multi-tiered framework at the level of the individual and at the level of larger systems (e.g., families, classrooms, schools, social service programs, and facilities (p. 3).

Core features of PBS, initially defined by Horner et al. (1990) and later extended by Carr et al. (2002) include: (a) a focus on quality of life; (b) collaborative partnership with key stakeholders, (c) functional assessment; (d) multicomponent behaviour support plans; (e) attention to contextual fit; (f) emphasis on preventive, teaching, and positive reinforcement; (g) eschewment of punishment strategies that cause loss of dignity, embarrassment or physical discomfort or pain; (h) evaluation of social validity; (i) the use of multiple research methodologies that assist in the development of knowledge about the delivery of PBS in family, school and community settings; and (j) the integration of other science-based disciplines and theoretical perspectives that contribute to the achievement of meaningful, durable and sustainable improvements in behaviour and quality of life. In addition to the core discipline of behaviour analysis, other science-based disciplines that inform PBS include systems theory, developmental theory, clinical and counselling psychology, positive psychology, cross-cultural anthropology, and implementation science.

PBS originally was developed as a tertiary approach to individualized assessment and intervention for individuals who engage in severe problem behaviour in home, school, and community settings (Lucyshyn et al., 2015). Since its inception, in addition to tertiary intensive intervention for severe problem behaviour, PBS has evolved into a multi-tiered preventative approach that includes universal strategies to prevent problem behaviour from emerging (Sugai & Horner, 2002), and secondary prevention strategies to prevent mild to moderate levels of problem behaviour from growing in frequency and intensity (Hawken & Horner, 2003; Lane et al., 2012). The emergence of universal and secondary prevention methods has primarily occurred in school settings with the empirical development of School-wide PBS (SWPBS), more recently referred to as Positive Behaviour Interventions and Supports (PBIS) (Horner et al., 2014; McIntosh, Mercer, Nese, Strickland-Cohen & Hoselton, 2016; Sugai & Horner, 2020).

PBS researchers have conducted numerous studies of PBS in school settings at the universal (Freeman et al., 2016; Horner et al., 2009; Sugai & Horner, 2009), secondary (Hawken, Wayman, & Stokes, 2020; McIntosh, Campbell, Carter, & Rossetto Dickey, 2009; Miller, Dufrene, Sterling, Olmi, & Bachmayer, 2015), and tertiary levels (Dunlap, Iovannone, Wilson, Kincaid, & Strain, 2010; Russell Carter, & Horner, 2009; Turton, Umbreit, & Mathur, 2011) of prevention and intervention. These studies have demonstrated the effectiveness of PBS when implemented by educators in schools at each tier of the multi-tiered system of behaviour support. For example, Horner et al. (2009) assessed the effects of implementing universal SWPBS in 30 elementary schools using a randomized, wait-list controlled trial. Results showed significant improvements in school safety, reading assessment scores and office discipline referrals within the experimental group in comparison to the control group. McIntosh et al. (2009) employed a pre-post comparison, quasi-experimental group design to investigate the effectiveness of the secondary prevention strategy of Check-In/Check-Out (CICO) among 36 elementary school students exhibiting attention-motivated problem behaviour. Results showed that the students had statistically significant decreases in problem behaviour and office discipline referrals, and increases in prosocial behaviour. Russell Carter and Horner (2009) used a single case multiple baseline design to investigate the effects of tertiary, function-based multicomponent behaviour support for three elementary school students, kindergarten to grade 1, who did not respond to First Steps to Success (FSS), a secondary prevention program for young students at risk for problem behaviour. Results documented a functional effect, with students showing a substantial decrease in problem behaviour and increase in academic engagement when comparing baseline to intervention phases.

1.1.1 Family-centered positive behaviour support (FCPBS)

PBS researchers in the 1990s developed a family centered approach to PBS with families of children with developmental disabilities and severe problem behaviour through a combination of single case design and qualitative research studies. Single case design studies provided initial evidence of the efficacy of the approach when implemented in collaboration with families in valued but problematic activity setting (i.e. routines) in the home (e.g., morning routine, dinner routine, bedtime routine) and community (e.g., shopping with parent, eating at restaurant with family; Clarke, Dunlap, & Vaughn, 1999; L. K. Koegel, Steibel, & Koegel, 1998; Lucyshyn, Albin & Nixon, 1997; Vaughn, Wilson & Dunlap, 2002). Qualitative studies provided parent perspective and valuable insights that validated or informed important features of the approach. These included the importance of building a trusting partnership with families, demonstrating genuine caring, providing emotional as well as instrumental support, providing family centered services, and incorporating behaviour supports into family routines such as transitions, leisure time, mealtime and bedtime (Fox, Vaughn, Dunlap & Bucy, 1997; Fox, Vaughn, Wyatte, & Dunlap, 2002; Ruef, A. P. Turnbull, Turnbull & Poston, 1999; Turnbull & Ruef, 1996; Turnbull & Ruef, 1997).

Reflecting the core features of a PBS approach, FCPBS is a science-informed, assessment-based approach to developing, in collaboration with family members, technically sound and contextually appropriate multicomponent behaviour support plans for individuals with problem behaviour in family contexts. Behaviour support plans emphasize preventative, skill-building, and positive reinforcement strategies to achieve meaningful, durable and sustainable improvements in child behaviour and family quality of life. FCPBS also takes into consideration family perspectives and family systems when developing supports and thus may include family focused supports as needed to strengthen the family as a whole. Across intervention studies implemented to date, FCPBS is characterized by eight core features (Lucyshyn, Miller, Cheremshynski, Lohrmann, & Zumbo, 2018): (a) development of a collaborative partnership with family members; (b) focus on improving child behaviour and child and family quality of life; (c) the family activity setting (routine) as a unit of analysis and intervention; (d) conduct of a functional assessment; (e) design of multicomponent behaviour support plans that emphasize prevention, teaching and positive reinforcement; (f) contextual fit with family life; (g) *in vivo* training and support to parents and other family members; (h) on-going evaluation of child and family outcomes, and (i) formative and summative evaluation of social validity.

In addition, PBS with families has integrated family centered practices into the approach (Allen & Petr, 1995; Dunst & Sherwindt, 2016). These practices include: (a) viewing the family as the unit of attention; (b) recognizing and building on family strengths; (c) developing authentic family-professional partnerships; (d) ensuring family choice and participation in decision-making; (e) mobilizing informal and formal supports and resources; and (f) actively involving the family in intervention by employing competency enhancing and empowering practices (Dunst, Trivette & Deal, 1988; Lucyshyn, Horner, Dunlap, Albin, & Ben, 2002). During a process of FCPBS, researchers and practitioners: (a) establish a collaborative partnership with family members; (b) conduct a functional assessment; (c) design, in collaboration with the family, a PBS plan that is technically sound and contextually appropriate; (d) provide implementation support to the family in valued but problematic target routines in home and community; (e) empower family members to solve behaviour problems in non-trained family settings; (f) gather ongoing data on parent implementation fidelity, child problem and adaptive behaviour, and social validity; and (g) make adjustments and revisions to the PBS plan and implementation support process to improve the plan's effectiveness, acceptability and sustainability.

Over the past two decades, FCPBS studies have documented the efficacy of the approach across a wide range of child and family characteristics and family settings in the home and community (Lucyshyn & Fossett, 2015). Between 1997 and 2019, single case design research on FCPBS has provided evidence of its effectiveness, acceptability and durability (Binnendyk & Lucyshyn, 2009; Buschbacher, Fox, & Clarke, 2004; Carr et al., 1999; Duda, Clarke, Fox, & Dunlap, 2008; Lucyshyn et al., 1997; Lucyshyn et al., 2007; Lucyshyn et al., 2015; Vaughn, Clarke, & Dunlap, 1997). These studies, taken together, have shown several positive outcomes for children with developmental disabilities and their families including: (a) fidelity of implementation by parents (Binnendyk & Lucyshyn, 2009; Cheremshynski, Lucyshyn & Olson, 2013; Duda et al., 2008; Fettig et al., 2015; Joseph, Strain, & Dunlap, 2019; Moskowitz, Carr, & Durand, 2011; Moskowitz et al., 2017); (b) a decrease in problem behaviour to low or near zero levels (Buschbacher et al., 2004; Cheremshynski et al., 2013; Duda et al., 2008; Fettig et al., 2015; Lucyshyn et al., 1997; Lucyshyn et al., 2007; Lucyshyn et al., 2015; Vaughn, Clarke, et al., 1997; Vaughn, Dunlap, et al., 1997; Vaughn et al., 2002); (c) improvements in child participation in valued family routines in the home and community (Binnendyk & Lucyshyn, 2009; Buschbacher et al., 2004; Cheremshynski et al., 2013; Duda et al., 2008; Lucyshyn et al., 2007; Lucyshyn et al., 2015; Vaughn, Clarke, et al., 1997; Vaughn et al., 2002); (d) improvements in the use of language to communicate wants and needs (Dunlap, Ester, Langhans & Fox, 2006; Moes & Frea, 2002); (e) improvements in child and family quality of life (Binnendyk & Lucyshyn, 2009; Lucyshyn et al., 1997; Lucyshyn et al., 2007; Lucyshyn et al., 2018); and (f) generalization of child behavioural outcomes to non-trained settings in the home and/or community (Carr et al., 1999; Cho Blair, Lee, Cho, & Dunlap, 2011; Lucyshyn et al., 1997; Lucyshyn et al., 2007; Lucyshyn et al., 2015; Moes & Frea, 2002; Sears et al., 2013). In addition, families consistently have rated highly the social validity of the goals, procedures and outcomes of the FCPBS approach (Binnendyk & Lucyshyn, 2009; Buschbacher et al., 2004; Duda et al., 2008; Fettig et al., 2015; Lucyshyn et al., 1997; Lucyshyn et al., 2007; Lucyshyn et al., 2015; Lucyshyn et al., 2018; Vaughn, Dunlap, et al., 1997).

1.2 Meta-analysis of Positive Behaviour Support

In the section below, I briefly summarize recent meta-analyses of multi-tiered PBS in school settings and justify the need for a meta-analysis of FCPBS.

1.2.1 Meta-analysis of PBS in schools

In the past decade, meta-analyses have been conducted to examine the overall effectiveness of PBS in school settings at the universal, secondary and tertiary levels of intervention. These meta-analyses have either included single case designs, group designs or a mix of single case and group designs. Among the meta-analyses conducted, the majority were at the universal level, and examined studies of SWPBS that employed group designs.

1.2.1.1 Meta-analyses at universal level

Solomon, Klein, Hintze, Cressey, and Peller (2012) conducted a meta-analysis across 20 studies of SWPBS at the universal level. These studies enrolled elementary (K-5) and middle school (6-8) students as participants and used single case designs to evaluate functional relations. The authors used the Allison-MT method, a regression-based procedure (Faith, Allison, & Gorman, 1996), to calculate effect sizes and reported them as R^2, the proportion of variance. Results showed that the effect sizes ranged from $r^2 = .27$ to $r^2 = .60$, and that mean outcomes across categories were moderate.

Lee and Gage (2019) conducted a meta-analysis of SWPBS across 29 group-design studies. They were conducted across 32 schools in Europe and United States, ranging from elementary school to high school. The investigators used robust variance estimation meta - analytic models to synthesize school-level behavioural outcomes. Results showed small to medium effect sizes, with SWPBS reducing school discipline exclusions by 0.26 standard deviation units and increasing academic achievement by 0.11 standard deviation units.

Gage, Whitford, and Katsiyannis (2018) conducted a fixed-effect meta-analytic model to examine the effect of SWPBS across 90 schools. Four studies using group experimental designs were identified. Dependent variables were office discipline referrals and school suspensions. Hedges' *g* (Hedges, 1981) was used as the effect size measure to adjust for small sample size. Results showed a significant treatment effect ($g = -.86$) for school suspensions but no effect for office discipline referrals.

1.2.1.2 Meta-analyses at secondary level

Drevon, Hixson, Wyse, and Rigney (2018) investigated 32 studies in which schools implemented the secondary prevention strategy Check-In, Check-Out (CICO). The studies employed either single case or between group research designs. The authors employed a between-case *d* statistic to compute initial effect sizes, and then corrected the *d* statistic for sampling bias using Hedge's *g* (Hedges, 1981). Results showed that CICO improved student outcomes by over one standard deviation, with an effect size (*g)* of 1.22 and 95% confidence interval [1.00, 1.44].

1.2.1.3 Meta-analyses at tertiary level

Goh and Bambara (2012) synthesized 83 school-based single case design studies of tertiary PBS with 185 students in which function-based multicomponent behaviour support plans were implemented. They employed percentage of non-overlapping data (PND) as their effect size measure (Scruggs & Mastropieri, 1998). Results showed that the individualized PBS plans were equally effective across students from diverse populations and educational settings, with moderate effect sizes for decreasing problem behaviour and increasing adaptative behaviour, with a median PND of 88%.

Gage, Lewis and Stichter (2012) investigated the effectiveness of FBA-based interventions for school-aged students (3-16 years) who were at risk for or had a designation of a behavioural disorder. The authors included 69 single case design studies with a total of 146 participants, and employed PND to calculate single case effect sizes. They then conducted a hierarchical linear modeling (HLM) meta-analysis, which provided significance levels for changes in mean, trend, and variability from baseline to intervention. Results showed that FBA-based interventions reduced problem behaviour by 70.5% from baseline to intervention. In doing so, the study largely replicated the results of Goh and Bambara (2012) for tertiary PBS in school settings.

1.2.2 Meta-analysis of PBS with families

Since 1997, PBS has been implemented with families of children with developmental disabilities such as ASD or intellectual disability, yielding over time a relatively large sample of participants who have been enrolled in these studies. Study enrollment of children and their family members (i.e., a case) has ranged from one case (e.g., Clarke, Dunlap, & Vaughn, 1999) to 10 cases (Lucyshyn et al., 2015). Because this is a low incidence population, in any given intervention study by a research team in a particular geographical area, the recruitment of a large number of child participants is difficult. For this reason, traditional, large-scale group design research has not been a feasible approach to investigating the effectiveness of FCPBS within this population. As a result, FCPBS intervention studies have utilized single case research designs which allow researchers to investigate the effects of an intervention on one or a few individuals (Ledford & Gast, 2018). In single case research, a functional relation between the intervention and behaviour change(s) in participants is determined by a visual analysis of changes in the level, trend, and variability of participant behaviour across baseline and intervention conditions (Horner et al., 2005; Horner, Swaminathan, Sugai, & Smolkowski, 2012). Although visual analysis provides researchers with a systematic way to evaluate the effectiveness of an intervention, the nature of single case research design makes it difficult to apply common statistical analyses (e.g., t-test, ANOVA, linear regression) to individual PBS studies to assess effect size (Matyas & Greenwood, 1996). This is due to the small sample size of single case studies, and the possibility of autocorrelation between data paths across baseline and intervention phases. Because visual analysis does not include the computation of effect sizes, it is necessary to conduct a meta-analysis. This is accomplished by generating an equal unit of measurement across studies, combining these data in baseline and intervention phases across studies, generating an effect size for each study, and then synthesizing individual study effect sizes using a statistical method such as regression analysis. A meta-analysis of single case research studies of an intervention informs us of the overall effectiveness of the intervention, and the extent to which this effect across studies is small, medium or large.

Lucyshyn and Fossett (2015), in a summary of the extant literature on PBS with families through 2015, identified 26 studies published in peer-reviewed journals. As noted above, these studies documented a wide range of positive child and family outcomes, including a decrease in problem behaviour, an increase in child participation in family routines, and improvements in child and family quality of life. However, in contrast to school-based implementation of PBS at the universal, secondary and tertiary levels of prevention and intervention, to date there have been no meta-analyses of FCPBS research. Given the number of FCPBS studies conducted to date, it behooves researchers of tertiary PBS with families of children with developmental disabilities to investigate the overall effectiveness of the approach with this population. Since a meta-analysis reviews results across studies that investigate the efficacy or effectiveness of the same or similar treatment (Glass, 1976), a meta-analysis of FCPBS is a necessary next step in the development of FCPBS as an empirically supported treatment.

1.3 Methodological Issues in Meta-Analysis of Single-Case Design Research

Methodologists have applied several statistical techniques (e.g., random effects regression, Hedges' *g*, between-case *d* statistics) and procedures to meta-analyze studies. However, to meta-analyze single case design studies, there are three issues that need to be solved before researchers can implement these techniques and procedures: (a) equalizing the data with different units of measurements; (b) extracting the data from graphs; and (c) evaluating the rigor of single case design studies.

1.3.1 Metrics of meta-analysis of single-case design research

One of the most critical problems that prevent researchers from conducting a meta-analysis of FCPBS using single case design is that the measurement of dependent variables is described in different units (e.g., rate, count, latency, percentage of interval) that are not directly comparable. For example, Lucyshyn et al. (2007) measured problem behaviour by responses per minute (rate) and latency in minutes to termination, while Sears et al. (2013) measured problem behaviour by percentage of intervals and adaptive behaviour by number of bites and steps completed (i.e., count). Currently, there are three metrics that can help researchers overcome this barrier and compute compound effect sizes by equalizing the data across different units of measurement (Moeyaert, Zimmerman, & Ledford, 2018). However, there is no agreed-upon effect size calculation metric in the meta-analysis of single case design. Examples of proposed metrics are shown in Table 1.

Table 1

Examples of Metrics Employed in Previous Meta-analyses of Single-case Design Studies

Metric	Examples
Overlap	1. Percentage of non-overlapping data (Scruggs & Mastropieri, 1998) 2. Tau-U for combining nonoverlap and trend with trend correction (Parker, Vannest, Davis, & Sauber, 2011)
Mean	1. Log response ratio (LRR, Pustejovsky, 2018) 2. Between-case standardized mean difference (Hedges, Pustejovsky, & Shadish, 2012; Hedges, Pustejovsky, & Shadish, 2013)
Regression	1. Bayesian linear regression (Swaminathan, Rogers, & Horner, 2014) 2. Hierarchical linear modeling (Moeyaert, Ferron, Beretvas, & Van den Noortgate, 2014)

An overlap-based metric refers to the degree of nonoverlap between adjacents phases, or evaluations of overlap at the maximum or minimum points in a phase (Parker, Vannest, & Davis, 2011). An overlap-based metric relies on visual analysis, and there is no need to consider the statistical features (i.e., mean, median, SD) of a given dataset. One of its merits is that it can be used across all types of single case designs, including alternating treatment designs which have been difficult to include in meta-analyses that use a mean-based or regression-based approach. Nevertheless, Scruggs and Mastropieri (1998) have pointed out that an overlap-based metric has shortcomings that prevent it from achieving a robust analysis of effect size. These include not being sensitive to changes in slope, not being independent of the number of observations, and being easily influenced by outliers.

A mean-based metric refers to the use of means to quantify the magnitude of change in dependent variables across different phases in one case or across multiple cases. Compared with an overlap-based metric, a mean-based metric can avoid the impact of outliers and more accurately quantify the magnitude of change in dependent variables. However, since there can be autocorrelation between baseline and intervention phases in single case design studies, inference with mean-based metrics can be difficult. In addition, single case design phases with a mean of zero in the baseline phase is not appropriate when calculating the log response ratio due to the nature of the logarithm (Pustejovsky, 2018).

A regression-based metric refers to the use of regression analysis to quantify the magnitude of change in dependent variables. A regression-based metric allows researchers to detect potential interaction effects of two or more variables, and to conduct multilevel meta-analyses. However, many single case design studies have insufficient data for regression analysis, fail to meet independent observation and normal distribution assumptions, and/or mix continuous and non-continuous dependent variables, which renders a meta-analysis problematic (Moeyaert, Zimmerman, & Ledford, 2018).

1.3.2 Data collection technique of meta-analysis of single-case design research

Another issue that is unique in meta-analysis of single case design research is that raw data are not available directly from journal articles. Most commonly, original data are presented in graphical form, and authors may provide additional information such as mean, range and standard deviation. On rare occasion, researchers can request and receive the original data from the study's authors. However, in most cases, researchers complete data extraction on their own. Software such as WebPlotDigitizer (Rohatgi, 2014) and XYit (Geomatix, 2005) make data extraction available so that researchers can proceed with subsequent effect size computation. Data obtained from WebPlotDigitizer have been shown to have relatively high reliability and validity, approximate to their original values (Moeyaert, Maggin, & Verkuilen, 2016).

1.3.3 Prerequisite of meta-analysis of single-case design research

Before completing the statistical analysis of effect size, it is essential to ensure that the study meets basic standards of methodological rigor. The higher the quality of the studies' methodologies, the more confident one can be in the results, which in turn act as the primary data of a meta-analysis (Ledford, Lane, & Tate, 2018). Including studies with inadequate rigor may compromise the outcomes of a meta-analysis. The What Works Clearinghouse (WWC) of the U.S. Institute of Educational Science (IES) provides design and evidence standards for the evaluation of single case design research studies in regard to methodological rigor. These include (a) a single case research design that is appropriate for evaluating the study's research question or hypothesis; (b) a sufficient number of replications of a basic effect to demonstrate a functional effect; (c) high reliability of the dependent variables; and (d) a sufficient amount of data within each phase for each participant (Hitchcock et al., 2014; Wolery, 2013). In addition, the WWC also has established criteria to determine whether an intervention has a sufficient number of high-quality studies to be considered an empirically supported treatment (EST). These criteria are: (a) a minimum of five single case experimental studies that meet design and evidence standards; (b) three or more independent research teams having conducted the studies; and (c) a total of twenty cases (e.g., one or more children and their families) across five or more studies. This is referred to as the 5-3-20 criteria for the establishment of an EST. In previous meta-analyses of secondary (i.e., CICO) and tertiary (i.e., function-based behaviour support) PBS interventions, the researchers used the WWC design and evidence standards to help determine which studies were included in their meta-analyses (Drevon et al., 2018; Walker, Chung, & Bonnet, 2018). Maggin, Pustejovsky, and Johnson (2017) also used the 5-3-20 criteria to determine whether a secondary prevention strategy, group-based positive contingencies, reached the point of becoming an EST.

A high-quality meta-analysis can contribute to an understanding of the generalizability of a treatment across different populations and settings, and also detect variations in overall treatment effects among studies (Moeyaert, Zimmerman, & Ledford, 2018). This can in turn increase the external validity of an individual study, as many of the targeted studies in a meta-analysis are direct or systematic replications of other studies. (Ledford & Gast, 2018). A meta-analysis with sufficient sample size may validate the determination that an intervention is an EST. Recently, there has been a concerted effort in the fields of education and psychology to develop criteria that define research studies of high-quality across different methodologies, and the range of high-quality studies of an intervention to be considered an EST. As noted above, to be qualified as an EST, studies of an intervention must meet the 5-3-20 criteria (Kratochwill et al., 2013). Given the relatively large number of FCPBS studies that have been conducted to date, there is a critical need to conduct a meta-analysis to determine the overall effectiveness and generalizability of FCPBS with families of children with developmental disabilities. There equally is a need to determine whether FCPBS studies conducted to date collectively meet the WWC criteria necessary to be considered an EST.

1.4 Research Questions

The purpose of this study is to: (a) conduct a meta-analysis to examine the effectiveness and generalizability of FCPBS implemented in family contexts to reduce problem behaviour and increase desired behaviour across a broad range of children with developmental disabilities, of families with varied characteristics, and home and community settings in which interventions were implemented; (b) evaluate the rigor of FCPBS studies based on the WWC design and evidence standards; and (c) determine whether single case design studies of FCPBS conducted to date meet the 5-3-20 criteria to be considered an EST. Given the purpose of the study, the following questions will be addressed:

(1) What is the overall effect of FCPBS across children and families?

(2) What is the overall rigor of FCPBS studies to date, according to the WWC design and evidence standards?

(3) Given the rigor of FCPBS studies to date, do they meet the WWC criteria necessary to be categorized as an EST?

(4) Do moderator variables such as age, gender, disability category and WWC rigor standards provide evidence of systematic variation in the magnitude of intervention effects?

Chapter 2: Method

2.1 Inclusion Criteria

The following criteria were used to identify FCPBS intervention studies for inclusion in the meta-analysis: (a) assessment and intervention process showed evidence of key features of FCPBS (i.e., functional assessment; multicomponent behaviour support plan; emphasis on proactive, educative and positive reinforcement strategies; collaboration in plan design; family routine as unit of analysis; attention to contextual fit; implementation support that included *in vivo* coaching or virtual coaching via telehealth); (b) dependent variables included problem behaviour and/or adaptive behaviour; (c) at least one child or youth in the study has a diagnosis of a developmental disability (e.g., ASD, intellectual disability); (d) behaviour support plan implemented in family settings in home and/or community; (e) single case design research methods used to determine whether basic effects (e.g., single baseline A-B-C design) or functional effects (e.g., multiple baseline design across three or more family routines) were documented; (f) problem behaviour and/or adaptive behaviour data were presented in graphical format, allowing for data extraction and effect size calculation; and (g) study was published in a peer-reviewed journal and written in English without publication time limits.

2.2 Literature Search

Two strategies were used to identify eligible studies for meta-analysis: (a) electronic reference database search, and (b) hand search. Electronic reference database searches included PsycINFO and Google Scholar. Keyword combination included (families AND positive behaviour support), (function-based intervention AND families), (autism AND positive behaviour support AND parents), (behaviour support AND parent training), (family-centered AND behaviour), (family routine AND behaviour), (family routine AND autism), (positive behaviour support AND infant AND problem behaviour), (collaboration AND family AND behaviour support), and (longitudinal AND behaviour support AND toddler). We did not set *Positive Behaviour Support* into the search term in each search as we found that not all PBS studies explicitly stated PBS in their titles or abstracts. Including "PBS" in every search tended to eliminate the outcomes of the search. To minimize confounding factors, the search keywords were not case or plural sensitive. For instance, we added * at the end of the keyword. The system then treated the asterisk as either single form or plural form (e.g., both "support" and "supports" would appear in the search outcomes).

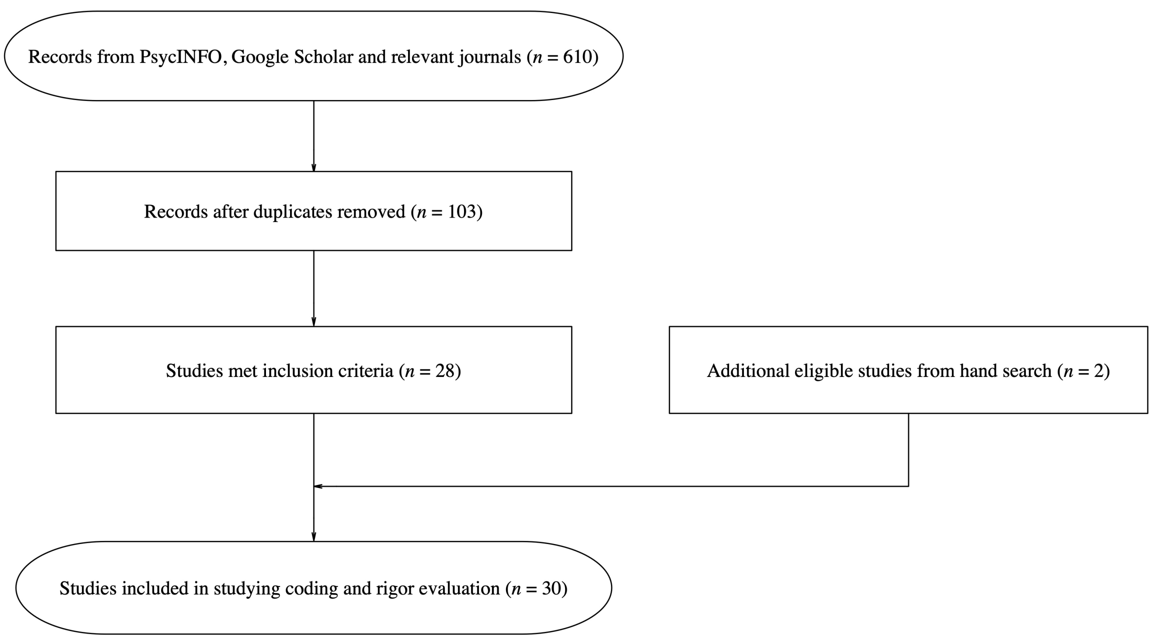
We also implemented a hand search from relevant journals to ensure that we did not miss any potential qualifying study. The following journals were reviewed: *Journal of Positive Behaviour Interventions*, *Behavioural Disorders*, *Topics in Early Childhood Special Education*, *Journal of Autism and Developmental Disorders*, *Journal of Child and Family Studies*, *Journal of the Association for Persons with Severe Handicaps*, *Education and Training in Autism and Developmental Disabilities*.

During initial screening of articles, the first and second screener worked together to search databases and journals, screenshot search outcomes, and read study abstracts. Then, the second screener determined the eligibility of each study by checking whether or not each study included each core component of FCPBS. He then explained his judgment and reached an agreement with the first screener. Finally, the first and second screeners checked the rest of the inclusion criteria together to make a final decision for each study. The first screener then reviewed his determination with the second screener, explained his judgement, as needed, and worked toward consensus.

After searching from PsycINFO, Google Scholar and relevant journals, 610 records were judged eligible by reading their titles and abstracts. After removing repetitive results, 103 studies remained. We then checked whether each study met inclusion criteria, and found that 74 studies missed at least one component (i.e., not all components of FCPBS were demonstrated in the study, not a single case design study, not published in a peer-reviewed journal, not written in English). Lastly, as of January 2020, we employed a hand search and found two additional qualifying studies. Therefore, 30 studies were included and then submitted to rigor evaluation. The process is illustrated in the flowchart (Figure 1) below.

Figure 1

Flowchart of Literature Search.



2.3 Data Collection

Data collection included two steps: study coding and data extraction. These steps are described below.

2.3.1 Study Coding

Study coding began immediately after all eligible studies were located. The first and the second coders extracted the following information across eight features of each study: (a) participants demographic characteristics (i.e., age, gender, ethnicity and disability category); (b) functional behaviour assessment (FBA) type (i.e., descriptive or experimental); (c) intervention type (i.e., multi-component, skills training, or other); (d) study design (i.e., withdrawal/reversal, multiple baseline, alternating treatment, or empirical case study); (e) topography of behaviour (i.e., positive or negative behaviour); (f) measurement of dependent variables (i.e., rate, count, time interval, or percentage of interval); (g) evaluation of social validity (i.e., yes or no); and (h) measurement of generalization and maintenance (i.e., yes or no).

In regard to demographic characteristics: (a) age refers to the age of participants at the beginning of the study; (b) gender refers to the gender of the participant, either male or female; (c) ethnicity refers to the ethnic group that participants belong to, either Caucasian, Asian, African, Aboriginal, Hispanic, Middle Eastern or not clear; and (d) disability category refers to the disability diagnosis of the participants, either ASD or other developmental disabilities.

FBA type and intervention type coding were adapted from a meta-analytical study of SWPBS by Goh and Bambara (2012). FBA type refers to FBA methods that were used to assess participants' problem behaviour and is made up of two categories: (a) experimental (e.g., functional analysis, structural analysis, hypothesis testing); and (b) descriptive (e.g., interview, direct observation, archival record review, rating scales). Intervention type refers to the FCPBS intervention implemented with participants and is made up of three categories: (a) multicomponent intervention (a combination of antecedent-based strategies, teaching strategies, and consequence-based strategies); (b) skills training (e.g. functional communication, self-management); and (c) other (i.e., not covered by the above).

Study design refers to the design type that the study used to document a functional relation or basic effect(s), including withdrawal/reversal, multiple baseline, alternating treatment, and empirical case study. In terms of the topography of behaviours, positive behaviour refers to the behaviour that is expected to increase following intervention and problem behaviour refers to the behaviour that is expected to decrease following intervention. Measurement of dependent variables refers to the unit that is used to quantify the occurrence of each behaviour and is comprised of count, rate, time interval and percentage of interval. Evaluation of social validity, and measurement of generalization and maintenance were coded as either "yes" or "no," with "yes" meaning the study has the component and "no" meaning the study does not have the component.

Intercoder agreement. The first screener served as the primary coder, while the second screener provided training and ongoing supervision and support. Once trained to establish interrater agreement (IRA) across the eight features of each study, the primary coder provided training to a secondary coder. After obtaining a minimum of 80% IRA across three consecutive FCPBS studies, the secondary coder independently scored a randomly selected set of 30% of all included FCPBS studies. After coding, the primary and secondary coders discussed their results and resolved disagreements. IRA was calculated by dividing the number of agreements by the number of agreements plus disagreements and then multiplying by 100. The mean IRA for study features was 93.4%.

2.3.2 Data extraction

Since raw data were not available from authors, the software program WebPlotDigitizer (Rohatgi, 2014) was used to extract the data from the published study's single case design graphs. Coordinates of each data point shown on WebPlotDigitizer for each study were converted to their original scale in preparation for future statistical analysis. The primary coder did all data extraction and corrected any errors between the studies' extracted data and graphed data. For example, he checked the extracted data and computed the mean for each phase of the study and then compared these means with those provided in the study to determine whether there were any potential errors during data extraction. The secondary coder checked all data extracted by the primary coder to ensure the data were 100% correct.

2.4 Data analysis

Effect sizes were calculated for each of the single case research studies. Following these within study analyses, a hierarchical meta-analysis model was used to synthesize effect size estimates. These analysis procedures are described below.

2.4.1 Effect sizes calculation

Following data collection, effect sizes were calculated for each case in preparation for meta-analysis. To do so, I used the log response ratio (LRR), a statistic that employs a mean-based metric, to compute effect sizes, as proposed and justified by Pustejovsky (2018) for single case research studies. LRR is defined as the natural logarithm of the ratio between the mean outcome in intervention phases divided by the mean outcome in baseline phases, both of which are positive numbers by nature. If there is no change from baseline to intervention, the LRR value should be zero. If the intervention results in an increase in the level of dependent variable(s), the LRR should be positive, and if the intervention results in a decrease in the level of dependent variable(s), the LRR should be negative.

The reason for employing the LRR is that it has three advantages when computing effect sizes from single case studies (Pustejovsky, 2018), compared to PND and BC-SMD. First, the LRR is not sensitive to the changes in the units of measurement of dependent variables. Even when the unit of measurement changes from rate per minute to percentage of intervals, for example, the effect sizes generated from the LRR will remain the same. Therefore, effect sizes will be comparable when using different types of units of measurement across studies. Second, the LRR is transformable to the metric of percentage change, which aids in the interpretation of study outcomes. Effect size differences between baseline and intervention phases computed by the LRR can be converted to percentage data so that the effectiveness of a given treatment can be evaluated in a more intuitive way. Third, the range of natural LRR can range from negative infinity to positive infinity, which removes potential ceiling effects.

To make the LRR meaningful for computing effect sizes, three conditions must be met (Pustejovsky, 2018). First, dependent variables must be continuous and signed where zero means the absence of the target behaviour. Second, either the baseline response or the intervention response cannot be exactly zero; otherwise the LRR would be infinite, due to the nature of logarithm. Third, outcomes have to be stable within each phase without auto-correlation, a requirement that sometimes cannot be met in single case design studies. The first and the second conditions are usually met in FCPBS studies, as can be seen, for example, in Lucyshyn et al. (1997), Buschbacher et al. (2004), Cho Blair et al. (2011), and Joseph et al. (2019). Although the third condition can be difficult to meet, Pustejovsky (2018) argued that the LRR method for computing effect sizes in single case research during a meta-analysis is still valid, as robust variance estimates mitigate this problem.

There are three issues that require attention when computing effect sizes for single case design research. First, the direction of change (i.e., positive or negative) of dependent variables may not be consistent with the direction of therapeutic improvement for the target behaviours of interest (Pustejovsky, 2018). Researchers cannot synthesize these outcome values until transformation of these values is completed. In some single case design studies, researchers examine desired behaviours which they aim to increase, such as peer interaction or engaged time whereas other studies examine problematic behaviours which they aim to decrease such as physical aggression or off-task behaviour. Still other single case researchers examine both desired and problematic behaviours. For example, within the FCPBS literature, Buschbacher et al. (2004) used percentage of intervals to describe changes in problem behaviour and in positive engagement in three target family routines. The former target behaviour was expected to decrease while the latter was expected to increase. In this proposed meta-analysis, when both types of behaviours were measured in a study, I transformed one of them to make all measurements uniform, after which effect sizes could be computed directly. The solution proposed by Pustejovsky (2018) for this transformation was followed. For measures that were count or frequency/rate, I reversed the sign of effect sizes to make all of these data consistent. For measures that were percentage of intervals, I subtracted the original data from 100% to make all of these data uniform.

The second issue is that some studies have more than two phases per case, which also prevents one from synthesizing outcome values. As noted above, the LRR is defined by calculating effect size using the LRR is the natural logarithm of the ratio between the mean outcome in intervention phases divided by the mean outcome in baseline phases. With this calculation, it is assumed that for each case, there is one baseline and one treatment phase. However, some studies use designs that have more than one baseline and intervention phase. FCPBS studies commonly use a variety of single case designs, such as a withdrawal/reversal (A-B-A-B) design, multiple baseline design, or alternating treatment design. For withdrawal/reversal designs, for each case there is one or more replications of baseline and treatment, which are contrasted with each other. For multiple baseline designs, there is one baseline phase and one intervention phase that are contrasted for two or more cases. For alternating treatment designs, sometimes there is no baseline phase, and one treatment is contrasted with a second treatment over time to determine which may be more effective. Compared to the other designs, the withdrawal/reversal design is problematic in regard to inclusion in a meta-analysis due to the presence of more than one baseline and intervention cycle. To address this problem, I computed the LRR for the A_1-B_1 (initial baseline and initial treatment) comparison and A_2-B_2 comparison (return to baseline and reintroduction of treatment) and then computed the average between the two sets of comparisons (Pustejovsky, 2018). Doing so utilizes all of the data, provides a precise estimate of treatment effects and maintains consistency with the logic of visual analysis.

The third issue is that some single case design studies have phases with a small number of observations (i.e., 3-4). In these cases, computing the LRR for these data without numerical adjustments will generate small-sample bias to the effect size estimates. To address this problem, I applied the bias-corrected estimator proposed by Pustejovsky (2015) in the LRR effect size calculation.

2.4.2 Meta-analysis

Meta-analysis was conducted based on the LRR effect sizes. I employed a three-level, hierarchical meta-analysis model introduced by Van den Noortgate and Onghena (2008) for synthesizing effect size estimates. The reason for employing the hierarchical meta-analysis model is that one single case design study may include multiple cases, which produces multiple case-level effect size estimates. Case-level effect size estimates nested within a study results in a hierarchical structure. When conducting a meta-analysis of a treatment for which group design methods were employed, each study generates an effect size estimate at the *study-level*. In contrast, when conducting a meta-analysis of a treatment for which a single case design was employed, each study generates multiple effect size estimates at the *case-level*. For example, in a study using a multiple baseline design across three participants, each participant is regarded as one case. Given this, there will be three case-level effect size estimates.

Due to the hierarchical structure that results from conducting a meta-analysis of single case design studies, the LRR estimates for a given case are affected by a case-level error, a study-level error, and a sampling error. The overall average effect size is computed as the weighted average of the study-level effect size estimates that are derived from the weighted case-level effect size estimates. For case-level effect size estimates, the weight is assigned by the variance of case-level error and the variance of sampling error where larger errors lead to fewer weights. For study-level effect size estimates, the weight is assigned by the variance of case-level error, the variance of study-level error and the variance of sampling error where larger errors lead to fewer weights, in which the variance of study-level error is more important in determining the weight. My primary interest in this meta-analysis is the overall average effect size, the variance of case-level error, and the variance of study-level error. These two errors play an important role in generating case-level and study-level effect size estimates and in determining heterogeneity. They are described below.

Case-level error refers to the discrepancy between the true effect size for a case and the average true effect size for all cases in its corresponding study. A large variance of case-level error indicates that intervention effects are less consistent across cases within a study, which may come from the characteristics of the population, fidelity of implementation of the intervention, reliability of the data, and/or length of the intervention across cases. Study-level error refers to the discrepancy between the average true effect size for a study and the overall average effect. A large variance of study-level error indicates that intervention effects are less consistent across studies. Similar to case-level error, a large variance of study-level error may come from the characteristics of the population, fidelity of implementation of the intervention, reliability of the data, and/or length of the intervention across studies. Sampling error refers to the discrepancy between the LRR effect size estimate and the true effect size parameter for a given case.

The variance of case-level error and the variance of study-level error were computed using the *metafor* package (Viechtbauer, 2010) in R software (R Core Team, 2019). Once the variance of case-level error and study-level error were computed, the case-level LRR, the study-level LRR and the overall average effect size would be generated. Meanwhile, the standard error was computed using a statistical technique referred to as *robust variance estimation*, which accounts for the correlation among effects sizes within a case. As mentioned above, auto-correlation may exist in single case design studies. For this reason, conventional standard error may be too large, and thus subsequent inferences (e.g., via confidence intervals) may be underpowered. To solve this problem, robust variance estimation, proposed by Hedges, Tipton, and Johnson (2010), was conducted using the *clubSandwich* package (robust variance estimators with small-sample corrections; Pustejovsky, 2017) in R software (R Core Team, 2019). Once a robust standard error was generated, a robust variance estimate and confidence intervals would be calculated. To help interpret the overall average effect size in an intuitive way, I converted LRR results into percentage change by $100\% \times [exp(\hat{\gamma}) - 1]$ (Pustejovsky, 2018).

2.4.3 Moderator analysis

To investigate whether other variables in the study have an impact on the magnitude of effect sizes, a moderator analysis was conducted. Moderators included age (a continuous variable), gender (a binary variable), disability category (a binary variable), and WWC quality standards evaluation (a binary variable). Meta-regression was used to compare the effect size estimates among subgroups of each moderator. These were: (a) the age of child participants from 2-years-old to 17-years-old; (b) male vs. female; (c) ASD vs. other developmental disabilities; and (d) studies that met WWC quality standards vs. studies that did not meet WWC quality standards.

2.5 Evaluation of Rigor

WWC procedures for evaluating single case research designs (Kratochwill et al., 2013) were used to assess the rigor of each of the FCPBS studies that met inclusion criteria. Rigor included WWC design standards and evidence standards, as described below.

2.5.1 Design standards

The first stage was to assess the design standards of all the studies that met FCPBS inclusion criteria. Design standards criteria included: (a) independent variable manipulated systematically; (b) systematic measurement of outcome variables over time by at least two assessors; (c) interobserver agreement (IOA) collected for at least 20% of sessions across each condition, with at least 80% agreement or equivalent (i.e., Cohen's Kappa of $\geqslant 0.60$); (d) at least three attempts to demonstrate an intervention effect at three different points in time; (e) sufficient phases (at least 4 for withdrawal/reversal design and at least 6 for multiple baseline design); (f) sufficient data points within each phase (minimum of 3); and (g) for multiple probe designs, overlap of initial baseline probe data points and probe data points present just prior to introducing independent variable.

According to WWC, studies that met all the above requirements, with no less than 5 data points in each phase, were rated as meeting standards without reservations. Studies that met all the above requirements, with 3 or 4 data points in any phase, were rated as meeting standards with reservations. Studies that failed to meet any of the above requirements were rated as not meeting design standards. For multiple probe designs, studies had to include three consecutive data points for each case within the first three sessions in the baseline phase, and within three sessions just prior to intervention to be qualified for meeting standards without reservations. Multiple probe design studies had to include one data point for each case within the first three sessions in the baseline phase, and within three sessions just prior to intervention to be qualified for meeting standards with reservations.

After initial screening of all the studies, we found that not every study explicitly mentioned how they allocated the proportion of interobserver agreement data across each condition, and thus the 20% IOA requirement could hardly be achieved. However, considering that most of the included studies were published prior to WWC standards for single case research and that some studies constrained data collection for some phases due to social validity and ethical concerns, we thought it unfair to rule out these studies as they may nevertheless demonstrate high quality. For this reason, we did not label "not meeting standards" to studies if they did not clearly state how they allocated IOA observations across each condition of the study. Instead, if studies demonstrated that IOA observations were conducted for 20% or more of observation sessions, and IOA was 80% or better, they were rated as meeting standards with reservation.

2.5.2 Evidence standards

The second stage was to assess the evidence standards from the single case design studies that were rated in the first assessment stage as either meeting standards without reservation or meeting standards with reservation. The purpose of this stage was to detect the existence of a causal effect (functional relation) through structured visual analysis. To demonstrate a functional relation, three basic effects had to be identified. Six indicators were used to determine a basic effect: level, trend, variability, immediacy of effect, overlap, and consistency of data in similar phases. Level represented the mean score for the data within a given phase. Trend represented the slope of the best-fitting line for the data within a given phase. Variability represented the range of data around the best-fitting line. Immediacy of effect represented the change in mean score between the last three data points in one phase and the first three data points in the following phase. Overlap represented the proportion of data points from one phase that overlaps with data points in the adjacent phase. Consistency of data in similar phases represented the consistency in the data for all phases within the same condition (e.g., all baseline phases).

These six indicators led to an overall conclusion about the strength of the evidence: (a) strong evidence (three or more basic effects without non-effect); (b) moderate evidence (three or more basic effects with at least one non-effect); and (c) no evidence. Studies with any of the following characteristics were deemed as showing no evidence of an effect: (a) inconsistent data pattern within any single phase (i.e., too much variability); (b) too much overlap between two adjacent phases; (c) long latency between the introduction of treatment and change in dependent variables; (d) inconsistent patterns across similar phases; (e) less than three basic effects; and (f) absence of a significant change in mean score between two adjacent conditions (i.e., little to no difference in level, trend, and/or variability).

Interrater agreement. The first screener served as the primary rater, while the second screener served as the secondary rater. The secondary rater provided training to the primary rater until a minimum of 80% IRA was obtained across six single case studies, balanced across single case research designs. Both raters then independently scored a randomly selected set of 30% of the included FCPBS studies. After scoring, the raters discussed their IRA results and resolved any disagreements. IRA was calculated by dividing the number of agreements by the number of agreements plus disagreements and then multiplying by 100. The mean IRA for WWC design standards was 84.98% and 100% for WWC evidence standards.

Chapter 3: Results

In this section, I will first present results from study coding at study-level and case-level. Next, I will present results from effect size calculation at study-level and case-level with forest plots followed by meta-analytic results. Then, I will present results from rigor evaluation for WWC design standards and evidence standards.

3.1 Characteristics of Included Studies and Cases

This subsection presents study-level and case-level characteristics from study coding.

3.1.1 Study-level characteristics

All of the 30 studies were published between 1997 and 2019. Table 2 provides a comprehensive delineation of key characteristics across these studies. This study-level coding provides an understanding of the methodological and clinical characteristics of each study. In regard to research design, a majority (*n* = 23, 76.7%) of the studies employed multiple baseline designs, while three studies (10%) employed withdrawal/reversal designs, one study (3.3%) employed an alternating treatment design, and three studies (10%) employed an empirical case study design (i.e., A-B). In regard to dependent variables, 29 studies (96.7%) included measures of problem behaviour and 20 studies (66.7%) included measures of adaptive behaviour such as positive engagement in routine (*n* = 7), routine steps completed (*n* = 4), appropriate behaviour (i.e. social interaction; *n* = 5) or replacement behaviour (i.e., functional communication; *n* = 6). During the FCPBS assessment process, 80% (*n* = 24) of the studies used a descriptive FBA such as a functional assessment interview, direct observations, archival record review, and rating scales. Twenty percent (*n* = 6) of the studies used an experimental FBA such as a functional analysis, structural analysis, or hypothesis testing. Following the FBA, in most of the studies (*n* = 27, 90%) a multicomponent plan was implemented, while in three studies (10%) skills training (functional communication) was implemented. After intervention, in 66.7% (*n* = 20) of the studies maintenance support was provided, and in 26.7% of the studies (*n* = 8) generalization was promoted. In 18 studies (60%) social validity data from families was collected. A summary of study level characteristics can be seen in Table 3. For several studies, some dependent variables were excluded from the summary and final meta-analysis because they measured adult behaviour (e.g., Carr et al., 1999) or repeated measures data were not presented (i.e., only mean, range and SD provided; Lucyshyn et al., 2015), which did not meet inclusion criteria.

Table 2

Key Characteristics of Studies Included in the Meta-analysis

Study	Research Design	Dependent Variables (Child)	Generalization Measured	Maintenance Measured	FBA Type	Intervention Type	Social Validity Collected
Bailey & Cho Blair (2015)	MB	Problem Behaviour Appropriate Behaviour	No	Yes	Descriptive	Multi-component	Yes
Barry & Singer (2001)	MB	Problem Behaviour Replacement Behaviour	No	Yes	Experimental	Multi-component	No
Binnendyk & Lucyshyn (2009)	ECS	Appropriate Behaviour Routine Duration	Yes	Yes	Descriptive	Multi-component	Yes
Buschbacher, Fox, & Clarke (2004)	MB	Problem Behaviour Positive Engagement	No	Yes	Descriptive	Multi-component	Yes
Carr et al. (1999)	MB	Problem Behaviour Positive Engagement Replacement Behaviour	Yes	Yes	Descriptive	Multi-component	Yes
Cheremshynski, Lucyshyn, & Olson (2013)	WR	Problem Behaviour Steps Completed Routine Duration	No	Yes	Descriptive	Multi-component	Yes
Cho Blair et al. (2011)	MB	Problem Behaviour Appropriate Behaviour	Yes	Yes	Descriptive	Multi-component	Yes
Chu (2012)	ECS	Problem Behaviour Routine Duration Appropriate Behaviour Steps Completed	No	Yes	Descriptive	Multi-component	Yes
Clarke, Dunlap, & Vaughn (1999)	WR	Problem Behaviour Positive Engagement	No	Yes	Descriptive	Multi-component	No
Duda et al. (2008)	MB	Problem Behaviour Positive Engagement	No	Yes	Descriptive	Multi-component	Yes
Dunlap & Fox (1999)	MB	Problem Behaviour	No	No	Descriptive	Multi-component	No
Dunlap, Ester, Langhans, & Fox (2006)	MB	Problem Behaviour Replacement Behaviour	No	No	Descriptive	FCT	Yes
Erbas (2010)	AT	Problem Behaviour	No	Yes	Descriptive	Multi-component	No
Fettig & Ostrosky (2011)	MB	Problem Behaviour	No	Yes	Descriptive	Multi-component	No
Fettig, Schultz, & Sreckovic (2015)	MB	Problem Behaviour	No	No	Descriptive	Multi-component	Yes

Table 2 Continued

Study	Research Design	Dependent Variables (Child)	Generalization Measured	Maintenance Measured	FBA Type	Intervention Type	Social Validity Collected
Frea & Hepburn (1999)	MB	Problem Behaviour Replacement Behaviour	No	No	Descriptive	FCT	No
Joseph et al. (2019)	WR	Problem Behaviour	No	No	Descriptive	Multi-component	No
Knowles et al. (2017)	MB	Problem Behaviour	No	No	Descriptive	Multi-component	Yes
L. K. Koegel, Steibel, & Koegel (1998)	MB	Problem Behaviour	No	Yes	Descriptive	Multi-component	Yes
Lucyshyn et al. (2007)	MB	Problem Behaviour Routine Duration	Yes	Yes	Experimental	Multi-component	Yes
Lucyshyn et al. (2015)	MB	Problem Behaviour Steps Completed	No	Yes	Experimental	Multi-component	Yes
Lucyshyn, Albin, & Nixon (1997)	MB	Problem Behaviour Routine Duration	Yes	Yes	Experimental	Multi-component	Yes
Moes & Frea (2000)	ECS	Problem Behaviour Positive Engagement Replacement Behaviour	Yes	Yes	Descriptive	Multi-component	No
Moes & Frea (2002)	MB	Problem Behaviour Replacement Behaviour	Yes	Yes	Experimental	FCT	No
Moskowitz, Carr, & Durand (2011)	MB	Problem Behaviour	No	No	Experimental	Multi-component	Yes
Moskowitz et al. (2017)	MB	Problem Behaviour	No	No	Descriptive	Multi-component	No
Sears et al. (2013)	MB	Problem Behaviour Appropriate Behaviour Steps Completed	Yes	No	Descriptive	Multi-component	Yes
Vaughn, Clarke, et al. (1997)	MB	Problem Behaviour Positive Engagement	No	Yes	Descriptive	Multi-component	No
Vaughn, Dunlap, et al. (1997)	MB	Problem Behaviour	No	Yes	Descriptive	Multi-component	Yes
Vaughn et al. (2002)	MB	Problem Behaviour Positive Engagement	No	No	Descriptive	Multi-component	No

Note. WR = Withdrawal/reversal; MB = Multiple baseline; AT = Alternating treatment; ECS = Empirical case study; FCT = Functional Communication Training.

Table 3

Summary of Study-level Coding

Study-Level Characteristics	*N* of Studies
Type of research design	
Multiple baseline	23
Withdrawal/reversal	3
Alternating treatment	1
Empirical case study	3
Dependent variables	
Problem behaviour	29
Adaptative behaviour	20
Type of FBA	
Descriptive	24
Experimental	6
Type of Intervention	
Multicomponent	27
FCT	3
Maintenance support	20
Generalization promotion	8
Social validity data collected	18

Note. $N = 30$.

3.1.2 Case-level characteristics

As mentioned above, one of the core features of FCPBS is that the family is the unit of analysis. Among the 30 studies in the meta-analysis, there were 56 focus children (i.e., children with a developmental disability) in 56 families who received the FCPBS intervention. Therefore, there were 56 cases in the final meta-analysis. Table 4 presents case-level demographic information on focus children and family involvement in the intervention. Of the 22 of 56 families (39.3%) whose ethnicity was note, there were 13 participants identified as Caucasian, six identified as Asian, one identified as African, one identified as Middle Eastern and one identified as Hispanic. The mean age of focus children at the beginning of each study was 5.4 years (range = 2-17 years). Most of the participants were males (n = 45, 80.4%), with fewer females (n = 10, 17.9%). As for disability diagnosis, 57.1% (n = 32) of the participants were diagnosed with ASD while 41.1% (n = 23) of the participants were diagnosed with developmental disabilities other than ASD (e.g., intellectual disability, developmental delay).

Across the majority of families (94.6%, n = 53), the child's mother was the implementer of the intervention in target family routines. For 27 families (48.2%), the child's father implemented the intervention or was a participant in the target routines. In 16 families (28.6%), siblings were participants in the target routines. In 17 families (32.1%), two members implemented the intervention, or one member implemented the intervention and the other member was involved. In 12 families (21.4%), the mother, father and one sibling implemented the intervention. Typically, mothers and/or fathers were responsible for implementing the behaviour support plan, while siblings were natural participants in the targeted routines, such as dinner, shopping, and restaurant. A summary of these features can be seen in Table 5. Some cases were excluded from the summary and meta-analysis because the focus person(s) did not have a diagnosis of a developmental disability or they were adults, which did not meet inclusion criteria.

Table 4

Key Characteristics of Cases Included in the Meta-analysis

Study	Child	Ethnicity	Age	Gender	Disability	Mother	Father	Siblings
Bailey & Cho Blair (2015)	1	Caucasian	7	Male	ASD	Involved	Not involved	Not involved
	2	Caucasian	6	Male	ASD	Involved	Not involved	Not involved
	3	Caucasian	5	Male	ASD	Involved	Not involved	Involved
Barry & Singer (2001)	1	Not Clear	10	Male	ASD	Involved	Involved	Involved
Binnendyk & Lucyshyn (2009)	1	Middle East	6	Male	ASD	Involved	Involved	Not involved
Buschbacher, Fox, & Clarke (2004)	1	Not Clear	7	Male	Other Disability	Involved	Involved	Not involved
Carr et al. (1999)	1	Not Clear	14	Male	Other Disability	Involved	Involved	Involved
	2	Not Clear	17	Male	Other Disability	Involved	Involved	Not involved
Cheremshynski, Lucyshyn, & Olson (2013)	1	Asian	5	Male	ASD	Involved	Not involved	Not involved
Cho Blair et al. (2011)	1	Asian	4	Female	ASD	Involved	Not involved	Not involved
	2	Asian	4	Male	ASD	Involved	Not involved	Not involved
	3	Asian	5	Male	ASD	Involved	Not involved	Not involved
Chu (2012)	1	Asian	6	Male	ASD	Not Involved	Involved	Not involved
Clarke, Dunlap, & Vaughn (1999)	1	Not Clear	10	Male	ASD	Involved	Not involved	Not involved
Duda et al. (2008)	1	Caucasian	2	Male	Other Disability	Involved	Not involved	Involved
Dunlap & Fox (1999)	1	Caucasian	2	Female	ASD	Involved	Involved	Not involved
	2	Caucasian	2	Male	ASD	Involved	Involved	Not involved
	3	Caucasian	3	Male	ASD	Involved	Involved	Not involved
	4	Caucasian	3	Male	ASD	Involved	Involved	Not involved
	5	Asian	2	Male	ASD	Involved	Involved	Not involved
	6	Hispanic	2	Male	ASD	Involved	Involved	Not involved
Dunlap, Ester, Langhans, & Fox (2006)	1	Not Clear	3	Female	Other Disability	Involved	Not involved	Not involved
	2	Not Clear	2	Female	Other Disability	Involved	Not involved	Not involved

Table 4 Continued

Study	Child	Ethnicity	Age	Gender	Disability	Mother	Father	Siblings
Erbas (2010)	1	Not Clear	4	Male	Other Disability	Involved	Not involved	Not involved
	2	Not Clear	5	Male	Other Disability	Involved	Not involved	Not involved
	3	Not Clear	5	Male	Other Disability	Involved	Not involved	Not involved
Fettig & Ostrosky (2011)	1	Not Clear	3	Male	Other Disability	Involved	Not involved	Involved
Fettig, Schultz, & Sreckovic (2015)	1	Not Clear	3	Male	ASD	Not involved	Involved	Not involved
	2	Not Clear	3	Female	Other Disability	Involved	Not involved	Not involved
	3	Not Clear	5	Male	ASD	Involved	Not involved	Not involved
Frea & Hepburn (1999)	1	Not Clear	4	Male	ASD	Involved	Not involved	Not involved
	2	Not Clear	4	Male	ASD	Involved	Not involved	Not involved
Joseph et al. (2019)	1	Caucasian	3	Female	Other Disability	Involved	Involved	Involved
Knowles et al. (2017)	1	African	2	Male	Other Disability	Involved	Not involved	Not involved
	2	Caucasian	3	Male	Other Disability	Involved	Not involved	Not involved
L. K. Koegel, Steibel, & Koegel (1998)	1	Not Clear	5	Female	ASD	Involved	Involved	Involved
	2	Not Clear	4	Male	Other Disability	Involved	Involved	Involved
	3	Not Clear	4	Male	ASD	Involved	Involved	Involved
Lucyshyn et al. (2007)	1	Caucasian	5	Female	ASD	Involved	Involved	Involved
Lucyshyn et al. (2015)[b]	1	-	-	-	-	-	-	-
Lucyshyn, Albin, & Nixon (1997)	1	Not Clear	14	Female	Other Disability	Involved	Involved	Not involved
Moes & Frea (2000)	1	Not Clear	3	Male	ASD	Involved	Involved	Involved
Moes & Frea (2002)	1	Not Clear	3	Female	ASD	Involved	Involved	Involved
	2	Not Clear	3	Male	ASD	Involved	Involved	Involved
	3	Not Clear	3	Male	ASD	Involved	Involved	Involved
Moskowitz, Carr, & Durand (2011)	1	Not Clear	10	Male	Other Disability	Involved	Not involved	Not involved
	2	Not Clear	7	Male	Other Disability	Involved	Not involved	Not involved
	3	Not Clear	9	Male	Other Disability	Involved	Not involved	Not involved

Table 4 Continued

Study	Child	Ethnicity	Age	Gender	Disability	Mother	Father	Siblings
Moskowitz et al. (2017)	1	Not Clear	6	Male	ASD	Involved	Involved	Not involved
	2	Not Clear	8	Male	ASD	Involved	Involved	Not involved
	3	Not Clear	9	Male	ASD	Involved	Involved	Not involved
Sears et al. (2013)	1	Caucasian	4	Male	Other Disability	Involved	Involved	Involved
	2	Caucasian	6	Male	ASD	Involved	Not involved	Not involved
Vaughn, Clarke, et al. (1997)	1	Not Clear	8	Male	Other Disability	Involved	Not involved	Not involved
Vaughn, Dunlap, et al. (1997)	1	Not Clear	9	Male	Other Disability	Involved	Not involved	Not involved
Vaughn et al. (2002)	1	Not Clear	7	Male	ASD	Involved	Not involved	Involved

Note. [b]Although this study included 10 children with a disability and their families, only one family's single case design graph was presented, allowing for a meta-analysis only for this family. Authors did not provide case-level information for this family.

Table 5

Summary of Case-level Coding

Case-Level Characteristics	*N* of Studies
Ethnicity	
Caucasian	13
Asian	6
African	1
Middle East	1
Hispanic	1
Not Clear	33
Gender	
Male	45
Female	10
Disability	
ASD	32
Other	23
Family involvement	
Mother involved	53
Father involved	27
Siblings involved	16

Note. $N = 56$.

3.2 Meta-analytic results

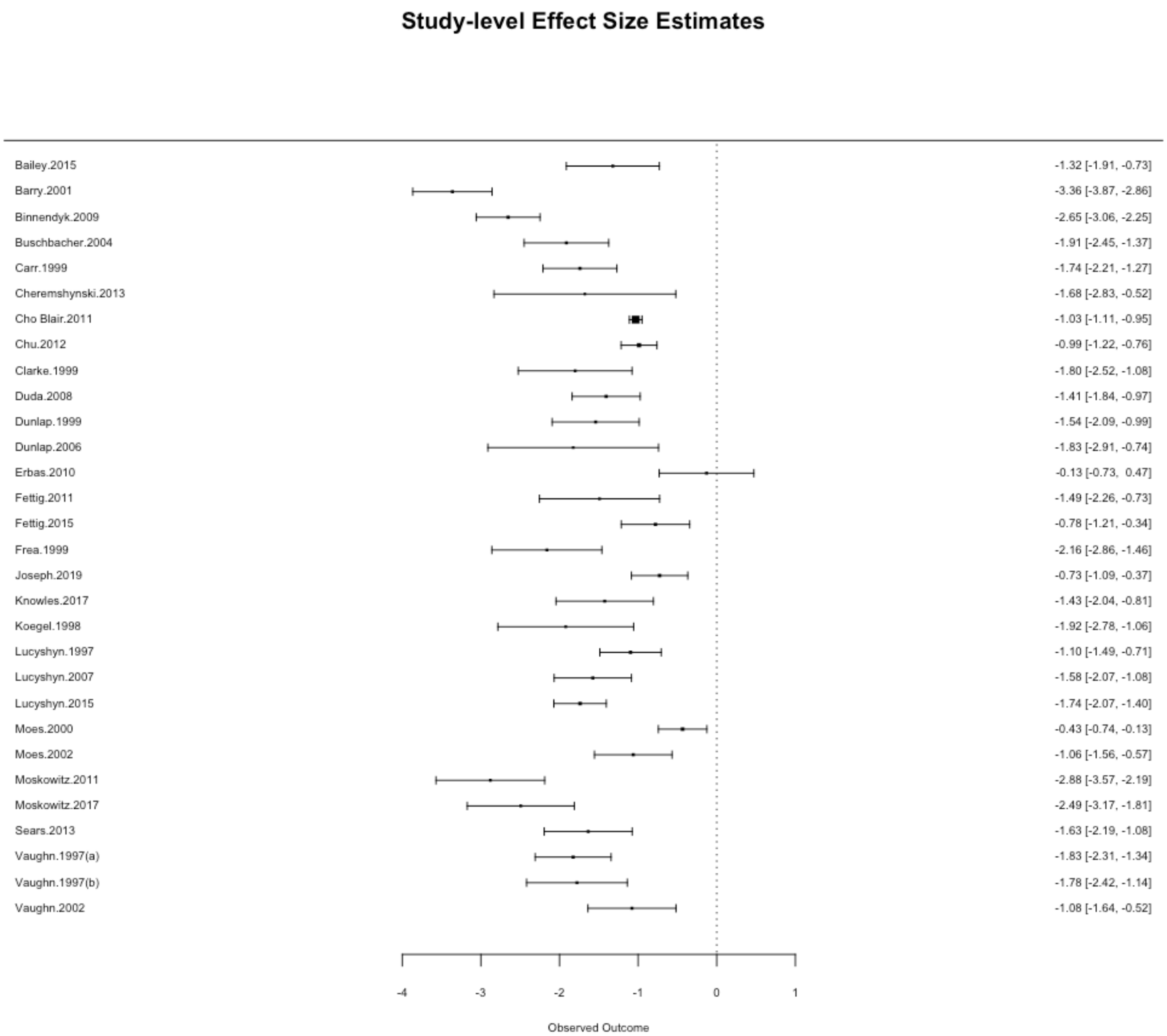
Figure 2 displays a forest plot of effect size estimates for each case and each dependent variable across 30 studies. Figure 3 and Figure 4 display forest plots of effect size estimates for each case. Since both positive-valence and negative-valence data were included, a transformation was required to conduct a meta-analysis. Since most studies included measures of problem behaviour, I transformed all data into the negative-valence form.

Using a robust random-effects model, the overall average effect size across all studies was –1.53 (95% CI [–1.79, –1.27], p < .001; see Table 6). After conversion to percentage data, this corresponded to an average decrease of problem behaviour by 78.3% (95% CI [–83.2%, –71.9%]) from baseline phases to intervention phases. Heterogeneity was detected (I^2 =23.8%), with the variance of study-level error (between-study variance) $\hat{\tau}^2 = 0.20$, and the variance of case-level error (within-study variance) $\widehat{\omega}^2 = 0.64$. I^2 suggests about 24% of the variance in effect sizes was due to between-study variation rather than sampling error.

Table 6 and Table 7 display the results for the moderator analysis. Results from the moderator analyses showed no significant differences of aggregate treatment effects across the grouping variables of gender, disability category, or WWC quality standards. A meta-regression also was conducted for the moderating variable of age. Table 7 shows that FCPBS was predicted to decrease the effect size of problem behaviour of focus children by 4.9% each year, all else held constant. However, this change across the age of children from early childhood to late adolescence was not statistically significant from a zero effect because of its associated standard error. Additionally, it was found that the age variable accounted for 8.1% of the between-study variation. For the other moderating variables, there was insufficient evidence to quantify the amount of between-study variation.

Figure 2

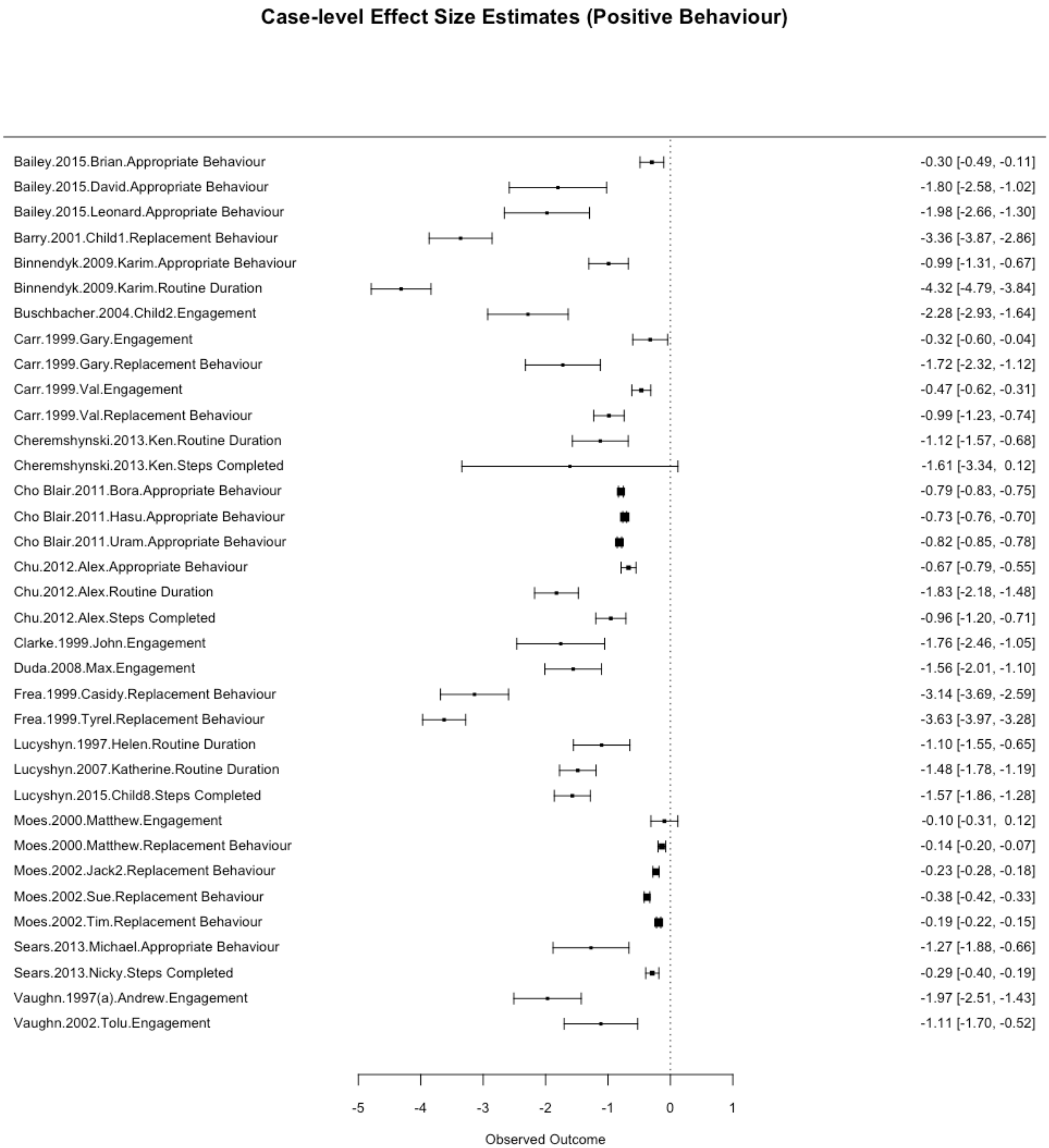
Forest Plot of Study-level Effect Size Estimates



Note. Studies are ordered alphabetically by the first author's last name.

Figure 3

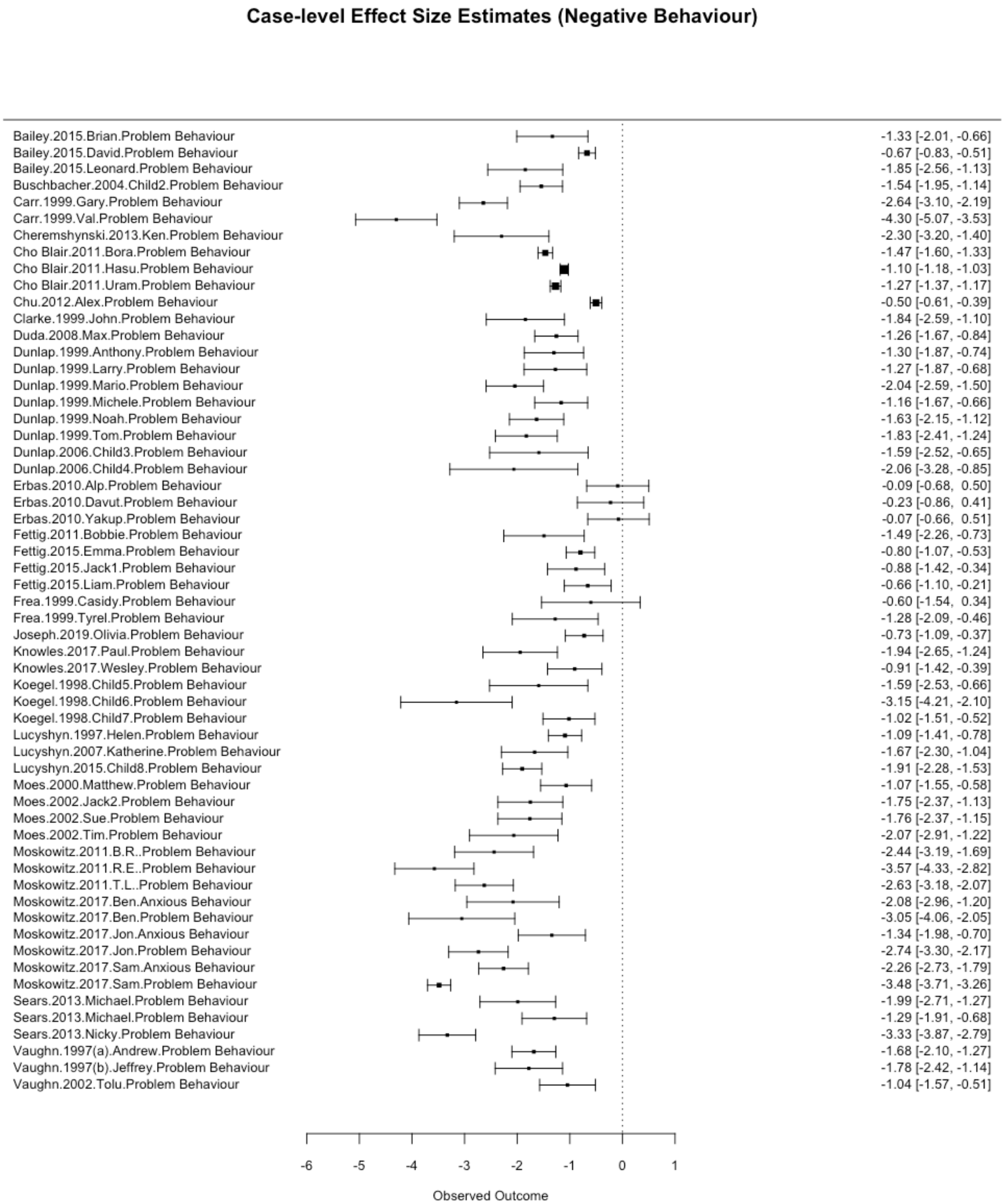
Forest Plot of Case-level Effect Size Estimates (Positive-Valence Behaviour)



Note. Studies are ordered alphabetically by the first author's last name.

Figure 4

Forest Plot of Case-level Effect Size Estimates (Negative-valence Behaviour)



Note. Studies are ordered alphabetically by the first author's last name.

Table 6

Summary Statistics for Overall Effect Sizes and for Effect Sizes of Binary Moderating Variables

Category	*N* of Cases	Estimate (%)	SE	Z-value	*p*-value	95% CI	95% Prediction Interval
Overall	56	–1.53 (–78.3%)	0.13	–12.18	<.0001	[–1.79, –1.27]	[–3.34, 0.29]
Gender							
Male	45	–1.55 (–78.7%)	0.14	–0.50	0.61	[–1.84, –1.25]	[–3.38, 0.29]
Female	10	–1.41 (–75.5%)	0.28			[–1.69, –1.13]	[–3.30, 0.49]
Disability							
ASD	32	–1.48 (–77.3%)	0.17	–0.40	0.69	[–1.85, –1.11]	[–3.34, 0.38]
Others	23	–1.58 (–79.4%)	0.19			[–1.98, –1.18]	[–3.45, 0.29]
WWC Standards							
Meet	25	–1.47 (–77.0%)	0.23	0.36	0.71	[–1.98, –0.95]	[–3.32, 0.39]
Does not meet	31	–1.57 (–79.3%)	0.14			[–1.88, –1.27]	[–3.42, 0.27]

Table 7

Summary Statistics for Effect Size of Moderating Variable of Age

Category	Estimate	Estimate (%)	SE	Z-value	*p*-value	95% CI	95% CI (%)
Intercept	–1.21	–70.2%	0.24	–4.94	<.0001	[–1.69, –0.73]	[–81.5%, –51.8%]
Age	–0.05	–4.9%	0.04	–1.48	0.1379	[–0.12, 0.02]	[–11.3%, 2.0%]

3.3 Rigor Evaluation Results

Rigor evaluation consists of WWC design standards evaluation and WWC evidence standards evaluation. Details are described below.

3.3.1 WWC design standards

After locating 30 eligible studies from the literature search, a two-stage rigor evaluation was conducted. Table 8 presents the results of applying WWC design standards to each of the 30 studies. In the first stage of rigor evaluation, there was one study (3.3%) meeting standards without reservations and 12 studies (40%) meeting standards with reservations. Among the 18 studies (56.7%) that did not meet standards, one study did not employ an experimental single case design and one study did not clearly state its IOA values. Six studies did not include at least three attempts to demonstrate an intervention effect at three different points. Six studies had insufficient (less than three) data points within each phase. Four studies did not fulfill additional requirements for a multiple probe design. Five studies had insufficient baseline and intervention phases corresponding to their single case design type (e.g., at least 6 phases for a multiple baseline design). Three studies employed an empirical case study design (i.e., A-B), which could not document a functional relation. There were some studies that did not meet design standards for more than one reason, and so totals do not add up to 18. Following this evaluation, the remaining 13 studies (43.3%) were submitted to the next stage of rigor evaluation.

Table 8

Summary of WWC Design Standards Evaluation

Study	Design	EXP	IV	DV	IOA	Probe	Phase	Data Points	Three Attempts	Conclusion
Bailey & Cho Blair (2015)	MB	Yes	Yes	Yes	Yes	N/A	≥6	4	Yes	Meet with R
Barry & Singer (2001)	MB	Yes	Yes	Yes	Yes	N/A	≥6	≤2	Yes	Does not meet
Binnendyk & Lucyshyn (2009)	ECS	Yes	Yes	Yes	Yes	Yes	3	4	No	Does not meet
Buschbacher, Fox, & Clarke (2004)	MB	Yes	Yes	Yes	Yes	Yes/R	≥6	4	Yes	Meet with R
Carr et al. (1999)	MB	Yes	Yes	Yes	Yes	No	≥6	4	Yes	Does not meet
Cheremshynski, Lucyshyn, & Olson (2013)	WR	Yes	Yes	Yes	Yes	N/A	4	3	Yes	Meet with R
Cho Blair et al. (2011)	MB	Yes	Yes	Yes	Yes	Yes/R	≥6	3	Yes	Meet with R
Chu (2012)	ECS	Yes	Yes	Yes	Yes	N/A	2	4	No	Does not meet
Clarke, Dunlap, & Vaughn (1999)	WR	Yes	Yes	Yes	Yes	N/A	4	≤2	Yes	Does not meet
Duda et al. (2008)	MB	Yes	Yes	Yes	Yes	No	≥6	3	Yes	Does not meet
Dunlap, Ester, Langhans, & Fox (2006)	MB	Yes	Yes	Yes	Yes	N/A	≥6	3	Yes	Meet with R
Dunlap & Fox (1999)	MB	No	Yes	Yes	Yes	Yes/R	≥6	3	No	Does not meet
Erbas (2010)	AT	Yes	Yes	Yes	Yes	N/A	≤2	≥6	Yes	Meet without R
Fettig & Ostrosky (2011)	MB	Yes	Yes	Yes	Yes	N/A	4	≥6	No	Does not meet
Fettig, Schultz, & Sreckovic (2015)	MB	Yes	Yes	Yes	Yes	N/A	≥6	3	Yes	Meet with R
Frea & Hepburn (1999)	MB	Yes	Yes	Yes	Yes	N/A	4	≥6	No	Does not meet
Joseph et al. (2019)	WR	Yes	Yes	Yes	Yes	N/A	4	≤2	Yes	Does not meet
Knowles et al. (2017)	MB	Yes	Yes	Yes	Yes	N/A	4	5	No	Does not meet

Table 8 Continued

L. K. Koegel, Steibel, & Koegel (1998)	MB	Yes	Yes	Yes	Yes	Yes/R	≥6	5	Yes	Meet with R
Lucyshyn et al. (2007)	MB	Yes	Yes	Yes	Yes	Yes/R	≥6	5	Yes	Meet with R
Lucyshyn et al. (2015)	MB	Yes	Yes	Yes	Yes	No	≥6	5	Yes	Does not meet
Lucyshyn, Albin, & Nixon (1997)	MB	Yes	Yes	Yes	Yes	Yes/R	≥6	4	Yes	Meet with R
Moes & Frea (2000)	ECS	Yes	Yes	Yes	Yes	N/A	2	2	No	Does not meet
Moes & Frea (2002)	MB	Yes	Yes	Yes	No	Yes/R	≥6	≤2	Yes	Does not meet
Moskowitz, Carr, & Durand (2011)	MB	Yes	Yes	Yes	Yes	N/A	≥6	3	Yes	Meet with R
Moskowitz et al. (2017)	MB	Yes	Yes	Yes	Yes	N/A	≥6	4	Yes	Meet with R
Sears et al. (2013)	MB	Yes	Yes	Yes	Yes	N/A	≥6	3	No	Does not meet
Vaughn, Clarke, et al. (1997)	MB	Yes	Yes	Yes	Yes	N/A	4	3	No	Does not meet
Vaughn, Dunlap, et al. (1997)	MB	Yes	Yes	Yes	Yes	No	≥6	≤2	Yes	Does not meet
Vaughn et al. (2002)	MB	Yes	Yes	Yes	Yes	Yes/R	≥6	3	Yes	Meet with R

Note. Design = What was the single case design employed (WR = Withdrawal/reversal, MB = Multiple baseline, AT = Alternating treatment, ECS = Empirical case study); EXP = Was the study an experimental single case design study; IV = Was the independent variable systematically manipulated by the researcher; DV = Were the outcome variables systematically measured by more than one assessor; IOA = Did the IOA in this study meet minimum acceptable values (at least 0.8 if measured by percentage and at least 0.6 if measured by Cohen's kappa); Probe = Did this study meet additional requirements (see in the method section) if it was a multiple probe design (N/A means not a multiple probe design); Phase = Minimum phases per case; Data Points = Minimum data points per phase; Three Attempts = Did this study have at least three attempts to demonstrate an intervention effect at three different points in time; Conclusion = "Meet WWC design standards with reservations" (Meet with R), or "Meet WWC design standards without reservations" (Meet without R), or "Does not meet WWC design standards" (Does not meet).

3.3.2 WWC evidence standards

Table 9 presents the results of applying WWC evidence standards for documenting a functional effect to each of the 13 studies. In the second stage of rigor evaluation, there were 11 studies (84.6%) rated as having strong evidence and one study (7.7%) rated as having moderate evidence. The latter study was rated as showing moderate evidence of a functional effect because only three of four baselines in the multiple baseline design documented a basic effect. The fourth baseline showed drift in the dependent variables across baseline and intervention phases. One study (7.7%) employing a multiple baseline design was rated as having no evidence because of drift in the dependent variable toward intervention levels in one of three baselines. Following the second stage of rigor evaluation, 12 studies (92.3%) documented evidence of an experimental effect of the FCPBS intervention. See the flowchart (Figure 5) below.

Table 9

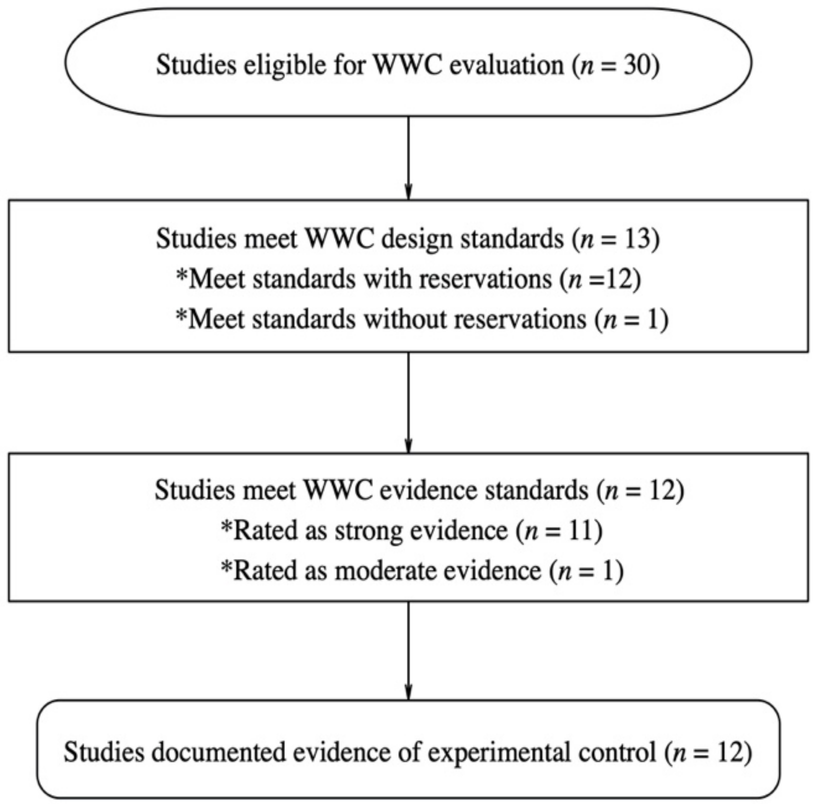
Summary of WWC Evidence Standards Evaluation

Study	Baseline data demonstrates clearly defined patterns that can assess effects of intervention	Visual analysis of level, trend, variability of intervention phase compared to baseline phase shows dramatic difference in direction of treatment, or clear separation of two treatments being compared	Overlap, immediacy of effect, and consistency of pattern are similar within same phases	Data in design across phases meet standard for documenting three demonstrations of effect at three different points in time (at least 4 clear separations in data paths for two treatments being compared)	Conclusion
Bailey & Cho Blair (2015)	Yes	Yes	Yes	Yes	Strong evidence
Buschbacher, Fox, & Clarke (2004)	Yes	Yes	Yes	Yes	Strong evidence
Cheremshynski, Lucyshyn, & Olson (2013)	Yes	Yes	Yes	Yes	Strong evidence
Cho Blair et al. (2011)	Yes	Yes	Yes	Yes	Strong evidence
Dunlap, Ester, Langhans, & Fox (2006)	Yes	Yes	Yes	Yes	Strong evidence
Erbas (2010)[a]	N/A	Yes	N/A	Yes	Strong evidence
Fettig, Schultz, & Sreckovic (2015)	Yes	Yes	Yes	Yes	Strong evidence
L. K. Koegel, Stiebel, & Koegel (1998)	No	No	No	No	No evidence
Lucyshyn et al. (2007)	Yes	Yes	Yes	Yes	Strong evidence
Lucyshyn, Albin, & Nixon (1997)	Yes	Yes	Yes	Yes	Moderate Evidence
Moskowitz, Carr, & Durand (2011)	Yes	Yes	Yes	Yes	Strong evidence
Moskowitz et al. (2017)	Yes	Yes	Yes	Yes	Strong evidence
Vaughn et al. (2002)	Yes	Yes	Yes	Yes	Strong evidence

Note. [a]There was no baseline phase in this study.

Figure 5

Flowchart of Rigor Evaluation



3.3.3 WWC evaluation of empirically supported treatment

Table 10 presents the results of the evaluation of the 12 studies that met WWC rigor standards in regard to whether these studies cumulatively met WWC criteria for an EST. As noted above, these criteria are as follows: a minimum of five single case research studies meeting WWC rigor standards conducted by three different research teams at three different locations without overlapping authors, with a total of at least 20 cases (5-3-20 criteria). A case is defined as one or more participants in one experimental single case design. Results show that the first criterion of 5 studies was exceeded, with a cumulative total of 12 studies surpassing this criterion. The second criterion of at least 3 different research teams in 3 different locations also was exceeded, with a cumulative total of six different research teams identified across seven geographic locations. The third criterion of at least 20 cases was not met, given the WWC definition of a case as an experimental single case research design that meets rigor standards. Table 10 shows that the cumulative number of cases across the 12 FCPBS studies was 15, thus falling short by 5 cases.

Table 10

Summary of 5-3-20 Criteria Features for Studies Meeting WWC Design and Evidence Standards

Year	Last Name of All Authors	Cumulative Research Teams	Location of Research	Cumulative Locations	Cumulative *N* of Cases
1997	Lucyshyn, Albin, & Nixon	1	OR, United States	1	1
2002	Vaughn, Wilson, & Dunlap	2	FL, United States	2	2
2004	Buschbacher, Fox, & Clarke	3	FL, United States	2	3
2006	Dunlap, Ester, Langhans, & Fox	3	FL, United States	2	5
2007	Lucyshyn, Albin, Horner, Jane Mann, James Mann, & Wadsworth	3	OR, United States	2	6
2010	Erbas	4	Turkey	3	9
2011	Cho Blair, Lee, Cho, & Dunlap	4	South Korea	4	10
2011	Moskowitz, Carr, & Durand	5	NY, United States	5	11
2013	Cheremshynski, Lucyshyn, & Olson	5	BC, Canada	5	12
2015	Bailey & Cho Blair	5	FL, United States	5	13
2015	Fettig, Schultz, & Sreckovic	6	East Central Region of United States	6	14
2017	Moskowitz, Walsh, Mulder, McLaughlin, Hajcak, Carr, & Zarcone	6	NY, United States	7	15

Chapter 4: Discussion

The primary purpose of this study was to conduct a meta-analysis of Family-Centered Positive Behaviour Support (FCPBS) across the extant literature comprised of single case design studies with families of children with developmental disabilities and problem behaviour in home and community settings. The secondary purpose of this study was to: (a) evaluate the scientific rigor of each included study based on What Works Clearinghouse (WWC) design and evidence standards for single case research; and (b) determine whether or not FCPBS met WWC 5-3-20 criteria for an empirically supported treatment (EST). The third purpose was to determine whether there were effect size differences across moderating variables of age, gender, disability category and WWC quality evaluation. A moderator analysis was conducted for this purpose. A summary and interpretation of results in light of these purposes are presented below.

4.1 Research Question 1: What is the overall effect of FCPBS across children and families?

Overall, results of the meta-analysis across 30 studies showed that FCPBS was effective at reducing problem behaviour of children with developmental disabilities and increasing their appropriate behaviour and/or positive engagement. Results revealed that the overall effect size of the 30 single case design studies was -1.53, which represents a large effect size. When converted to a percentage of change, FCPBS improved child behaviour in target routines from baseline to intervention phases by 78.3%. The magnitude of this overall average effect size estimate is similar to that of meta-analyses of PBS in school settings at the tertiary level using the LRR estimate (Common et al., 2017; Pustejovsky, 2018).

This overall effect size and percentage of change was associated with several common features of each study implementing FCPBS. First, researchers conducted a functional assessment that informed the design of a multicomponent behaviour support plan. Second, families were collaborative partners in the development of the plan, which increased the likelihood that behaviour support plans were both technically sound in terms of the laws of behaviour and contextually appropriate given the ecology and/or culture of the family. Third, the multicomponent behaviour support plan for each child emphasized preventive, teaching and positive reinforcement strategies as well as consequence strategies that removed or diminished functional reinforcers for problem behaviour. Fourth, the child parent(s) implemented the behaviour support plans in family routines in the home and/or community that were valued by the family but problematic due to child problem behaviour. Fifth, researchers provided implementation support to families in the form of coaching either directly in the target routines or indirectly during meetings. Sixth, researchers conducted ongoing evaluation of child progress in target routines, and adjusted plan components and implementation supports based on the unfolding data in the single case repeated measures designs.

Although follow-up data were not included in the meta-analysis, which is common practice, I also conducted a descriptive analysis of follow-up data cross the 30 FCPBS studies because these data provide important evidence of the durability of FCPBS outcomes. Although evidence of a large effect is important, if these changes during an intervention phase do not maintain over time after the interventionists are no longer providing implementation support, it would be difficult to argue that FCPBS is a truly effective practice. Across the 30 studies, the duration of maintenance ranged from 2 weeks to 7 years. Follow-up phase results documented relatively stable trends with high levels of adaptive behaviour and low levels of problem behaviour. For example, some studies showed improvement in child behaviour at maximum levels (e.g., 100% of routine steps completed) for adaptive behaviour (e.g., Binnendyk &

Lucyshyn, 2009; Buschbacher et al., 2004; Fettig & Ostrosky, 2011) or at minimum levels (e.g., zero percentage of intervals) for problem behaviour (Carr et al., 1999; Lucyshyn et al., 2007; Lucyshyn et al., 2015; Moes & Frea, 2002). These results offer additional descriptive support for the durability of the large overall effect size estimate.

One caution that needs to be noted is that meta-analysis findings are specific to children with developmental disabilities, as studies of FCPBS with typically developing children and adults with developmental disabilities living in group home settings did not meet inclusion criteria. In addition, although two studies of FCPBS included an adult with a developmental disability, the sample size was too small to make broad inferences about the effectiveness of FCPBS with this population. In summary, overall findings of the meta-analysis of FCPBS provide strong and reliable evidence that FCPBS is effective at addressing the behavioural support needs of families raising children with developmental disabilities who engage in problem behaviour in family contexts.

4.2 Research Question 2: What is the overall rigor of FCPBS studies to date, according to the WWC design standards and evidence standards?

Evaluation of WWC design standards indicated that 43.3% of included FCPBS studies (i.e., 13 of 30) met WWC design standards. This percentage of adherence is similar to the percentage reported by Maggin et al. (2017) in their meta-analysis of group contingencies (39%, 81 of 229 cases), a secondary level of prevention. In contrast, Walker et al. (2018), in a meta-analysis of PBS in school settings at the tertiary level, reported that 73% (33 out of 45) of the included studies met WWC standards with or without reservations, much higher than the percentage found in this study. Similarly, Carr (2016) in another meta-analysis of PBS in school settings at the tertiary level, reported that 80% (35 out of 44) of the included studies met WWC standards with or without reservations

The main reasons for FCPBS studies not meeting WWC design standards were: (a) insufficient phases necessary to document a functional effect (e.g., multiple baseline design across two family routines); (b) insufficient number of data points in a phase necessary to document either stability or trend (e.g., baseline phase with only 2 data points); and (c) insufficient immediacy of observation sessions just prior to intervention phase in multiple probe designs, which diminishes documentation of the immediacy of effect from baseline to intervention phases.

Many of the FCPBS studies that did not meet WWC design standards were conducted before 2013, the year in which WWC single case design standards were first published. Given this, a logical question is whether single case research studies of FCPBS conducted *before* WWC design standards were published were more likely to fall short of design standards compared to FCPBS studies conducted after the WWC standards were published. To answer this question, I examined Table 8. Table 8 shows that 22 FCPBS studies were conducted before 2013 and 8 FCPBS studies were conducted after 2013. Eight of the 22 studies published before 2013 met design standards (36.4%), while 4 of the 8 studies (50%) published after 2013 met design standards. This difference (13.6%) offers modest, tentative evidence in support of the hypothesis that studies published before 2013 were more likely to be evaluated as not meeting design standards. However, more FCPBS studies published after 2013 are needed to have a sample size large enough to test the hypothesis that studies published after 2013 are more likely to meet design standards.

As noted above, 13 of 30 FCPBS studies met WWC *design* standards. When I evaluated WWC *evidence* standards all but one of the 13 studies met this standard. Twelve of the studies demonstrated moderate to strong evidence of a functional effect when evaluating level, trend, variability, immediacy of effect, overlap, and consistency of data in similar phases. It is noteworthy that the studies that met design standards for the most part also met evidence standards. Overall, 40.0% of the included studies of FCPBS demonstrated evidence of methodological rigor and causal inference.

4.3 Research Question 3: Given the FCPBS studies to date that meet WWC design and evidence standards, do they collectively meet WWC criteria for an EST?

Analysis of the 5-3-20 criteria indicated that the 12 studies meeting WWC design and evidence standards satisfied the first criterion of a minimum of five rigorous single case research studies. The analysis also showed that the 12 studies exceeded the second criterion of at least three different research teams across three different locations. However, the 12 studies documented a total of 15 cases, falling short of the third criterion of at least 20 cases (i.e., participants within an experimental single case research design).

It is noteworthy that the WWC 5-3-20 criteria do not include additional criteria that allow for the classification of a practice based on the extent to which the three criterion have been met in the extant literature. In contrast, the Council for Exceptional Children (CEC) offers additional criteria for determining the extent to which a practice approaches classification as an evidence-based practice. CEC *Standards for Evidence-Based Practices in Special Education* established criteria for five classifications: evidence-based practice, potential evidence-based practice, mixed effects, insufficient evidence, or negative effects (Cook et al., 2014). A practice with at least five methodologically sound single case design studies demonstrating positive effects is considered either an evidence-based practice (equivalent to an EST) or a potential evidence-based practice, depending on the total number of participants across the studies. If the number of participants is equal to or greater than 20, the practice is classified as an evidence-based practice. Otherwise, the practice is classified as a potential evidence-based practice. Similarly, the National Autism Center (NAC, 2015) has developed a three-level classification system consisting of established practice, emerging practice and unestablished practice.

The merit of such classification systems is that they inform the research community where along a continuum of methodologically sound studies does a particular practice stand in relation to the ultimate aim of establishing the practice as an EST. For FCPBS researchers, with two of the three 5-3-20 criterion met, and five more cases needed to meet the third criterion of 20 cases, it would be reasonable, given the CEC or NAC classification systems, to classify FCPBS as a potential or emerging evidence-based practice or EST.

4.4 Research Question 4: Do moderator variables such as age, gender, disability category and WWC rigor standards provide evidence of systematic variation in the magnitude of intervention effects?

The moderator analysis that was conducted across the 30 FCPBS studies for the moderator variables of age, gender, disability category and WWC rigor standards did not show significant differences in effect sizes based on the focus individuals' age, gender, disability category or WWC rigor standards. In regard to child age and FCPBS impact on problem behaviour, although meta-regression showed that the magnitude of effect sizes decreased as focus children were older, this change was not statistically significant. This is in contrast to a meta-analysis of studies of the antecedent strategy of *offering choices by* Shogren et al. (2004). The authors generated mean effect sizes for participants 4-7 years old and 8-50 years old. Results showed that the age 4-7 group had a significantly larger effect size than the age 8-50 group. These findings suggest that *offering choices* is more powerful among individuals at younger ages. Although the moderator analysis of age in this study was in the same direction as that of the Shogren study, the lack of statistical significance requires this inference for FCPBS to be viewed with caution.

The moderator analysis for gender, disability and WWC rigor standards showed little difference between effect sizes for these moderator variables, with gender, disability and WWC rigor standards evidencing differences of 3.2%, 2.1%, 2.3% respectively. Although statistical analyses of differences between moderator subgroups were not statistically significant, application of statistical tests should still be treated with caution for the following reasons. First, sample sizes were small, with *N* of cases for most subgroups less than 30. Second, the data in the binary subgroups were skewed, and thus did not meet the assumption of a normal distribution, required for statistical analysis of differences between two groups. Third, the data in the binary subgroups were correlated, and thus did not meet the assumption of independent observations, also required for statistical analysis of differences between two groups. Skewness and autocorrelation were endemic to the employment of single case research designs. Normally distributed data across baseline and intervention phases would indicate the absence of a basic effect. However, the 30 FCPBS studies examined all documented basic effects at the case level. Autocorrelation also was associated with single case research designs. Behavioural data collected from focus individuals in each study were not independent of each other.

4.5 Limitations

First, in the selection of studies that met WWC design standards, I excluded one design standard; that is, that IOA data were gathered in a minimum of 20% observation sessions in each phase of the single case design study (Kratochwill et al., 2013). Although in most of the 30 FCPBS studies, authors reported that IOA was gathered in more than 20% of total observation sessions, they often did not state the percentage of IOA allocation across phases, or reported measuring IOA in less than 20% of observations in some phases (see, for example, Barry & Singer, 2001; Dunlap & Fox, 1999; Vaughn, Clarke, & Dunlap, 1997). Authors typically reported that IOA was gathered in 30% or more of observation sessions, a standard for IOA data collection established decades ago for single case design methods (see Kazdin, 1982; Wolery & Gast, 1982). For example, Sears et al. (2013) reported that "… fifty percent of the sessions were assessed for IOA" (pp. 1008). Moskowitz et al. (2017) reported that IOA was gathered in "… 67% of the videotaped sessions distributed across baseline and intervention sessions" (pp. 10). In both examples, although IOA measurement exceeded the historical standard of 30% or more of sessions, it is unclear whether a minimum of 20% of observations were gathered in each study phase. I made this adjustment to WWC design standards for two reasons. If I were to include this standard, almost all FCPBS studies would have been excluded from the meta-analysis. This seemed particularly unfair to the 26 studies published between 1996 and 2013, before WWC standards were established and available to inform researchers of the newly established standards for rigor. Second, I believed it was reasonable to assume that if studies reported IOA as being over 80% for more than 30% of the data points gathered, the data gathered in these studies were more likely to be reliable than unreliable.

A second limitation is that although data extraction software such as WebPlotDigitizer has high reliability and validity and their results approach the true value of extracted data (Moeyaert, Maggin, & Verkuilen, 2016), errors in data extraction nevertheless may occur. One possible source of error is in the formatting of single case design graphs. For example, in some graphs the geometric center of the data point (for data presented as circles) represents the true value on the *y*-axis (i.e., the level of the dependent variable). In other graphs the mid-point of the bottom line of the data point (for data presented as triangles or squares) represents the true value on the *y*-axis. However, this information was not provided in the studies examined, and so coders were not aware of these possible differences in geometric center when extracting data from the graphs. Another example of potential error due to formatting can occur in single case design graphs that are printed in a small area on the page but with a large scale for the *y*-axis (e.g., 100% divided by 5 units). These potential data extraction errors were prevented or minimized for studies that report the mean and range for each case and phase. For these studies, I compared the minimum value, maximum value and mean between the extracted data and the information provided by authors. I then corrected any errors that were present in the extracted data. Data extraction errors were typically uniform and represented an approximately 1%-3% departure from the estimated true values of the study's dataset.

Third, autocorrelation of data in single case design studies has made applications of statistical approaches problematic (Pustejovsky, 2018). Even though statisticians have worked on minimizing the impact of autocorrelation, current statistical approaches cannot entirely solve this issue and researchers have to accept that results derived by these approaches may have some unpredictable errors. Although autocorrelation has little to no impact on effect size estimates, it can diminish the statistical power necessary to detect differences and model fits. In single case design research, autocorrelation can be found at the study level and case level. At the case level, within a phase, data of the next observation session is dependent on data from the previous observation session. For example, in this meta-analysis of FCBPS studies, Lucyshyn et al. (1997), Erbas (2010), and Cho Blair et al. (2011) all showed clear serial dependency within their baseline phases and intervention phases. Although problematic, correlation of this kind is endemic to single case design research and cannot be overcome. At the study level, autocorrelation may be seen across studies with one or more of the same authors. For example, in this meta-analysis of FCBPS studies, Vaughn, Wilson, & Dunlap (2002), Buschbacher, Fox, & Clarke (2004), and Dunlap, Ester, Langhans, & Fox (2006) shared one or more authors. Overlapping authorship indicates that these studies were conducted by at least one member of the same research team and at the same institution. This may have created dependency across these studies in assessment, plan design, data collection and decision-making processes. Autocorrelation of this kind can be overcome by non-overlapping authorship of single case design studies. In this meta-analysis, the presence of six independent research groups suggests that study-level autocorrelation was moderated and thus did not strongly influence the meta-analytic results, in contrast to case-level autocorrelation.

Fourth, results showed mild heterogeneity. Heterogeneity can be generated from between-study and within-study factors. Between-study factors may include different units of measurement of dependent variables, different target routines, different single case research designs and different data patterns. For example, Vaughn et al. (2002), Buschbacher et al. (2004), and Duda et al. (2008) conducted multiple baseline designs *across three settings* with one participant in which three A-B effects were documented. In contrast, Koegel et al. (1998), Dunlap et al. (1999), and Bailey and Cho Blair (2015) conducted multiple baseline designs *across three participants* in which each participant showed one A-B effect. As a result, the former cases had more chances to produce an expected change than the latter cases. Another example is the study by Cho Blair et al. (2011) in which stable data patterns created less variation compared to the study by Joseph et al. (2019) in which fluctuating data patterns created more variation. These difference in variability contributed to the heterogeneity of the complete dataset. Within-study factors also can contribute to the heterogeneity. For example, in this study, there were differences in age (i.e., 2 years old to 17 years old), gender (i.e., male and female), disability diagnosis (e.g., autism, intellectual disability, Fragile-X Syndrome) and family involvement (e.g., mother; father; mother and father). If small effects exist between these factors, they can accumulate to create heterogeneity when aggregating effect sizes of these cases. Though heterogeneity is not the most optimal condition for the conduct of a meta-analysis, heterogeneity is nearly inevitable in single case design studies and thus is considered acceptable (Pustejovsky, 2018).

4.6 Future Directions

The study and its findings suggest two considerations for future meta-analytic studies of FCPBS and four considerations for future FCPBS intervention research. First, future meta-analytic studies of FCPBS should consider comparing effect size calculation metrics. As noted in the introduction, overlap-based metrics such as percentage of nonoverlapping data (PND) also can be employed to calculate effect sizes. It would be worth exploring whether effect sizes generated from overlap-based metrics (e.g., PND) and from mean-based metrics (e.g., LRR) result in different meta-analytic results when examining FCPBS studies.

Second, future meta-analytic studies of FCPBS should consider employing different approaches to evaluating methodological rigor. For example, the Council for Exceptional Children (CEC) published a set of standards for evaluating the methodological rigor of single case design research studies that differs from WWC quality criteria (Cook et al., 2014). Although WWC and CEC single case design standards have much in common, there are differences in design requirements. For example, CEC requires that studies report implementation fidelity data; the extent to which the intervention was implemented as intended. WWC criteria currently does not include this design standard. On the other hand, while WWC requires a minimum of three data point in each phase of baseline and intervention conditions, CEC allows for fewer data points in a phase if the reason is based on ethical and social validity considerations. In the meta-analytic literature that has included an evaluation of the rigor of PBS studies, a few authors have applied both WWC and CEC standards in their evaluation. For example, Mitchell, Hatton, and Lewis (2018) applied WWC and CEC standards to evaluate five group design studies of SWPBS. Results showed that four studies met WWC standards while no studies met CEC quality indicators. In a future meta-analysis of FCPBS, applying both WWC standards and CEC quality indicators would offer a more complete understanding of the effectiveness of FCPBS, and provide additional direction for the future conduct of FCPBS intervention studies in regard to both methodological and clinical rigor.

The findings of this study also offer recommendations for future FCPBS research that employs single case research designs. First, future studies should prevent common methodological shortcomings that were identified in this study and adhere to the full set of WWC design and evidence standards. Design standards that future FCPBS studies should adhere to include: (a) describing the percentage of IOA allocation across phases and baselines and measure IOA in a minimum of 20% of observations in all phases and baselines; (b) employing experimental single case designs that have at least the minimum number of phases necessary to document a functional relation (e.g., four phases for withdrawal/reversal designs; six phases for multiple baseline designs); (c) gathering at least three data points in each phase to establish the minimum number necessary to document stability or trend; and (d) ensuring that in multiple probe design studies at least one baseline phase data point is gathered just before transitioning to the intervention phase. An evidence standard to which FCPBS studies should adhere is ensuring that drift toward treatment levels during a baseline phase either do not occur or compromise the documentation of a functional effect.

Second, given the increasing ethnic diversity of families in the U.S. and Canada, future FCPBS research should inform readers of the ethnicity of study participants. Across the 30 FCPBS studies identified, only 27% of the studies reported the ethnic background of participants. Given that the reported ethnicity of participating families was primarily Caucasian and Asian, information on ethnicity may contribute to future FCPBS researchers actively recruiting a broader diversity of families, thus contributing to an understanding of the generalizability of the FCPBS approach.

Third, given the nine features of a positive behaviour support approach as described in the seminal article by Carr et al. (2002), FCPBS studies will do well to integrate additional features that are less represented in the extant FCPBS literature. These include measurement of child and family quality of life, measurement of social validity as perceived by family members, and longer follow-up data (i.e., > 6 months). This latter feature would offer stronger evidence of PBS practitioners' commitment to a lifespan perspective as documented by the durability of child and family outcomes.

Fourth, given 12 studies, 6 research teams across 7 locations, and 15 cases to date that meet WWC design and evidence standards, as noted above FCPBS may be classified as an emerging evidence-based practice (National Autism Center, 2015) or potential evidence-based practice (Cook et al., 2014). This finding should encourage FCPBS researchers to continue to investigate the efficacy of FCPBS with particular attention to ensuring that the single case design methodology employed meets WWC design and evidence standards.

Summary

This study represents the first meta-analysis that examined the effectiveness of FCPBS across families of children with developmental disabilities and problem behaviour. The overall treatment effect generated from the meta-analysis provides initial evidence of the effectiveness of FCPBS for reducing problem behaviour and increasing adaptive behaviour among children with developmental disabilities in family routines in the home and community. An evaluation of WWC rigor standards showed that 40% of included studies (12 of 30) met quality standards. An evaluation of the 5-3-20 criteria for an empirically supported treatment showed that although the 5 single case study criterion and 3 independent research team criterion were met, the 20 cases criterion was not met, with only 15 of 20 cases meeting this criterion across the 12 studies with sufficient rigor. A moderator analysis across age, gender, disability and WWC rigor standards did not show statistically significant differences within each moderating variable. Study limitations included moderation of one WWC design standard, data extraction errors, autocorrelated data, and large heterogeneity. Future directions for research and practice are recommended that will strengthen the rigor of FCPBS research employing single case design methods, and advance FCPBS toward achieving classification as an established EST.

Bibliography

References marked with an asterisk indicate studies included in the final meta-analysis.

Allen, R., & Petr, C. (1995). Family-centered service delivery. *The University of Kansas: Beach Center on Families and Disability*.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis 1. *Journal of applied behavior analysis*, *1*(1), 91-97.

*Bailey, K. M., & Blair, K. S. C. (2015). Feasibility and potential efficacy of the family-centered Prevent-Teach-Reinforce model with families of children with developmental disorders. *Research in developmental disabilities*, *47*, 218-233. doi:10.1016/j.ridd.2015.09.019

Bambara, L. M. (2005). Overview of the behavior support process. In Bambara, L. M., & Kern, L. (Eds), *Individualized supports for students with problem behaviors: designing positive behavior plans* (pp. 47-70). New York, NY: Guilford Press.

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change* (2nd ed.). New York, NY: Pergamon Press.

*Barry, L. M., & Singer, G. H. S. (2001). A family in crisis: Replacing the aggressive behavior of a child with autism toward an infant sibling. *Journal of Positive Behavior Interventions*, *3*(1), 28-38. doi:10.1177/109830070100300105

Barton, E. E., Pustejovsky, J. E., Maggin, D. M., & Reichow, B. (2017). Technology-aided instruction and intervention for students with ASD: A meta-analysis using novel methods of estimating effect sizes for single-case research. *Remedial and Special Education*, *38*(6), 371-386. doi:10.1177/0741932517729508

*Binnendyk, L., & Lucyshyn, J. M. (2009). A family-centered positive behavior support approach to the amelioration of food refusal behavior: An empirical case study. *Journal of Positive Behavior Interventions*, *11*(1), 47-62. doi:10.1177/1098300708318965

*Blair, K. C., Lee, I., Cho, S., & Dunlap, G. (2011). Positive behavior support through Family–School collaboration for young children with autism. *Topics in Early Childhood Special Education*, *31*(1), 22-36. doi:10.1177/0271121410377510

Boden, L. J., Jolivette, K., & Alberto, P. A. (2018). The effects of check-in, check-up, check-out for students with moderate intellectual disability during on- and off-site vocational training. *Journal of Classroom Interaction*, *53*(1), 4.

Brown, F., Anderson, J., & De Pry, R. L. (Ed.). (2014). *Individual positive behavior supports: A standards-based guide to practices in school and community settings*. Baltimore, MD: Paul H. Brookes Publishing.

*Buschbacher, P., Fox, L., & Clarke, S. (2004). Recapturing desired family routines: A parent-professional behavioral collaboration. *Research and Practice for Persons with Severe Disabilities*, *29*(1), 25-39. doi:10.2511/rpsd.29.1.25

Carr, M. E. (2016). Self‐management of challenging behaviours associated with autism spectrum disorder: A Meta‐Analysis. *Australian Psychologist*, *51*(4), 316-333. doi:10.1111/ap.12227

Carr, E. G. (2007). The expanding vision of positive behavior support: Research perspectives on happiness, helpfulness, and hopefulness. *Journal of Positive Behavior Interventions*, *9*, 3–14. doi:10.1177/10983007070090010201

Carr, E. G., Dunlap, G., Horner, R. H., Koegel, R. L., Turnbull, A. P., Sailor, W., . . . Fox, L. (2002). Positive behavior support: Evolution of an applied science. *Journal of Positive Behavior Interventions*, *4*, 4–16. doi:10.1177/109830070200400102

*Carr, E. G., Levin, L., McConnachie, G., Carlson, J. I., Kemp, D. C., Smith, C. E., & McLaughlin, D. M. (1999). Comprehensive multisituational intervention for problem behavior in the community: Long-term maintenance and social validation. *Journal of Positive Behavior Interventions*, *1*(1), 5-25. doi:10.1177/109830079900100103

Carroll, D. W. (2013). *Families of children with developmental disabilities: Understanding stress and opportunities for growth*. Washington, D.C.: American Psychological Association.

Centers for Disease Control and Prevention, & Autism and Developmental Disabilities Monitoring Network Surveillance Year 2002 Principal Investigators. (2007). Prevalence of autism spectrum disorders — autism and developmental disabilities monitoring network, 14 sites, united states, 2002. *MMWR. Surveillance Summaries*, *56*(SS-1), 12-28.

*Cheremshynski, C., Lucyshyn, J. M., & Olson, D. L. (2013). Implementation of a culturally appropriate positive behavior support plan with a Japanese mother of a child with autism: An experimental and qualitative analysis. *Journal of Positive Behavior Interventions*, *15*(4), 242-253. doi:10.1177/1098300712459904

*Chu, S. Y. (2012). A family-centered positive behavior support approach to intervene food refusal behavior for a child with autism. *Journal of Special Education*, *36*, 37-60.

*Clarke, S., Dunlap, G., & Vaughn, B. (1999). Family-centered, assessment-based intervention to improve behavior during an early morning routine. *Journal of Positive Behavior Interventions*, *1*(4), 235-241. doi:10.1177/109830079900100406

Common, E. A., Lane, K. L., Pustejovsky, J. E., Johnson, A. H., & Johl, L. E. (2017). Functional Assessment–Based interventions for students with or at-risk for high-incidence disabilities: Field testing single-case synthesis methods. *Remedial and Special Education*, *38*(6), 331-352. doi:10.1177/0741932517693320

Cook, B., Buysse, V., Klingner, J., Landrum, T., McWilliam, R., Tankersley, M., & Test, D. (2014). Council for exceptional children: Standards for evidence-based practices in special education. *TEACHING Exceptional Children*, *46*(6), 206-212. doi:10.1177/0040059914531389

Drevon, D. D., Hixson, M. D., Wyse, R. D., & Rigney, A. M. (2019). A meta ‐ analytic review of the evidence for check ‐ in check ‐ out. *Psychology in the Schools*, *56*(3), 393-412. doi:10.1002/pits.22195

*Duda, M. A., Clarke, S., Fox, L., & Dunlap, G. (2008). Implementation of positive behavior support with a sibling set in a home environment. *Journal of Early Intervention*, *30*(3), 213-236. doi:10.1177/1053815108319124

Duda, M. A., Dunlap, G., Fox, L., Lentini, R., & Clarke, S. (2004). An experimental evaluation of positive behavior support in a community preschool program. *Topics in Early Childhood Special Education*, *24*(3), 143–155. https://doi.org/10.1177/02711214040240030201

Dufrene, B. A., Doggett, R. A., Henington, C., & Watson, T. S. (2007). Functional assessment and intervention for disruptive classroom behaviors in preschool and head start classrooms. *Journal of Behavioral Education*, *16*(4), 368-388. doi:10.1007/s10864-007-9048-8

*Dunlap, G., Ester, T., Langhans, S., & Fox, L. (2006). Functional communication training with toddlers in home environments. *Journal of Early Intervention*, *28*(2), 81-96. doi:10.1177/105381510602800201

*Dunlap, G., & Fox, L. (1999). A demonstration of behavioral support for young children with autism. *Journal of Positive Behavior Interventions*, *1*(2), 77-87. doi:10.1177/109830079900100202

Dunlap, G., Iovannone, R., Wilson, K. J., Kincaid, D. K., & Strain, P. (2010). Prevent-teach-reinforce: A standardized model of school-based behavioral intervention. *Journal of Positive Behavior Interventions*, *12*(1), 9-22. doi:10.1177/1098300708330880

Dunlap, G., Sailor, W., Horner, R. H., & Sugai, G. (2009). Overview and history of positive behavior support. In Sailor, W., Dunlap, G., Sugai, G., & Horner, R. (Eds.), *Handbook of positive behavior support* (pp. 3-16). New York, NY: Springer.

Dunst, C. J. (2002). Family-centered practices: Birth through high school. *The Journal of Special Education*, *36*(3), 141-149. doi:10.1177/00224669020360030401

Dunst, C. J., & Espe-Sherwindt, M. (2016). Family-centered practices in early childhood intervention. In Reichow, B., Boyd, B. A., Barton, E. E., & Odom, S. L. (Eds.), *Handbook of early childhood special education* (pp. 37-55). Cham, Switzerland: Springer International Publishing.

Dunst, C. J., Trivette, C. M., & Deal, A. G. (1988). Enabling and empowering families. In Dunst, C. J., Trivette, C. M., & Deal, A. G. (Eds.), *Strengthening and supporting families: Methods, strategies, and practices*, (Vol. 1, pp. 2–11). Cambridge, MA: Brookline Books.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*, 405–432. doi:10.1111/j.1467-8624.2010.01564.x

*Erbas, D. (2010). A collaborative approach to implement positive behavior support plans for children with problem behaviors: A comparison of consultation versus consultation and feedback approach. *Education and Training in Autism and Developmental Disabilities*, *45*(1), 94-106.

Faith M.S., Allison D.B., Gorman B.S. (1996). Meta-analysis of single-case research. In R.D. Franklin, D.B. Allison, B.S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 245-277). Mahwah, NJ: Lawrence Erlbaum Associates.

Feldman, M. A., Condillac, R. A., Tough, S., Hunt, S., & Griffiths, D. (2002). Effectiveness of community positive behavioral intervention for persons with developmental disabilities and severe behavior disorders. *Behavior Therapy*, *33*(3), 377-398. doi:10.1016/S0005-7894(02)80034-X

*Fettig, A., & Ostrosky, M. M. (2011). Collaborating with parents in reducing children's challenging behaviors: Linking functional assessment to intervention. *Child Development Research*, *2011*, 1-10. doi:10.1155/2011/835941

*Fettig, A., Schultz, T. R., & Sreckovic, M. A. (2015). Effects of coaching on the implementation of functional assessment–based parent intervention in reducing challenging behaviors. *Journal of Positive Behavior Interventions*, *17*(3), 170-180. doi:10.1177/1098300714564164

Floyd, F., & Gallagher, E. (1997). Parental Stress, care demands, and use of support services for school-age children with disabilities and behavior problems. *Family Relations*, *46*(4), 359-371. doi:10.2307/585096

Fox, L., Vaughn, B. J., Dunlap, G., & Bucy, M. (1997). Parent-professional partnership in behavioral support: A qualitative analysis of one family's experience. *Journal of the Association for Persons with Severe Handicaps*, *22*(4), 198-207. doi:10.1177/154079699702200403

Fox, L., Vaughn, B. J., Wyatte, M. L., & Dunlap, G. (2002). “We can't expect other people to understand”: Family perspectives on problem behavior. *Exceptional Children*, *68*(4), 437-450. doi:10.1177/001440290206800402

*Frea, W. D., & Hepburn, S. L. (1999). Teaching parents of children with autism to perform functional assessments to plan interventions for extremely disruptive behaviors. *Journal of Positive Behavior Interventions*, *1*(2), 112-122. doi:10.1177/109830079900100205

Freeman, J., Simonsen, B., McCoach, D. B., Sugai, G., Lombardi, A., & Horner, R. (2016). Relationship between school-wide positive behavior interventions and supports and academic, attendance, and behavior outcomes in high schools. *Journal of Positive Behavior Interventions*, *18*(1), 41-51. doi:10.1177/1098300715580992

Gage, N. A., Lewis, T. J., & Stichter, J. P. (2012). Functional behavioral assessment-based interventions for students with or at risk for emotional and/or behavioral disorders in school: A hierarchical linear modeling meta-analysis. *Behavioral Disorders*, *37*(2), 55-77. doi:10.1177/019874291203700201

Gage, N. A., Whitford, D. K., & Katsiyannis, A. (2018). *A review of schoolwide positive behavior interventions and supports as a framework for reducing disciplinary exclusions.* Los Angeles, CA: SAGE Publications. doi:10.1177/0022466918767847

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*(10), 3-8. doi:10.2307/1174772

Goh, A. E., & Bambara, L. M. (2012). Individualized positive behavior support in school settings: A meta-analysis. *Remedial and Special Education*, *33*(5), 271-286. doi:10.1177/0741932510383990

Ha, J. H., Hong, J., Seltzer, M. M., & Greenberg, J. S. (2008). Age and gender differences in the well-being of midlife and aging parents with children with mental health or developmental problems: Report of a national study. *Journal of Health and Social Behavior*, *49*, 301-316. doi:10.1177/002214650804900305

Hawken, L. S., & Horner, R. H. (2003). Evaluation of a targeted intervention within a schoolwide system of behavior support. *Journal of Behavioral Education*, *12*(3), 225-240. doi:10.1023/A:1025512411930

Hawken, L. S., Wayman, G., & Stokes, K. (2020). Engaging Students in Appropriate Social Behavior Using Check-In, Check-Out (CICO). In *Student Engagement* (pp. 185-199). Cham, Switzerland: Springer International Publishing.

Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*(2), 107-128. doi:10.3102/10769986006002107

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, *3*(3), 224-239. doi:10.1002/jrsm.1052

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods*, *4*(4), 324-341. doi:10.1002/jrsm.1086

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta‐regression with dependent effect size estimates. *Research synthesis methods*, *1*(1), 39-65. doi:10.1002/jrsm.5

Hitchcock, J. H., Horner, R. H., Kratochwill, T. R., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2014). The What Works Clearinghouse single-case design pilot standards: Who will guard the guards? *Remedial and Special Education*, *35*(3), 145-152. doi:10.1177/0741932513518979

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional children*, *71*(2), 165-179. doi:10.1177/001440290507100203

Horner, R. H., Dunlap, G., Koegel, R. L., Carr, E. G., Sailor, W., Anderson, J., ... & O'Neill, R. E. (1990). Toward a technology of “nonaversive” behavioral support. *Journal of the Association for Persons with Severe Handicaps*, *15*(3), 125-132. doi:10.1177/154079699001500301

Horner, R. H., Kincaid, D., Sugai, G., Lewis, T., Eber, L., Barrett, S., . . . Johnson, N. (2014). Scaling up school-wide positive behavioral interventions and supports: Experiences of seven states with documented success. *Journal of Positive Behavior Interventions*, *16*(4), 197-208. doi:10.1177/1098300713503685

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions*, *11*(3), 133-144. doi:10.1177/1098300709332067

Horner, R. H., Sugai, G., Todd, A. W., & Lewis-Palmer, T. (2005). Schoolwide positive behavior support. In L. M. Bambara & L. Kern (Eds.), *Individualized supports for students with problem behaviors: Designing positive behavior plans* (pp. 359–390). New York, NY: Guilford Press.

Horner, R. H., Swaminathan, H., Sugai, G., & Smolkowski, K. (2012). Considerations for the systematic analysis and use of single-case research. *Education and Treatment of Children*, *35*(2), 269-290. doi:10.1353/etc.2012.0011

*Joseph, J. D., Strain, P. S., & Dunlap, G. (2019). An experimental analysis of prevent teach reinforce for families (PTR-F). *Topics in Early Childhood Special Education*. doi:10.1177/0271121419847760

Kincaid, D., & Dunlap, G. (2015). *The definition of positive behavior support: Moving toward greater precision.* Paper presented at the 12th International Conference of the Association for Positive Behavior Support, Boston, MA.

*Knowles, C., Blakely, A., Hansen, S., & Machalicek, W. (2017). Parents with intellectual disabilities experiencing challenging child routines: A pilot study using embedded self‐determination practices. *Journal of Applied Research in Intellectual Disabilities*, *30*(3), 433-444. doi:10.1111/jar.12303

Koegel, L. K., Koegel, R. L., & Dunlap, G. (1996). *Positive behavioral support: Including people with difficult behavior in the community*. Baltimore, MD: P.H. Brookes Publishing.

*Koegel, L. K., Stiebel, D., & Koegel, R. L. (1998). Reducing aggression in children with autism toward infant or toddler siblings. *Journal of the Association for Persons with Severe Handicaps*, *23*(2), 111-118. doi:10.2511/rpsd.23.2.111

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. doi:10.1177/0741932512452794

Lane, K. L. , Menzies, H. M. , Oakes, W. P. , Lambert, W. , Cox, M. , & Hankins, K. (2012). A validation of the student risk screening scale for internalizing and externalizing behaviors: Patterns in rural and urban elementary schools. *Behavioral Disorders*, *37*, 244–270. doi:10.1177/019874291203700405

Ledford. J. R, & Gast, D. L. (2018). Replication. In Ledford, J. R., Gast, D. L. (Ed.), *Single case research methodology: Applications in special education and behavioral sciences* (3rd ed., pp. 77-96). New York, NY: Routledge.

Ledford, J. R., Lane, J. D., & Tate, R. (2018). Evaluating quality and rigor in single case research. In Ledford, J. R., & Gast, D. L. (Ed.), *Single case research methodology: Applications in special education and behavioral sciences* (3rd ed., pp. 365-392). New York, NY: Routledge.

*Lucyshyn, J. M., Albin, R. W., Horner, R. H., Mann, J. C., Mann, J. A., & Wadsworth, G. (2007). Family implementation of positive behavior support for a child with autism: Longitudinal, single-case, experimental, and descriptive replication and extension. *Journal of Positive Behavior Interventions*, *9*(3), 131-150. doi:10.1177/10983007070090030201

*Lucyshyn, J. M., Albin, R. W., & Nixon, C. D. (1997). Embedding comprehensive behavioral support in family ecology: An experimental, single-case analysis. *Journal of Consulting and Clinical Psychology*, *65*(2), 241-251. doi:10.1037/0022-006X.65.2.241

Lucyshyn, J.M., Dunlap, G., & Freeman, R. (2015). A historical perspective on the evolution of positive behavior support as a science-based discipline. In F. Brown, J. Anderson, & R. DePry (Eds.), *Individual positive behavior supports: A standards-based guide to practices in school and community-based settings* (pp. 27–54). Baltimore, MD: Paul H. Brookes Publishing.

Lucyshyn, J. M., & Fossett, B. (2015). *Twenty years of PBS research with families: Lesson learned and future directions.* Paper presented at the 12th Annual International Conference on Positive Behavior Support, Boston, MA.

*Lucyshyn, J. M., Fossett, B., Bakeman, R., Cheremshynski, C., Miller, L., Lohrmann, S., . . . Irvin, L. K. (2015). Transforming Parent–Child interaction in family routines: Longitudinal analysis with families of children with developmental disabilities. *Journal of Child and Family Studies*, *24*(12), 3526-3541. doi:10.1007/s10826-015-0154-2

Lucyshyn, J. M., Horner, R. H., Dunlap, G., Albin, R. W., Ben, K. R. (2002). Positive behavior support with families. In Lucyshyn, J. M., Dunlap, G., Albin, R. W. (Eds.), *Families and positive behavior support: Addressing problem behavior in family contexts* (pp. 3–44). Baltimore, MD: Paul H. Brookes Publishing.

Lucyshyn, J. M., Miller, L. D., Cheremshynski, C., Lohrmann, S., & Zumbo, B. D. (2018). Transforming coercive processes in family routines: Family functioning outcomes for families of children with developmental disabilities. *Journal of Child and Family Studies*, *27*(9), 2844-2861. doi:10.1007/s10826-018-1113-5

Maggin, D. M., Pustejovsky, J. E., & Johnson, A. H. (2017). A meta-analysis of school-based group contingency interventions for students with challenging behavior: An update. *Remedial and Special Education*, *38*(6), 353-370. doi:10.1177/0741932517716900

Matyas, T. A., & Greenwood, K. M. (1996). Serial dependency in single-case time series. In R. D. Franklin, Allison, D. B., & Gorman, B. S. (Eds.), *Design and Analysis of Single-Case Research* (pp. 215-243). Mahwah, NJ: Lawrence Erlbaum Associates.

Marquis, S., McGrail, K., Hayes, M., & Tasker, S. (2018). Estimating the prevalence of children who have a developmental disability and live in the province of British Columbia. *Journal on Developmental Disabilities*, *23*(3), 46-56.

McIntosh, K., Campbell, A. L., Carter, D. R., & Rossetto Dickey, C. (2009). Differential effects of a tier two behavior intervention based on function of problem behavior. *Journal of Positive Behavior Interventions*, *11*(2), 82-93. doi:10.1177/1098300708319127McIntosh, K., Mercer, S. H., Nese, R. N., Strickland-Cohen, M. K., & Hoselton, R. (2016). Predictors of sustained implementation of school-wide positive behavioral interventions and supports. *Journal of Positive Behavior Interventions*, *18*(4), 209-218. doi:10.1177/1098300715599737

Meadan, H., Ayvazo, S., & Ostrosky, M. M. (2016). The ABCs of challenging behavior: Understanding basic concepts. *Young Exceptional Children*, *19*(1), 3-15. doi:10.1177/1096250614523969

Miller, L. M., Dufrene, B. A., Sterling, H. E., Olmi, D. J., & Bachmayer, E. (2015). The effects of check-In/Check-out on problem behavior and academic engagement in elementary school students. *Journal of Positive Behavior Interventions*, *17*(1), 28-38. doi:10.1177/1098300713517141

Mitchell, B. S., Hatton, H., & Lewis, T. J. (2018). An examination of the evidence-base of school-wide positive behavior interventions and supports through two quality appraisal processes. *Journal of Positive Behavior Interventions*, *20*(4), 239-250. doi:10.1177/1098300718768217

*Moes, D. R., & Frea, W. D. (2000). Using family context to inform intervention planning for the treatment of a child with autism. *Journal of Positive Behavior Interventions*, *2*(1), 40-46. doi:10.1177/109830070000200106

*Moes, D. R., & Frea, W. D. (2002). Contextualized behavioral support in early intervention for children with autism and their families. *Journal of Autism and Developmental Disorders*, *32*(6), 519-533. doi:10.1023/A:1021298729297

Moeyaert, M., Ferron, J. M., Beretvas, S. N., & Van den Noortgate, W. (2014). From a single-level analysis to a multilevel analysis of single-case experimental designs. *Journal of School Psychology*, *52*(2), 191-211. doi:10.1016/j.jsp.2013.11.003

Moeyaert, M., Maggin, D., & Verkuilen, J. (2016). Reliability, validity, and usability of data extraction programs for single-case research designs. *Behavior Modification*, *40*(6), 874-900. doi:10.1177/0145445516645763

Moeyaert, M., Zimmerman, K. N., & Ledlord, J. R. (2018). Synthesis and meta-analysis of single case research. In Ledford, J. R., & Gast, D. L. (Ed.), *Single case research methodology: Applications in special education and behavioral sciences* (3rd ed., pp. 393-416). New York, NY: Routledge.

*Moskowitz, L. J., Carr, E. G., & Durand, V. M. (2011). Behavioral intervention for problem behavior in children with fragile X syndrome. *American Journal on Intellectual and Developmental Disabilities*, *116*(6), 457-478. doi:10.1352/1944-7558-116.6.457

*Moskowitz, L. J., Walsh, C. E., Mulder, E., McLaughlin, D. M., Hajcak, G., Carr, E. G., & Zarcone, J. R. (2017). Intervention for anxiety and problem behavior in children with autism spectrum disorder and intellectual disability. *Journal of Autism and Developmental Disorders*, *47*(12), 3930-3948. doi:10.1007/s10803-017-3070-z

Muir, K., & Strnadova, I. (2014). Whose responsibility? Resilience in families of children with developmental disabilities. *Disability & Society*, *29*(6), 922-937. doi:10.1080/09687599.2014.886555

National Autism Center. (2015). *Findings and Conclusions: National Standards Project, Phase 2*. Retrieved from http://www.autismdiagnostics.com/assets/Resources/NSP2.pdf

Ncube, B. L., Perry, A., & Weiss, J. A. (2018). The quality of life of children with severe developmental disabilities: Quality of life. *Journal of Intellectual Disability Research*, *62*(3), 237-244. doi:10.1111/jir.12460

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, *35*(4), 303-322. doi:10.1177/0145445511399147

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, *42*(2), 284-299. doi:10.1016/j.beth.2010.08.006

Petrou, S., Yiu, H. H., & Kwon, J. (2019). Economic consequences of preterm birth: A systematic review of the recent literature (2009–2017). *Archives of Disease in Childhood*, *104*(5), 456-465. doi:10.1136/archdischild-2018-315778

Poppes, P., Van der Putten, A. J. J., & Vlaskamp, C. (2010). Frequency and severity of challenging behaviour in people with profound intellectual and multiple disabilities. *Research in Developmental Disabilities*, *31*(6), 1269-1275. doi:10.1016/j.ridd.2010.07.017

Pustejovsky, J. E. (2015). Measurement-comparable effect sizes for single-case studies of free-operant behavior. *Psychological Methods*, *20*(3), 342-359. doi:10.1037/met0000019

Pustejovsky, J. E. (2017). clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections. *R package version 0.2.3*.

Pustejovsky, J. E. (2018). Using response ratios for meta-analyzing single-case designs with behavioral outcomes. *Journal of School Psychology*, *68*, 99-112. doi:10.1016/j.jsp.2018.02.003

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Roberts, C., Mazzucchelli, T., Studman, L., & Sanders, M. R. (2006). Behavioral family intervention for children with developmental disabilities and behavioral problems. *Journal of Clinical Child and Adolescent Psychology*, *35*, 180–193. doi:10.1207/s15374424jccp3502_2

Rohatgi, A. (2015). *WebPlotDigitizer user manual version 3.4*. Retrieved from https://automeris.io/WebPlotDigitizer/userManual.pdf

Ruef, M. B., Turnbull, A. P., Turnbull, H. R., & Poston, D. (1999). Perspectives of five stakeholder groups: Challenging behavior of individuals with mental retardation and/or autism. *Journal of Positive Behavior Interventions*, *1*(1), 43-58. doi:10.1177/109830079900100106

Russell Carter, D., & Horner, R. H. (2009). Adding function-based behavioral support to first step to success: Integrating individualized and manualized practices. *Journal of Positive Behavior Interventions*, *11*(1), 22-34. doi:10.1177/1098300708319125

Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: Issues and applications. *Behavior modification*, *22*(3), 221-242. doi:10.1177/01454455980223001

*Sears, K. M., Blair, K. C., Iovannone, R., & Crosland, K. (2013). Using the prevent-teach-reinforce model with families of young children with ASD. *Journal of Autism and Developmental Disorders*, *43*(5), 1005-1016. doi:10.1007/s10803-012-1646-1

Shadish, W. R. (2014). Analysis and meta-analysis of single-case designs: An introduction. *Journal of School Psychology*, *52*(2), 109-122. doi:10.1016/j.jsp.2013.11.009

Shogren, K. A., Faggella-Luby, M. N., Bae, S. J., & Wehmeyer, M. L. (2004). The effect of choice-making as an intervention for problem behavior: A meta-analysis. *Journal of Positive Behavior Interventions*, *6*(4), 228-237. doi:10.1177/10983007040060040401

Singer, G. H. S., Goldberg-Hamblin, S. E., Peckman-Hardin, K. D., Barry, L., & Santarelli, G. E. (2002). Toward a synthesis of family support practices and positive behavior support. In Lucyshyn, J. M., Dunlap, G., Albin, R. W. (Eds.), *Families and positive behavior support: Addressing problem change in family contexts* (pp. 155–183). Baltimore, MD: Paul H. Brookes Publishing.

Solomon, B. G., Klein, S. A., Hintze, J. M., Cressey, J. M., & Peller, S. L. (2012). A meta‐analysis of school‐wide positive behavior support: An exploratory study using single‐case synthesis. *Psychology in the Schools*, *49*(2), 105-121. doi:10.1002/pits.20625

Sugai, G., & Horner, R. H. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Child & Family Behavior Therapy*, *24*(1-2), 23-50. doi:10.1300/J019v24n01_03

Sugai, G., & Horner, R. H. (2009). Responsiveness-to-intervention and school-wide positive behavior supports: Integration of multi-tiered system approaches. *Exceptionality*, *17*(4), 223-237. doi:10.1080/09362830903235375

Sugai, G., & Horner, R. H. (2020). Sustaining and scaling positive behavioral interventions and supports: Implementation drivers, outcomes, and considerations. *Exceptional Children*, *86*(2), 120-136. doi:10.1177/0014402919855331

Swaminathan, H., Rogers, H. J., & Horner, R. H. (2014). An effect size measure and Bayesian analysis of single-case designs. *Journal of School Psychology*, *52*(2), 213-230. doi:10.1016/j.jsp.2013.12.002

Turnbull, A. P., & Ruef, M. B. (1996). Family perspectives on problem behavior. *Mental Retardation*, *34*, 280–293.

Turnbull, A. P., & Ruef, M. B. (1997). Family perspectives on inclusive lifestyles for people with problem behavior. *Exceptional Children*, *63*(2), 211.

Turton, A. M., Umbreit, J., & Mathur, S. R. (2011). Systematic function-based intervention for adolescents with emotional and behavioral disorders in an alternative setting: Broadening the context. *Behavioral Disorders*, *36*(2), 117-128. doi:10.1177/019874291103600203

Van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Evidence-Based Communication Assessment and Intervention*, *2*(3), 142-151. doi:10.1080/17489530802505362

*Vaughn, B. J., Clarke, S., & Dunlap, G. (1997). Assessment-based intervention for severe behavior problems in a natural family context. *Journal of Applied Behavior Analysis*, *30*(4), 713-716. doi:10.1901/jaba.1997.30-713

*Vaughn, B. J., Dunlap, G., Fox, L., Clarke, S., & Bucy, M. (1997). Parent-professional partnership in behavioral support: A case study of community-based intervention. *Journal of the Association for Persons with Severe Handicaps*, *22*(4), 186-197. doi:10.1177/154079699702200402

*Vaughn, B. J., Wilson, D., & Dunlap, G. (2002). Family-centered intervention to resolve problem behaviors in a fast-food restaurant. *Journal of Positive Behavior Interventions*, *4*(1), 38-45. doi:10.1177/109830070200400107

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1-48.

Walker, V. L., Chung, Y., & Bonnet, L. K. (2018). Function-based intervention in inclusive school settings: A meta-analysis. *Journal of Positive Behavior Interventions*, *20*(4), 203-216. doi:10.1177/1098300717718350

Wang, M. & Singer, G. H. S. (2016). *Supporting families of children with developmental disabilities: Evidence-based and emerging practices*. New York, NY: Oxford University Press.

Wolery, M. (2013). A commentary: Single-case design technical document of the What Works Clearinghouse. *Remedial and Special Education*, *34*(1), 39-43. doi:10.1177/0741932512468038

XYit (3.1.4) [Computer Software]. Pennington, NJ: Geomatix.

Appendix

WWC Methodology Standards for Single Case Design Studies

<table>
<tr><th colspan="3">Methodology Standards</th></tr>
<tr><td colspan="2">Independent variable systematically manipulated</td><td>YES NO</td></tr>
<tr><td colspan="2">Each outcome variable measured systematically over time by more than one assessor</td><td>YES NO</td></tr>
<tr><td colspan="2">IOA agreement sessions for 20% of data points in each condition</td><td>YES NO</td></tr>
<tr><td colspan="2">Agreement score of ≥ 80%; Cohen’s Kappa of ≥ 0.60</td><td>YES NO</td></tr>
<tr><td colspan="2">At least 3 attempts to demonstrate intervention effect at 3 different points in time.</td><td>YES NO</td></tr>
<tr><th colspan="3">Choose either A, B, C under the design that was used in the study</th></tr>
<tr><td colspan="2">Withdrawal/reversal design
A. Meets design standards: at least 4 phases with 5 data points per phase
B. Meets design standards with reservations: at least 4 phases with 3 to 4 data points per phase
C. Does not meet design standards: less than 4 phases, and/or less than 3 data points per phase</td><td>A
B
C</td></tr>
<tr><td colspan="2">Multiple baseline design
A. Meets design standards: at least 6 phases & 5 data points per phase
B. Meets design standards with reservations: at least 6 phases with 3-4 data points per phase
C. Does not meet design standards: less than 5 phases and/or less than 3 data points per phase</td><td>A
B
C</td></tr>
<tr><td colspan="2">Alternating treatment design
A. Meets design standards: at least 5 repetitions of the alternating sequence of the interventions, and no more than 2 repetitions of an intervention prior to alternating with the other intervention(s).
B. Meets design standards with reservations: at least 4 repetitions of the alternating sequence of the interventions with no more than 2 repetitions of an intervention prior to alternating with the other intervention(s).
C. Does not meet design standards: less than 4 repetitions of alternating sequence of interventions and/or more than 2 repetitions of an intervention prior to alternating with the other intervention(s).</td><td>A
B
C</td></tr>
<tr><th colspan="3">Additional criterion ONLY applicable to multiple probe design (a special case of multiple baselines)</th></tr>
<tr><td colspan="2">Initial pre-intervention sessions must overlap vertically.
A. Meets SCD Standards: Within first 3 sessions, design must include 3 consecutive data points for each case.
B. Meets SCD Standards with Reservations: Within first 3 sessions, design must include at least 1 probe data point for each case
C. Does not meet design standards: Within first 3 sessions, no data point for each case.</td><td>A
B
C</td></tr>
<tr><td colspan="2">Probe data points must be available just prior to introducing independent variable.
A. Meets SCD Standards Without Reservations: Within 3 sessions just prior to introducing independent variable, design must include 3 consecutive data points for each case.
B. Meets SCD Standards with Reservations: Within 3 sessions just prior to introducing independent variable, design includes at least 1 probe data point for each case.
C. Does not meet design standards: Within 3 sessions just prior to introducing independent variable, design includes no data point for each case.</td><td>A
B
C</td></tr>
<tr><th colspan="3">OVERALL EVALUATION: Select one based on your evaluation of standards. This study:</th></tr>
<tr><td>meets standards without reservations
(All YES, All A)</td><td>meets standards with reservations
(All YES, No C, at least one B)</td><td>does NOT meet standards
(At least one NO or C)</td></tr>
<tr><td></td><td></td><td></td></tr>
</table>