



University of Kentucky
UKnowledge

Theses and Dissertations--Early Childhood,
Special Education, and Counselor Education

Early Childhood, Special Education, and
Counselor Education

2015

A SYSTEMATIC REVIEW OF USING WEIGHTED VESTS WITH INDIVIDUALS WITH AUTISM SPECTRUM DISORDER

Christopher J. Taylor
University of Kentucky, cjtaylor@uky.edu

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

Recommended Citation

Taylor, Christopher J., "A SYSTEMATIC REVIEW OF USING WEIGHTED VESTS WITH INDIVIDUALS WITH AUTISM SPECTRUM DISORDER" (2015). *Theses and Dissertations--Early Childhood, Special Education, and Counselor Education*. 20.

https://uknowledge.uky.edu/edsrc_etds/20

This Master's Thesis is brought to you for free and open access by the Early Childhood, Special Education, and Counselor Education at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Early Childhood, Special Education, and Counselor Education by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Christopher J. Taylor, Student

Dr. Amy D. Spriggs, Major Professor

Dr. Ralph Crystal, Director of Graduate Studies

A SYSTEMATIC REVIEW OF USING
WEIGHTED VESTS WITH INDIVIDUALS WITH AUTISM SPECTRUM DISORDER

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in the
College of Education at the University of Kentucky

By

Christopher J. Taylor

Bardstown, Kentucky

Director: Dr. Amy D. Spriggs, Professor of Special Education

Lexington, Kentucky

2015

Copyright © Christopher J. Taylor, 2015

ABSTRACT OF THESIS

A SYSTEMATIC REVIEW OF USING WEIGHTED VESTS WITH INDIVIDUALS WITH AUTISM SPECTRUM DISORDER

The purpose of the study was to evaluate the current literature on the use of weighted vests with individuals with autism spectrum disorder. A literature review using the What Works Clearinghouse Standards was conducted. The results of the review show that the use of weighed vests with individuals with autism spectrum disorder is not an evidence-based practice.

KEYWORDS: Autism spectrum disorder, ASD, weighted vests, evidence-based practices

Christopher J. Taylor

November 8, 2015

A SYSTEMATIC REVIEW OF USING
WEIGHTED VESTS WITH INDIVIDUALS WITH AUTISM SPECTRUM

By

Christopher J. Taylor

Dr. Amy D. Spriggs
Co-Director of Thesis

Dr. Ralph Crystal
Director of Graduate Studies

November 8, 2015

TABLE OF CONTENTS

LIST OF FIGURES.....	v
LIST OF TABLES.....	vi
Section 1: Introduction.....	1
Autism Spectrum Disorders.....	7
Weighted Vests.....	8
Section 2: Purpose.....	11
Section 3: Method.....	12
Search Procedures.....	12
Inclusion criteria.....	12
What Works Clearinghouse (2010) indicators.....	13
Descriptive analysis.....	13
Determination of an evidence base for using weighted vest.....	13
Interrater reliability on quality indicators and study outcomes.....	14
Section 4: Study Characteristics.....	16
Quality of the Single Subject Studies.....	16
Participants.....	16
Settings.....	17
Target Behavior.....	17
Dependent Variable.....	18
Data Collection.....	18
Weight of Vests.....	18
Single Subject Research Designs.....	19

Section 5: Study Results and Outcomes.....	26
Student Outcomes.....	26
Determination of Evidence-based Practice.....	26
Section 6: Discussion.....	28
Implications for Future Practices.....	29
Limitations and Conclusions.....	29
References.....	32
Vita.....	35

LIST OF FIGURES

Figure 1, Flowchart for Evaluation of SCRD Studies.....	15
---	----

LIST OF TABLES

Table 1, Evaluation of Studies Using WWC Guidelines.....	20
Table 2, Descriptive Information from Studies Rated as MDS and MDwR.....	26

Section 1: Introduction

Evaluating the quality of interventions used in school settings is garnering more attention from district administrators and researchers due to laws that emphasize accountability, like the No Child Left Behind Act of 2001, and the evolution of the field of special education. Initially consumers of research relied on narrative descriptions of interventions to evaluate their quality, but as educational research shifted to more quantitative models the need for a method of evaluating studies emerged (Wong et al., 2015).

Quality interventions are discovered and validated with quantitative research. Two popular methodologies in the behavioral sciences, like psychology and education, are group and single case research designs (SCRD). Special education is a field that focuses on the behavior of individuals, and due to the heterogeneity of its population, SCRDR research is uniquely suited to examine issues in its field. Single case designs provide rigorous investigations of independent variables through the measurement of dependent variables across a variety of conditions. Specifically, the effects the independent variable are examined through repeated and systematic applications with concurrent data collection on the dependent variable (Kratochwill et al., 2010).

The three identifying features of research using SCRDR methodology are the use of an individual case, repeated measurement of a dependent variable across multiple conditions, and designs that allow individuals to serve as their own control. An individual case can be comprised of a single individual or a group of individuals. The conditions under which a dependent variable is measured includes environments where the independent variable is not present (i.e., baseline) and environments where the

independent variable is present (i.e., intervention). Individuals or groups can serve as their own control by measuring the dependent variable in the absence and presence of the independent variable (Kratochwill et al., 2010).

All of the designs that use single case methodology rely on systematic manipulation of the independent variable. In SCRD research this can be achieved in three ways. The independent variable can be introduced and withdrawn, introduced iteratively, or introduced in staggered stages. When the independent variable is introduced and withdrawn it occurs within an ABAB design or a derivation of an ABAB design (e.g., ABABAB). When the independent variable is introduced iteratively, it occurs an alternating treatment design or a derivation of an alternating treatment design (e.g., adapted alternating treatment design, parallel treatment design). When the independent variable is introduced in staggered stages, it occurs in multiple baseline or multiple probe designs (Kratochwill et al., 2010).

A unique feature of SCRD research is its use of visual analysis to interpret data. Representing data visually is an effective practice because it makes information accessible to people without specialized training. Typically, when representing data visually, researchers will present time related data (e.g., number of sessions, dates) in the abscissa (i.e., the x axis) and performance data (e.g., frequency count of target behavior) in the ordinate (i.e., the y axis). Once the data have been placed into a graph, it can be analyzed to identify its level, trend, and variability. Level represents the mean value of data in a specific phase; trend represents the slope or best-fit line of the data; and variability refers to the stability of the data. Each of these factors are analyzed closely when interpreting data (Kratochwill et al., 2010).

There is not a universally accepted definition of what constitutes an evidence-based practice in SCRD; however, a number of researchers have proposed different standards to help consumers evaluate the quality of published research. Of those researchers, criteria proposed Reichow et al. (2008), Horner et al. (2005), the National Professional Development Center for Autism Spectrum Disorder and the What Works Clearinghouse (WWC) have emerged as popular choices for individuals interested in examining the research on a specific intervention.

In 2008, Reichow, Volkmar, and Cicchetti published two sets of methods, one for group research and one for SCRD research, to evaluate practices used for children with autism spectrum disorders (ASD). Reichow and colleagues recommended using three tools when evaluating practices used with individuals with ASD: (1) a rubric to evaluate the rigor of individual studies, (2) guidelines to generate research report strength, and (3) criteria to determine if the evidence on a single intervention is sufficient to be determined as an evidence-based practice. The tools used for examining group research will not be discussed in detail because no studies included in this review used group designs. For a detailed description of the tools recommended by Reichow et al. (2008), see the original article.

The rubric for evaluating the experimental rigor of studies looks at primary and secondary quality indicators. Primary quality indicators are factors of research that are necessary for demonstrating validity and secondary quality indicators are factors of research that are important but not necessary for demonstrating validity. The primary quality indicators rubric rates studies as “high quality”, “acceptable quality”, or “unacceptable quality.” The secondary quality indicators rubric examines the presence or

absence or specific factors. The guidelines to generate a research report uses the information gleaned from the primary and secondary quality indicator rubrics to rate a study's research strength as "strong," "adequate," or "weak." The criteria for evaluating if an intervention is an evidence-based practice categorizes practices as "established evidence-based practices" or "promising evidence-based practices". An established evidence-based practice is one with at least five SCRD studies with a "strong" rating that were conducted by three different research teams in three different locations with at least 15 participants or at least 10 SCRD studies with an "adequate" rating conducted by three different research teams in three different locations with at least 30 participants. For SCRD studies, a promising evidence-based practice has at least three studies with "adequate" research strength conducted by two different research teams in two different locations with at least nine participants.

In 2005, Horner and colleagues published a set of criteria to help researchers and practitioners evaluate the quality of single-case research studies and determine if a practice was evidence-based. In their paper, the authors described quality indicators for "acceptable" studies. The criteria for "acceptable" studies focused on the description of participants and settings, dependent variable, independent variable, baseline condition, and internal, external, and social validity. In addition to outlining the criteria for "acceptable" studies, the paper also recommended that studies examining an intervention collectively include at least five studies completed in at least three unique geographic regions by three different research teams, including at minimum of 20 participants to be considered an evidence-based practice (Horner et al., 2005).

To aid in the use of evidence-based practices, the Office of Special Education Programs in the US Department of Education provided funding for the National Professional Development Center on Autism Spectrum Disorder (NPDC). The organization provides resources to educators, administrators and researchers through comprehensive professional development. One of the resources provided by the center is a list of evidence-based practices. In its most recent literature review, the NPDC used the guidelines described by Wong et al. (2015). In their article, the authors wrote that the panel used a combination of the standards recommended by Horner et al. (2005) and select criteria from the standards proposed by WWC to evaluate the quality of studies; although the authors did not cite which WWC standards were used. The NPDC determined if a practice was evidenced-based by using the following criteria: five high quality SCRD studies conducted by at least three research teams with at least 20 participants or two high quality quasi-experimental or experimental studies conducted by at least two research teams or a combination of three SCRD and randomized and one quasi-experimental/randomized study that was conducted by at least three different research teams (Wong et al., 2015).

Amid differing opinions of what constitutes an evidence-based practice, many scholars and consumers of single-case research are looking to WWC to find clarity about methodological debates in the behavioral sciences. In 2002, the Institute for Education Sciences created the WWC to conduct independent examinations of psychology- and education-based interventions (Kratochwill et al., 2013).

When evaluating a practice with the WWC standards, examiners must first evaluate if a study uses a SCRD. A study is determined to use a SCRD if (1) it uses single

participant or cluster of participants, like a classroom; (2) the participants serve as their own control group (i.e., data on the dependent variable is collected before and after the introduction of the independent variable); and (3) the dependent or outcome variable is measured repeatedly across different phases. Although the standards were designed to be applied to wide variety of SCRD (e.g., ABAB designs, changing criterion designs, multiple baseline designs), they were only intended to be used with core SCRDs and not recommended for use with augmented independent comparison SCRDs (Kratochwill et al., 2013).

Next, if the study meets the criteria of a SCRD design, examiners will review all aspects of the study and classify it into one of three categories: “meets evidence standards”, “meets evidence standards with reservations”, or “does not meet evidence standards”. Specifically, the criteria examines if the independent variable was systematically manipulated, if the dependent variable was measured systematically, if a sufficient number of demonstrations of effect were present, and the number of data points within each phase (Kratochwill et al., 2013).

Finally, studies identified as “meets evidence standards” or “meets evidence standards with reservations”, undergo an evaluation of their visual analysis. The visual analysis of acceptable studies can be classified into one of three categories: “strong evidence”, “moderate evidence”, and “no evidence”. Specifically, the criteria examines the level, trend, and variability of the data points within each phase, the immediacy of effects when the independent variable is manipulated, the percent of overlap between adjacent phases, and the data patterns across similar phases (i.e., does the data show the same level, trend, and variability in the first and second baseline phases of an ABAB

design). To receive a rating of “strong evidence”, a study must document at least three demonstrations of effect with no non-effects by (1) documenting consistent trend, level, and variability across phases; (2) documenting immediacy of effect, acceptable levels of overlap between data points, and consistency across similar phases; and (3) accounting for confounding variables and anomalies. If a study documents at least three demonstrations of effect with at least one demonstration of non-effect, it is rated as “moderate evidence”. If a study does not document at least three demonstrations of effect it is rated as “no evidence” (Kratochwill et al., 2013).

In regards to combining the results of multiple studies, the panel recommended only describing the results in a single summary if (1) there are minimum of five SCRD studies that are classified as “meets evidence standards” or “meets evidence with reservations”, (2) the SCRD studies included were conducted in at least three different geographic regions, (3) and the total number of participants equals at least 20 (Kratochwill et al., 2013).

Autism Spectrum Disorder

Due to the increase in diagnoses for autism spectrum disorder (ASD), there is a greater need to identify evidence-based practices (Wong et al., 2015). According to the *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* (DSM-5), ASD has two behavioral domains: social communication deficits and repetitive/stereotypic behaviors. Social communication deficits are divided into three subdomains: deficits in social-emotional reciprocity; deficits in non-verbal communicative behaviors for social interaction; and deficits in developing, maintaining, and understanding relationships. Repetitive/stereotypic behaviors are divided into four subgroups: stereotyped or repetitive

motor movements, use of objects, or speech; insistence of sameness or inflexible adherence to routines; high restricted, fixated interest; and hyper- or hypo-activity to sensory input or unusual interest in sensory aspects of the environment (Fung & Hardan, 2014).

Weighted Vests

To reduce the negative effects of repetitive/stereotypic behaviors and increase on-task behavior, occupational therapists regularly prescribe individuals with ASD to wear weighted vests (WV) during specific activities or times (Morrison, 2007). A WV is a garment that adds an even distribution of up to 10% of an individual's body weight to a person (Stephenson & Carter, 2009). Those who prescribe their use purport them to be physically calming, assist in the organization of sensory input information by providing deep pressure, promote increased levels of the neurotransmitters serotonin and dopamine, affect deep brain structures, and reduce purposeless movements (Kane, Luiselli, Dearborn, & Young, 2004-05; Olson & Moulton, 2004; Morrison, 2007; Stephenson & Carter, 2009). Weighted vests have been used to reduce stereotypic behavior (Fertel-Daly, Bedell, & Hinojosa, 2001; Kane et al., 2004-2005) and increase attention to task for individuals with ASD (Fertel-Daly, Bedell, & Hinojosa, 2001; Kane et al., 2004-2005; VandenBerg, 2001).

Maintained, deep pressure, like the kind produced by WVs, claim to create calming effects by increasing parasympathetic or relaxed tone. Deep pressure can also produce calming effects by providing input to the thalamus, reticular formation, and the parietal lobe, which is located in the cerebral cortex (VandenBerg, 2001).

Weighted vests are a commonly used tool by many occupational therapists. Olson and Moulton (2004) administered a 43-item survey to 514 occupational therapists who were members of the American Occupational Therapy Association. Of the 514 randomly selected participants, 349 returned the mail survey. The results of survey indicated that a majority (82%) of the respondents use or have used WV to address the sensory needs of children with ASD. The respondents also reported anecdotal data that indicated the use of WV provided calming effects, increased students' attention to tasks, and reduced stereotypic behavior. Despite the overall positive opinions of WV, some respondents expressed concerns about lack of research examining the effectiveness of the practice.

Morrison (2007) conducted a review of the research on the use of WV on children with ASD. The criteria for inclusion in the review was publication between 1980 and 2006, that the article was written in English, inclusion of participants with ASD, and examination of the dependent variables attention to task and/or on-task behavior. Of the 37 articles found, five were included in the review (including Olson and Moulton [2004] study), and only three used experimental designs. The findings of the three studies that used experimental designs produced mixed results. Fertal-Daly, Bedell, and Hinojosa (2001) reported moderate improvements in attention to task and distractive behaviors when the participants wore WV. Kane, Wiselli, Dearorn and Young (2004-2005) reported no improvements in stereotypic behavior or attention to task when the participants wore WV and three of the four participants demonstrated negative outcomes when wearing the vests. The third article, by Myles et al. (2004), evaluated the results of three single case studies that examined the use of WV on students' with ASD. The results of the second study showed a negative effect on the dependent variable (i.e., on-task behavior) and the

results of the first and third studies reported positive outcomes on the dependent variables (i.e., attending, pressure-seeking behaviors).

Building off of the work of Morrison (2007), Stephenson and Carter (2009) also examined the research on autism and WV. In their review, the authors evaluated seven studies that used WV to improve the behavior of children with ASD and other developmental disabilities. The review included articles with empirical data that were published in peer-reviewed and non-refereed journals that examined the use of WV to improve the behavior of children with disabilities. A total of seven studies, five peer-reviewed, one non-refereed article, and one poster presentation were reviewed. The authors found methodological flaws with many of the studies, like inadequate participant description, experimental designs that could not be used to establish a functional relation (e.g., AB or ABA designs), and an insufficient amount of reliability data. The results of their analysis found there was not sufficient evidence to support the use of WV with children with ASD to improve their behavior.

Section 2: Purpose

Despite the results of previous literature reviews, WV continue to be used by occupational therapists and special educators. The purpose of this comprehensive literature review is to build off the work of Morrison (2007) and Stephenson and Carter (2009) and evaluate if newly conducted research has produced enough information to determine if the use of WV with individuals with ASD is an evidence-based practice, based on the criteria suggested by WWC.

Section 3: Methods

Search Procedures

The authors reviewed existent literature to evaluate the evidence-base for using WV to improve the behavior of children with ASD. The authors used the search terms *sensorimotor dysfunction, sensorimotor therapy, attention, sensory modulation, sensory integration, weighted vests, autis**, ASD, PDD, Aspergers within an electronic search of the following search engines: PsychInfo, ERIC, Academic Search Complete, Psychology and Behavior Sciences Collection, MEDLINE, and MasterFile Premier. The authors also conducted a hand search of the following journals: *The American Journal of Occupational Therapy, Journal of Occupational Science, Journal of Autism and Developmental Disorders, Education and Training in Developmental Disabilities*, and the *Journal of Special Education*. Finally, the authors examined the reference lists of the studies found through the electronic and hand searches and completed an ancestral search of their citations.

Inclusion criteria. The studies included in the review met the following criteria: (a) use of a group or single case research design; (b) inclusion of at least one individual with ASD as defined by the *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision* or DSM-5 (e.g., autism, PDD-NOS, PDD, Asperger's syndrome); (c) examination of the effects of WV on a particular dependent variable (e.g., aggressive behavior, attention to task); and (d) publication in English in a peer-refereed journal in the past 25 years. For purposes of this review, "weighted vests" were considered to be a wearable garment that added at least one pound to a person's weight. In the reference list, a single asterisk was used identify studies included in the review and two asterisk were

used to identify studies rated as “meets evidence standards” or “meets evidence standards with reservations” and retained for further analysis.

What Works Clearinghouse (2010) indicators. The authors used the quality indicators recommended by the WWC (2010) to evaluate each of the studies (see Table 1). A researcher-created data sheet was used to determine the presence or absence of each indicator. The data sheet was comprised of eight categories: (1) systematic manipulation of the independent variable, (2) collection of interobserver data for at least 20% of all sessions, (3) interobserver agreement of at least 80% of all sessions, (4) at least three demonstrations of effect, (5) at least five data points per condition, (6) at least three data points per condition, (7) clarification of design standards, and (8) classification of evidence for effectiveness.

Descriptive analysis. After the authors evaluated studies using the quality indicators recommended by the WWC (2010), the authors reported the characteristics of the studies that were categorized as “meets evidence standards” and “meets evidence standards with reservations”. The information (see Table 2) provided information about the following study components: (a) reference; (b) participant information (i.e., age, diagnostic label); (c) setting; (d) target behavior; (e) dependent variables; (f) weight of vest; (g) dosage of vest; (h) experimental design; and (g) findings. A master’s student was the primary coder for each of the studies. A researcher from a local university collected reliability data. The coded information is represented in Tables 1 and 2, which were created by the primary coder.

Determination of an evidence base for using weighted vests. The authors evaluated the studies rated as “meets evidence standards” and “meets evidence standards

with reservations” collectively against the criteria for evidence-based practices recommend by WWC (2010). Their criteria were (a) a minimum of five studies categorized as “meets evidence standards” and “meets evidence standards with reservations”, (b) the practice be examined by at least three different research teams, (c) the total number of participants included in the studies was at least 20, and (d) the studies be conducted in at least three geographic regions. A flowchart detailing the process of inclusion in this review can be found in Figure 1. While the authors elected to use the criteria recommended by the WWC (2010), it is important to note that other criteria for evaluating if a practice is evidence-based exist.

Interrater reliability on quality indicators and study characteristics. The authors coded a randomly selected article using the quality indicators to ensure reliability. The first author then coded all of the articles using the quality indicators and reported their descriptive characteristics. The second author, who is a researcher from the local university, coded at least 30% of the articles to examine interrater reliability. The studies examined for interrater reliability were selected randomly and the coders were unaware of the other’s scoring. Interrater reliability was calculated by using a point-by-point reliability method. Specifically, the number of agreements was divided by the number of agreements plus the number of disagreements, then the quotient was multiplied by 100 to convert the number into a percentage. A summary of the information from the descriptive analysis can be found in Table 2.

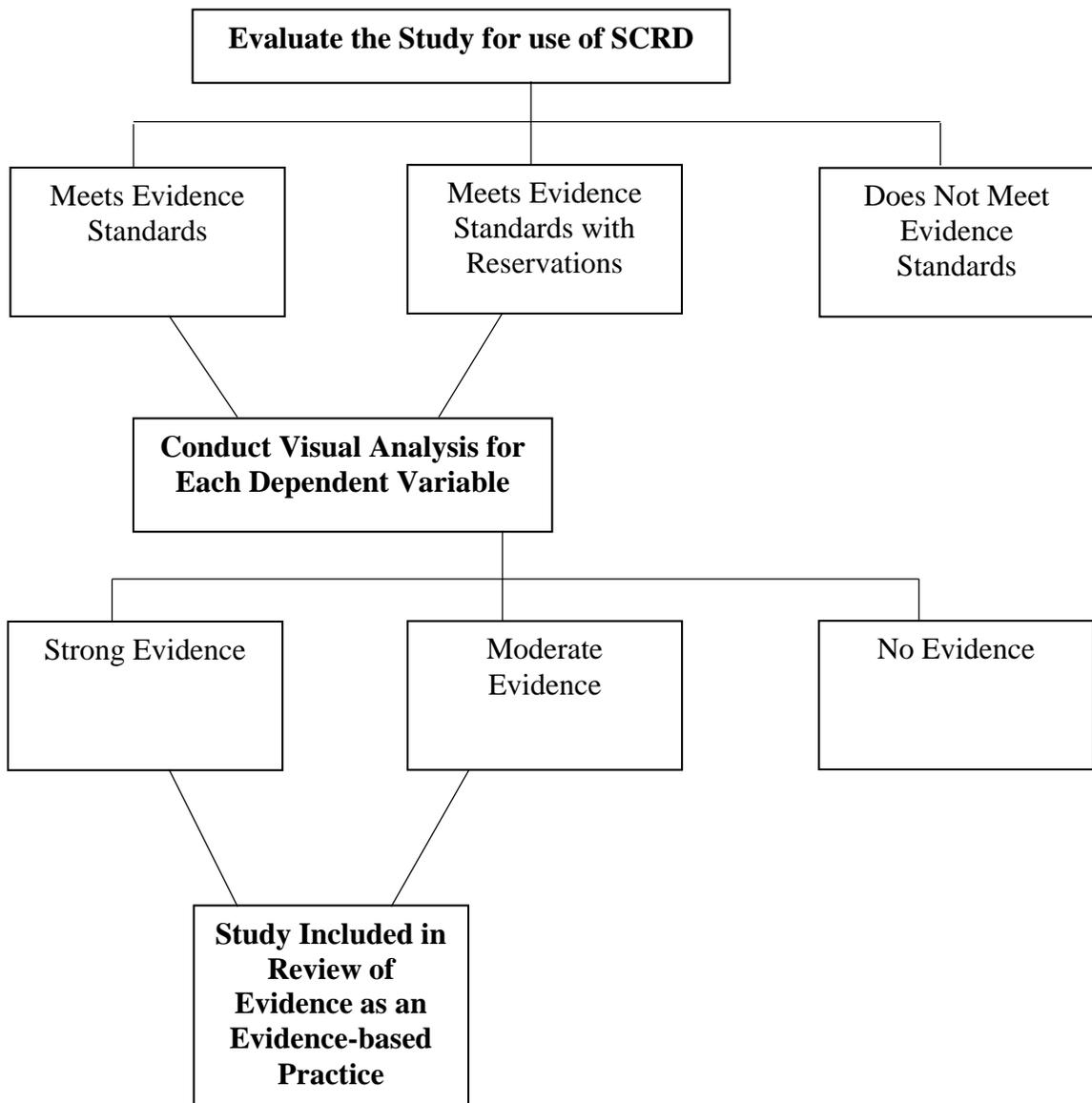


Figure 1: Flowchart for evaluation of SCRD adapted from Kratochwill et al. (2010).

Section 4: Study Characteristics

Quality of the Single Subject Studies

A total of 32 studies met inclusion criteria for this review. Data on the studies can be found in Table 1. Of the 32, four (12%) were rated as “meets evidence standards” and six (19%) were rated as “meets evidence standards with reservations”. Twenty two (69%) studies were rated as “does not meet evidence standards”. The studies rated as “does not meet evidence standards” were not retained for further review. Studies were rated “does not meet evidence standards” due to failure to apply the independent variable in a systematic fashion, absence of IOA data, IOA below an acceptable level, absence of at least three attempts of demonstrations of effect, and/or insufficient amount of data points in each phase. The most common reason studies were not rated as “meets evidence standards” or “meets evidence standards with reservations” was failure to report IOA data, which occurred in nine (32%) of studies.

Eight studies were retained to examine their descriptive characteristics. Data on the eight studies retained for further analysis can be found on Table 2. Researchers examined studies rated as “meets evidence standards” and “meets evidence standards with reservations” for descriptive information, including: (a) participants, (b) setting, (c) target behavior, (d) dependent variable, (e) weight of vest, (f) dosage of vest, (g) experimental design, and (h) findings.

Participants

A total of 8 children (4-10 years) with ASD participated in the studies rated as meeting evidence standards or meeting evidence standards with reservations. Six participants were male and two participants were female. Five of the studies (Cox et al.,

2009; Hodgetts et al. 2011) referenced the diagnostic tool, such as the Gilliam Autism Rating Scale (GARS), that was used to make the diagnosis of autism and three studies provided descriptions of the severity of the individual's diagnosis (Cox et al., 2009). Five of the studies provided scores from the Short Sensory Profile (Dunn, 1999), which is a metric used to measure sensory differences (Cox et al., 2009; Hodgetts et al. 2011). All of the participants had a total score of 141 or lower, which is below the total typical score of 190 (Cox et al., 2009). Five studies did not report any measure of sensory processing.

Settings

All of the studies were conducted in the participants' classrooms. Five studies took place in public elementary schools (Cox et al., 2009; Hodgetts et al. 2011), three studies took place in a university-affiliated early childhood center (Reichow et al., 2010), and two studies took place in early childhood special education classrooms (Myles et al., 2004). Four studies (Myles et al., 2004; Reichow et al., 2010) were conducted in an integrated setting. Three were conducted in self-contained classrooms for students with ASD (Hodgetts et al., 2011; Myles et al., 2004). Three studies were conducted in self-contained classrooms for students with unspecified disabilities (Cox et al., 2009). The majority (70%) of data collection occurred during group activities (Cox et al., 2009; Myles et al. 2004; Reichow et al., 2010). Two studies (Hodgetts et al., 2011) collected data only during individual activities. One study collected data during both group and individual activities (Myles et al., 2001).

Target Behaviors

The target behaviors examined in the studies measured a variety of behaviors commonly targeted by interventions for individuals with ASD, including: in-seat

behavior (Cox et al., 2009), off-task behavior (Hodgetts et al., 2011), on-task behavior (Myles et al., 2004), engagement (Reichow et al., 2010), stereotypic behavior (Reichow et al., 2010), and problem behavior (Reichow et al., 2010). In-seat behavior was the most frequently examined behavior. All studies reported operational definitions of the target behaviors that were observable and measurable.

Dependent Variables

Of the 10 studies retained for descriptive analysis, three measured the dependent variable using percent of intervals with appropriate in-seat behavior (Cox et al., 2009). Two studies measured the dependent variable with percent of intervals engaging in off-task behavior (Hodgetts et al., 2001). Two studies measured the duration of on-task behavior in seconds (Myles et al., 2004). One study measured the dependent variable using percent of intervals of engaged behavior (Reichow et al., 2010). One study measured the percent of intervals with stereotypic behavior (Reichow et al., 2010). One study examined the percent of intervals with problem behavior (Reichow et al., 2010).

Data Collection

Eight studies recorded data on the participants through videotaped footage (Cox et al., 2009; Hodgetts et al., 2011; Reichow et al., 2010). Two studies recorded data on the participants in vivo (Myles et al., 2004). Of the eight studies that examined the dependent variable using percent of intervals, three used momentary time sampling (Reichow et al., 2010). Five studies used whole-interval recording (Cox et al., 2009; Hodgetts et al., 2011). The two studies that measured the dependent variable using duration with a stopwatch (Myles et al., 2001).

Weight of Vest

All of the participants donned WV that weighed 5-10 percent of their body weight. The majority of the participants (88%) wore WV that weighed 5% of their body weight. One participant wore a WV that weighed 10% of his body weight (Myles et al., 2004). Of the eight vests worn by the participants, seven were made from denim (Cox et al., 2009; Hodgetts et al., 2011; Myles et al., 2004). The material that one vest was made from was not specified (Reichow et al., 2010). Seven vests were equipped with four pockets to hold weighted materials (Cox et al., 2009; Hodgetts et al., 2011; Myles et al., 2004; Reichow et al., 2010). One vest was equipped with nine pockets to hold weighted materials (Myles et al., 2004). The majority of the participants (88%) wore the WVs while completing activities in the classroom (Cox et al., 2009; Hodgetts et al., 2011; Myles et al., 2004; Reichow et al., 2010). One participant (Myles et al., 2004) wore a WV for 30 minutes prior to completing an activity in the classroom.

Single Subject Research Designs

All of the studies included in this review examined their research questions using SCRDS. Of the 10 studies retained for descriptive analysis, six used alternating treatment designs (Cox et al., 2009; Reichow et al., 2010). Four studies used withdrawal designs (Hodgetts et al., 2011; Myles et al., 2004).

Table 1: Evaluation of Studies using What Works Clearinghouse Guidelines

Authors (Design)	Systematic Manipulation of IV	IOA for 20% of sessions	IOA at or above 80%	At least 3 Demonstr. Of Effect	5 Data Points per Condition	3 Data Points per Condition	Classification of Design Standards	Classification of Evidence of Effectiveness
Demonstration Studies								
Fertel-Daly et al. 2001 (Withdrawal) 1	Y	N	Y	N	Y	Y	--	--
Fertel-Daly et al. 2001 (Withdrawal) 2	Y	N	Y	N	Y	Y	--	--
Fertel-Daly et al. 2001 (Withdrawal) 3	Y	N	Y	N	Y	Y	--	--
Fertel-Daly et al. 2001 (Withdrawal) 4	Y	N	Y	N	Y	Y	--	--
Fertel-Daly et al. 2011 (Withdrawal) 5	Y	N	Y	N	Y	Y	--	--
Hodgetts et al. 2011 (Withdrawal) 1	Y	Y	Y	Y	N	Y	MDwR	Moderate
Hodgetts et al. 2011 (Withdrawal) 2	Y	Y	Y	Y	N	N	--	--
Hodgetts et al. 2011 (Withdrawal) 3	Y	Y	Y	Y	N	Y	MDwR	Moderate
Hodgetts et al. 2011 (Withdrawal) 4	Y	Y	Y	Y	N	N	--	--
Hodgetts et al. 2011 (Withdrawal) 5	N	Y	Y	Y	N	N	--	--

Table 1 continued

	Hodgetts et al. 2011 (Withdrawal) 6	N	Y	Y	Y	N	N	--	--
	Hodgetts et al. 2011 (Withdrawal) 7	Y	Y	Y	Y	N	N	--	--
	Hodgetts et al. 2011 (Withdrawal) 8	Y	Y	Y	Y	N	N	--	--
	Hodgetts et al. 2011 (Withdrawal) 9	N	Y	Y	Y	N	N	--	--
	Hodgetts et al. 2011 (Withdrawal) 10	N	Y	Y	Y	N	Y	--	--
	Kane et al. 2004-2005 (ABC) 1	Y	N	N	N	N	Y	--	--
	Kane et al. 2004-2005 (ABC) 2	Y	N	N	N	N	Y	--	--
	Kane et al. 2004-2005 (ABC) 3	Y	N	N	N	N	Y	--	--
21	Kane et al. 2004-2005 (ABC) 4	Y	N	N	N	N	Y	--	--
	Leew et al. 2010 (MB-P)	Y	Y	Y	Y	N	N	--	--
	Myles et al. 2004 (Withdrawal) 1	Y	Y	Y	Y	N	Y	MDwR	Strong
	Myles et al. 2004 (Withdrawal) 2	Y	N	N	Y	N	Y	--	--
	Myles et al. 2004 (Withdrawal) 3	Y	Y	Y	Y	N	Y	MDwR	Moderate
<hr/>									
Comparison Studies									
	Cox et al. 2009 (ATD) 1	Y	Y	Y	Y	Y	Y	MDS	None

Table 1 continued

Cox et al. 2009 (ATD) 2	Y	Y	Y	Y	N	Y	MDwR	None
Cox et al. 2009 (ATD) 3	Y	Y	Y	Y	N	Y	MDwR	None
Reichow et al. 2010 (ATD) 1	Y	Y	Y	Y	N	N	--	--
Reichow et al. 2010 (ATD) 2	Y	Y	Y	Y	N	N	--	--
Reichow et al. 2010 (ATD) 3	Y	Y	Y	Y	N	N	--	--
Reichow et al. 2010 (ATD) 4	Y	Y	Y	Y	Y	Y	MDS	None
Reichow et al. 2010 (ATD) 5	Y	Y	Y	Y	Y	Y	MDS	None
Reichow et al. 2010 (ATD) 6	Y	Y	Y	Y	Y	Y	MDS	None

Notes: MDS = meets design standards; MDwR = meets design standards with reservations; MB-P = multiple baseline across participants; ATD = alternating treatment design

Table 2: Descriptive Information from Studies Rated as MDS and MDwR

	Participants	Setting (Activities)	Target Behavior	Dependent Variable	Weight of Vest	Dosage of Vest	Experimental Design	Findings
23	Cox et al. (2009) 1	5 years, 7 months GARS-35 th percentile (probable autism)	Self-contained classroom (circle time)	In-seat behavior	% of intervals of appropriate in-seat behavior	5% of body weight	Vest worn during activity (30 minutes)	Alternating treatment design No Evidence High percent of overlap, no immediacy of effect
	Cox et al. (2009) 2	6 years, 8 months CARS-severe autism	Self-contained classroom (circle time)	In-seat behavior	% of intervals of appropriate in-seat behavior	5% of body weight	Vest worn during activity (30 minutes)	Alternating treatment design No Evidence High percent of overlap, no immediacy of effect
	Cox et al. (2009) 3	9 years, 3 months GARS-45 th percentile (probable autism)	Self-contained classroom (circle time)	In-seat behavior	% of intervals of appropriate in-seat behavior	5% of body weight	Vest worn during activity (30 minutes)	Alternating treatment design No Evidence High percent of overlap, no immediacy of effect
	Hodgetts et al. (2011) 1	8 years, 0 months ADOS	Self-contained classroom (fine motor table-top activity)	Off-task behavior	% of intervals with off-task behavior	5% of body weight	Vest worn during activity (20 minutes)	Withdrawal (ABCBC) Moderate Evidence Some overlap between B and C conditions

Table 2 continued

Hodgetts et al. (2011) ³	10 years, 1 months ADOS	Self-contained classroom (fine motor table-top activity)	Off-task behavior	% of intervals with off-task behavior	5% of body weight	Vest worn during activity (20 minutes)	Withdrawal (ABCBC)	Moderate Evidence
Myles et al. (2004) ¹	5 years, 7 months ASD	Self-contained early childhood education class	On-task behavior	Duration (seconds)	10% of body weight	Vest worn during activities	Withdrawal (ABAB)	Some overlap between B and C conditions Strong Evidence
Myles et al. (2004) ³	4 years, 11 months ASD	Early childhood special education program	On-task behavior	Duration (seconds)	5% of body weight	Vest only worn for 30 minutes prior to instruction	Withdrawal (ABAB)	No overlap of data points, immediacy of effect Moderate Evidence
Reichow et al (2010) ⁴	5 years old ASD	University-affiliated, early childhood center (table time activity)	Engagement	% of intervals engaged	5% of body weight	Vest worn during activities	Alternating treatment design	Low overlap, high variability in first intervention phase No evidence
								High overall across conditions, no immediacy of effect

Table 2 continued

Reichow et al (2010) 5	5 years old ASD	University-affiliated, early childhood center (table time activity)	Stereotypic behavior	% of intervals with stereotypic behavior	5% of body weight	Vest worn during activities	Alternating treatment design	No Evidence 100% overlap across conditions
Reichow et al (2010) 6	5 years old ASD	University-affiliated, early childhood center (table time activity)	Problem behavior	% of intervals with problem behavior	5% of body weight	Vest worn during activities	Alternating treatment design	No Evidence 100% overlap across conditions

Notes: GARS = Gilliam Autism Rating Scale; CARS = Childhood Autism Rating Scales; ADOS = Autism Diagnostic Observation Schedule; PEP-R = Psychoeducational Profile-Revise

Section 5: Study Results and Outcomes

Student Outcomes

The effects of WV on student outcomes were determined through visual analysis. Of the 32 studies included in the review, 10 were retained for further analysis. Those 10 studies were retained because they were rated as “meets evidence standards” or “meets evidence standards with reservations”, which indicated they had an acceptable level of experimental rigor. Six of the 10 studies demonstrated no support for the use of WV and four studies did show support for the use WV. The four studies rated as “meets evidence standards” indicated that there was “no evidence” to support the use of WV with individuals with ASD (Cox et al., 2009; Reichow et al., 2010). Two studies rated as “meets evidence standards with reservations” also showed “no evidence” to support the use of WV with individuals with ASD (Cox et al., 2009). Three studies rated as “meets evidence standards with reservations” showed “moderate support” for the use of WVs (Hodgetts et al., 2011; Myles et al., 2004). One study rated as “meets evidence standards with reservations” showed “strong evidence” for the use of WV with individuals with ASD (Myles et al., 2004).

Cox et al. (2009) demonstrated that the use of WV had no effect on in-seat behavior. Reichow et al. (2010) demonstrated that the use of WV had no effect on engagement, stereotypic behavior, or problem behavior. Hodgetts et al. (2011) showed “moderate evidence” that the use of WV decreased off-task behavior in children with ASD. Myles et al. (2004) showed “strong evidence” that the use of WV increased on task behavior with a child with ASD.

Determination of an Evidence-based Practice

The four studies that demonstrated moderate to strong evidence for the use of WV with children with ASD were evaluated using the criteria for rating an evidence-based practice recommended by WWC. The criteria are (a) a minimum of five studies categorized as “meets evidence standards” and “meets evidence standards with reservations”, (b) the practice be examined by at least three different research teams, (c) the total number of participants included in the studies was at least 20, and (d) the studies be conducted in at least three geographic regions. First, the criteria requires that a minimum of five studies categorized as “meets evidence standards” or “meets evidence standards with reservations” show support for the practice. The results of this review show that only four studies with acceptable experimental rigor meet this standard. Second, the criteria require that the studies be conducted by at least three different research teams. The results of this review show that the four studies were only conducted by two research teams (Myles et al., 2004; Hodgetts et al., 2011). Third, the criteria require that the results be demonstrated across a minimum of 20 participants. The results of this review show that the results were only demonstrated across four participants. Fourth, the criteria requires that the studies be conducted across three different geographic regions. The four studies did not report the geographic region that the experiments were conducted. Based on the results of this review, the use of WV with children with ASD is not an evidence-based practice

Section 6: Discussion

The purpose of this literature review was to examine the findings on published studies examining the use of WV with children with ASD. This review expanded on previous reviews by including published studies conducted after 2004 and evaluating the literature base using WWC criteria. According to Kratochwill et al. (2013), evidence-based practices can be identified through a two-step process. First, researchers and practitioners must conduct an analysis of their experimental rigor, then they must examine the descriptive features of studies with acceptable experimental rigor and compare them to the criteria for an evidence-based practice. Specifically, studies with acceptable levels of experimental rigor (a) systematically manipulate the independent variable that they examine, (b) collect IOA data for at least 20% of sessions, (c) have IOA data at or above 80%, (d) use experimental designs that are capable of documenting at least three demonstrations of effect at three different times, (e) and collect a sufficient amount of data in each phase of the study. Sixty-nine percent of the studies examined in this literature review lacked the experimental rigor to be retained for further analysis. Fifty percent of studies (N=11) rated as “does not meet evidence standards” qualified for that classification due to an insufficient number of data points. However, several studies were rated “does not meet evidence standards” due to multiple methodological limitations (Fertel-Daly et al., 2001; Hodgetts et al., 2011; Kane et al., 2004; Myles et al., 2004). The 10 studies rated as “meets evidence standards” or “meets evidence standards with reservations” failed to meet the criteria for an evidence-based practice as defined by WWC. Specifically, they lacked the requisite number of acceptable studies, were not conducted by three independent research teams in three different geographic regions, and

did not include at least 20 participants. At this time, it appears that the use of WV with children with ASD is not an evidence-based practice.

Implications for Practice

Current legislation requires educators to use evidence-based practices as the impetus for making educational decisions. Unfortunately, for educators working with children with ASD, their choices are limited. Therefore, it is incumbent to use practices validated by rigorous research to ensure positive student outcomes and avoid interventions that may cause negative effects (Knight et al., 2015). The information gleaned from this review indicates that the use of WV with children with ASD is not an evidence-based practice. Practitioners should be aware of the literature examining WV when designing interventions for children with ASD.

Limitations and Conclusions

One limitation of this study was the criteria used for evaluating experimental rigor did not examine procedural fidelity. According to Barnett et al. (2014), procedural fidelity, also referred to as procedural reliability, treatment integrity, treatment delivery, and intervention delivery, describes the degree that an intervention is executed as planned. Procedural fidelity is regularly used as part of criteria for establishing a causal relationship between the independent and dependent variables. When a high degree of procedural fidelity is established, researchers can be more confident that the change in dependent variable was a result of the independent variable, opposed to an outside or confounding variable (Barnett et al., 2014).

Although the use of WV is popular among occupational therapists and educators, the results of this review indicate that the use of WV with children with ASD is not an

evidence-based practice. The results of this review are consistent with the findings of Stephen and Carter (2009). If researchers wish to extend the research on the use of WV in children with ASD, they should focus on establishing operationally defined criteria for the use of WV and increasing the experimental rigor of their studies. There is not a well-defined protocol for the use of WV. Indeed, much of protocol for using WV is based on anecdotal evidence (Reichow et al., 2010). Future research examining WV should begin including measures of procedural fidelity to help establish best practices on how to use the intervention. By providing an operationalized definition of how the intervention was used, researchers would be better equipped to replicate previous studies and practitioners would have more information of how to use the intervention. Of the 10 studies retained for further analysis, only 50% reported data on procedural fidelity (Cox et al., 2009; Hodgetts et al., 2011).

In addition to including a measure of procedural fidelity, future research should adhere to the standards recommended by WWC to increase experimental rigor. Unfortunately, a majority (69%) of the studies found through electronic, hand, and ancestral searches did not meet the criteria for further analysis. This review could have provided a more definitive answer on the efficacy on the use of WV with children with ASD if more published studies had adequate experimental rigor. An additional 22 studies could have been added to the descriptive analysis phase of this review if they had manipulated the independent variable in a systematic fashion, included adequate measures of IOA, used experimental designs capable of demonstrating at least three attempts of demonstration of effect (e.g., withdrawal designs that return to a baseline

condition), and included a sufficient number of data points in each phase of the experiment.

In addition to the legal obligations to use evidence-based practices, educators have an ethical obligation to use practices that will produce positive outcomes for their students. Although some practices are ubiquitous in special education, it is important to continually evaluate their effects. While the results of this review are limited due to the scarce number of studies with adequate experimental rigor, professionals working with children with ASD should be familiar with the literature examining WV and be mindful when using them. The results of this review highlight the fact that popularity and effectiveness are not synonymous.

References

*Indicates articles included in review.

**Indicates articles retained for further analysis.

Barnett, D., Hawkins, R., McCoy, D., Wahl, E., Shier, A., Denune, H., & Kimener, L. (2014). Methods used to document procedural fidelity in school-based intervention research. *Journal of Behavioral Education, 23*, 89-107.

**Cox, A. L., Gast, D., Luscre, D., & Ayres, K. M. (2009). The effects of weighted vests on appropriate in-seat behaviors of elementary-aged students with autism and severe to profound intellectual disabilities. *Focus on Autism and Other Developmental Disabilities, 24*(1), 17-26.

Dunn, W. (1999). Short sensory profile. San Antonio, TX: Psychological Corporation.

*Fertel-Daly, D., Bedell, G., & Hinojosa, J. (2001). Effects of a weighted vest on attention to task and self-stimulatory behavior in preschoolers with pervasive developmental disorders. *The American Journal of Occupational Therapy, 55*(6), 629-640.

Fung, L. K., & Hardan, A. Y. (2014). Autism in DSM-5 under the microscope: Implications to patients, families, clinicians, and researchers. *Asian Journal of Psychiatry, 11*, 93-37.

**Hodgetts, S., Magill-Evans, J., & Misiaszek, J. (2011). Effects of weighted vests on classroom behavior for children with autism and cognitive impairments. *Research in Autism Spectrum Disorders, 5*, 495-505.

- Horner, R. H., & Kratochwill, T. R. (2012). Synthesizing single-case research to identify evidence-based practices: Some brief reflections. *Journal of Behavioral Education, 21*, 266-272.
- 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.
- *Kane, A., Luiselli, J. K., Dearborn, S., & Young, N. (2004-05). Wearing a weighted vest as intervention for children with autism/pervasive developmental disorder: Behavioral assessment of stereotypy and attention to task. *The Scientific Review of Mental Health Practice, 3*(2), 19-24.
- Knight, V., Sartini, E., & Spriggs, A. D. (2015). Evaluating visual activity schedules as evidence-based practice for individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 45*, 157-178.
- 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.
- *Leew, S. V., Stein, N. G., & Gibbard, W. B. (2010). Weighted vests' effect on social attention for toddlers with autism spectrum disorders. *Canadian Journal of Occupational Therapy, 77*(2), 113-124.
- Morrison, E. E. (2009). A review of research on the use of weighted vests with children on the autism spectrum. *Education, 127*(3), 323-327.
- **Myles, B. S., Simpson, R. L., Carlson, J., Laurant, M., Gentry, A., Cook, K. T., & Earles-Vollrath, T. L. (2004). Examining the effects of the use of weighted vests

for addressing behaviors of children with autism spectrum disorders. *Journal of the International Association of Special Educators*, 5(1), 47-62.

Olson, L. J., & Moulton, H. J. (2004). Use of weighted vests in pediatric occupational therapy practice. *Physical and Occupational Therapy in Pediatrics*, 24(3), 45-60.

**Reichow, B., Barton, E. E., Sewell, J. N., Good, L., & Wolery, M. (2010). Effects of weighted vests on the engagement of children with developmental delays and autism. *Focus on Autism and Other Developmental Disabilities*, 25(1), 3-11.

Reichow, B., Volkmar, F. R., & Cicchetti, D. V. (2008). Development of the evaluative method for evaluating and determining evidence-based practices in autism. *Journal of Developmental Disorders*, 38, 1311-1319.

Stephenson, J., & Carter, M. (2009). The use of weighted vests with children with autism spectrum disorders and other disabilities. *Journal of Autism and Developmental Disorders*, 39, 105-114.

VandenBerg, N. L. (2001). The use of a weighted vest to increase the on-task behavior of children with attention difficulties. *American Journal of Occupational Therapy*, 55, 621-628.

Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., . . .

Schultz, T. R. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism and Developmental Disorders*, 45, 1951-1966.

Vita

Christopher J. Taylor

Barton College 2003-2004

University of Kentucky 2004-2008
Bachelor of Arts in Psychology

University of Kentucky 2009-2012
Bachelor of Science in Education

