

**How can we effectively develop independent functioning of children and
adults with Autism Spectrum Disorder?**

A thesis including published works

by

Monica E. Carr

Bachelor of Applied Science (Swinburne University of Technology)

Master in Organisational Leadership (Monash University)

Submitted in total fulfilment of the requirements for the degree of

Doctor of Philosophy

Faculty of Education



MONASH University

April 2015

Copyright © 2015, Monica E. Carr

Notice 1

Under the Copyright Act 1968, this thesis must be used only under the normal conditions of scholarly fair dealing. In particular no results or conclusions should be extracted from it, nor should it be copied or closely paraphrased in whole or in part without the written consent of the author. Proper written acknowledgement should be made for any assistance obtained from this thesis.

PART A: General Declaration

Monash University

Declaration for thesis based or partially based on conjointly published or unpublished work

General Declaration

In accordance with Monash University Doctorate Regulation 17.2 Doctor of Philosophy and Research Master's regulations the following declarations are made:

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes four original papers published in peer reviewed journals. The core theme of the thesis is identifying evidence based treatment to develop independent functioning of children and adults with Autism Spectrum Disorder. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the candidate, working within the Faculty of Education, Krongold Center under the supervision of Professor Dennis W. Moore and Dr Angelika Anderson.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of Chapters 3, 4, 5, and 6 my contribution to the work involved the following:

Thesis chapter	Publication title	Publication status	Nature and extent of candidate's contribution
Chapter 3	How should we determine treatment effectiveness with single-case design research for participants with autism spectrum disorder?	Published in Review Journal of Autism and Developmental Disorders (2014) doi:10.1001/s40489-014-0030-9	I have conceptualized and developed this work independently, identified and analyzed the data independently, and prepared the final manuscript under the guidance and collaboration of Prof Dennis Moore. Dr Angelika Anderson and Dr William Evans have provided additional guidance and collaboration of the final manuscript.
Chapter 4	A sensitivity analysis of three nonparametric treatment effect scores for single-case research for participants with autism.	Published in Review Journal of Autism and Developmental Disorders (2014) doi:10.1007/s40489-014-0037-2	I have conceptualized this work independently, identified and analyzed the data independently, and sole authored this publication. Prof Dennis Moore has provided collaboration and guidance regarding

			the selection of treatment effect calculations included in the study, the methods adopted in the analysis of the data, and has completed calculations in order to derive an inter-observer agreement report.
Chapter 5	Self-management interventions on students with autism: A meta-analysis of single-subject research.	Published in Exceptional Children (2014) doi: 10.1177/0014402914532235	I have identified this research topic under the guidance of Prof Dennis Moore and Dr Angelika Anderson as a systematic literature review. Although I have independently expanded the project into a meta-analysis, Prof Moore and Dr Anderson have provided guidance on apply quality assessment guidelines to evaluate the evidence. I have analyzed the data independently, and conducted the inter-observer agreement with a PhD student.

			Prof Moore and Dr Anderson provided guidance and collaboration on how to organize and prepare the final written paper.
Chapter 6	Goal setting interventions: Implications for participants on the autism spectrum.	Published in Review Journal of Autism and Developmental Disorders (2014) doi:10.1007/s40489-014-0022-9	I have conceptualized this research topic independently, Prof Moore provided guidance and collaboration on what data to include in this systematic search, and Dr Anderson provided guidance and collaboration on the thematic analysis of emergent themes. Subsequently, I have independently analyzed the data. The final paper was written under the guidance of Dr Anderson and Prof Moore.

I have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Signed:

A black rectangular box redacting the signature.

Date:

10 April, 2015

Acknowledgements

I would like to thank both Prof Dennis Moore and Dr Angelika Anderson, who have supervised my entire PhD journey. I believe I have benefited tremendously from their extensive knowledge of my research topic, and the world of academia. Despite the stress of the process, both Prof Moore and Dr Anderson have been continuously supportive, opened their hearts and their home, and have gone to great lengths to ensure that we have had fun along the way.

I would also like to thank Associate Prof Wee Tiong Seah for his ongoing support and technical guidance particularly with respect to the statistical analysis included in my study. In addition, I would like to thank Associate Prof Allie Clemans, who during her visits to Singapore in the early stages of my studies helped me develop an understanding of the world of social science research. Their continuous encouragement has been most appreciated.

An important component of my knowledge acquisition throughout the course of my studies has been gained through conference attendance. I would like to thank Dr Lynda Geller in New York, whom I have had the privilege of hearing speak at Adelphi University and in training in her office in Manhattan. Further, my depth of knowledge has developed tremendously after twice attending the Pivotal Response Training Conference at University of California, Santa Barbara. I would like to extend my thanks to the Koegel Autism Centre faculty and conference presenters.

Finally, I would like to thank my sons Andrew, Noah and Matthew who have been patient beyond their years. I would also like to thank my friends Aik Cheow, Stacey, Sarah and Dennis. Their unwavering support has been the motivation I needed to persevere during moments of doubt.

TABLE OF CONTENTS

GENERAL DECLARATION	ii
ACKNOWLEDGEMENTS	vii
TABLE OF CONTENTS	viii
LIST OF ABBREVIATIONS	xi
ABSTRACT	xiii
CHAPTER 1: INTRODUCTION	1
1.1 Overview	1
1.1.1 Titles of papers included in thesis	1
1.1.2 Research questions covered by the thesis	2
1.2 Theoretical framework	4
1.3 History of definitions of ASD	6
1.3.1 Table 1. Definitions of Autism Over Time	7
1.4 Prevalence estimates of ASD	8
1.5 Applied Behaviour Analysis	9
1.6 Summary	10
CHAPTER 2: LITERATURE REVIEW	12
2.1 Introduction	12
2.2 Younger children	12
2.3 Adolescents and adults	15
2.4 An important problem	16
2.5 Evaluating the evidence	18
2.5.1 Measuring treatment effect	26
2.6 Summary	30

CHAPTER 3: PUBLISHED PAPER 1	32
3.1 Declaration	33
3.2 How should we determine treatment effectiveness with single-case design research for participants with autism spectrum disorder?	34
CHAPTER 4: PUBLISHED PAPER 2	47
4.1 A sensitivity analysis of three nonparametric treatment effect scores for single-case research for participants with autism	48
CHAPTER 5: PUBLISHED PAPER 3	62
5.1 Declaration	63
5.2 Self-management interventions on students with autism: A meta-analysis of single- subject research	64
CHAPTER 6: PUBLISHED PAPER 4	81
6.1 Declaration	82
6.2 Goal setting interventions: Implications for participants on the autism spectrum	83
CHAPTER 7: DISCUSSION	102
7.1 Methodological implications of treatment effect calculations	102
7.2 Developing independent functioning	104
7.3 Findings from self-management and goal-setting systematic reviews	106
7.4 Summary	108
CHAPTER 8: CONCLUSION	110
8.1 Recount of study	110
8.2 Significance of findings	112
8.3 Implications for practice	113
8.4 Limitations of the study	115
8.5 Recommendations for future research	117

8.6 Contribution to knowledge	118
REFERENCE LIST	122

List of Abbreviations

ABA	Applied Behaviour Analysis
AD	Autistic Disorder
ADDM	Autism and Developmental Disabilities Monitoring network
APA	American Psychiatric Association
AS	Asperger's Syndrome
ASD	Autism Spectrum Disorder
CDC	United States Centers for Disease Control and Prevention
CDD	Childhood Disintegrative Disorder
CV	Confounding Variable
DSM	Diagnostic and Statistical Manual
DV	Dependent Variable
EBM	Evidence Based Medicine
EBP	Evidence Based Practice
EBPP	Evidence Based Practice in Psychology
FBA	Functional Behaviour Assessment
FDA	Food and Drug Administration, United States
HFA	High Functioning Autism
IDEA	Individuals with Disabilities Education Act
IES	Institute of Education Sciences, United States Department of Education
IOA	Inter-Observer Agreement
IV	Independent Variable
MEIP	Murdoch Early Intervention Program
NAP	Non-overlap of All Pairs

NRC	National Research Council (2001)
NSR	National Standards Report (2009)
PBS	Positive Behaviour Support
PDD-NOS	Pervasive Developmental Disorder – Not Otherwise Specified
PEM	Percentage Exceeding the Median
PND	Percentage of Non-overlapping Data
PAND	Percentage of All Non-overlapping Data
RD	Rett’s Disorder
SMD	Standardised Mean Difference
SMRS	Scientific Merit Rating Scale
WWC	What Works Clearinghouse (2010; 2011; 2013), United States Department of Education

Abstract

This thesis presents an exploration of how to develop independent functioning for children and adults who are diagnosed with Autism Spectrum Disorder (ASD). Two core areas of research are presented in a series of four published papers. Together with the introduction, literature review, discussion and conclusion chapters, eight chapters form the thesis via publication. All data included in this thesis has been extracted from systematic searches of peer-reviewed published literature and has drawn from self-management, exercise, and goal setting interventions. The first area of research explored how to measure the strength of treatment effect in single-case design (SCD) research. Over two sequential studies, data collection trends over time, and a subsequent sensitivity analysis of three appropriate treatment effect calculation methods was conducted. The widely reported percentage of non-overlapping data (PND) served as the basis of comparison, and in addition both the percentage of all non-overlapping data (PAND) and non-overlap of all pairs (NAP) were examined. The second area of research explored the use of self-management techniques to increase independent functioning for individuals diagnosed with ASD. A meta-analysis of these SCD interventions was conducted for studies that targeted the development of a new skill or the improvement of an existing skill. The U.S. Department of Education, What Works Clearinghouse (WWC) single-case design technical documentation guidelines were adopted to evaluate the quality of studies included in this review. PND was used to measure strength of treatment effect. One important finding of the meta-analysis indicated that goal setting skills were infrequently reported within the self-management literature for participants with ASD. It was argued that goal setting skills are an important component of independent functioning for children and adults with ASD, and a systematic review of the broader goal setting literature was conducted. Results were examined with respect to their implications for participants on the autism spectrum, and a future research agenda was proposed.

CHAPTER 1: INTRODUCTION

1.1 Overview

This thesis presents an exploration of how to best develop independent functioning for children and adults who are diagnosed on the autism spectrum. Two core areas of research have been conducted, and are presented in depth in a series of four published papers. Each published paper is presented as a stand-alone chapter. In addition to this introductory chapter, a literature review, discussion and conclusion provide the framing and linking text to present the entire study in a cohesive format of eight chapters that addresses the overall aim, and the specific research questions of the thesis.

The first core area addressed in Chapters 3 and 4 concerned technical issues that relate to how to measure the strength of treatment effect in single-case design (SCD) research, with a specific focus on the characteristics of participants with Autism Spectrum Disorder (ASD), and what these findings may mean for this population. The second core area addressed in Chapters 5 and 6 has explored interventions that may be beneficial for children and adults with ASD when developing skills that are required to function independently in educational, social and working environments, by critically evaluating the evidence base represented in studies using a SCD research paradigm.

All data used to inform this thesis has been extracted from peer-reviewed published literature. Literature on the topics of self-management, exercise, and goal-setting interventions has been used to build the data sets that have subsequently been analysed. A contribution to knowledge has been made via analysis and interpretation of these data as presented in the published papers included.

1.1.1 Titles of Papers Included in Thesis

Paper 1

How should we determine treatment effectiveness with single-case design research for participants with autism spectrum disorder? *Review Journal of Autism and Developmental Disorders* (2014) Monica E. Carr, Angelika Anderson, Dennis W. Moore, William H. Evans. DOI:10.1001/s40489-014-0030-9

Paper 2

A sensitivity analysis of three nonparametric treatment effect scores for single-case research for participants with autism. *Review Journal of Autism and Developmental Disorders* (2014) Monica E. Carr. DOI:10.1007/s40489-014-0037-2

Paper 3

Self-management interventions on students with autism: A meta-analysis of single-subject research. *Exceptional Children* (2014) Monica E. Carr, Dennis W. Moore, Angelika Anderson. DOI:10.1177/0014402914532235

Paper 4

Goal setting interventions: Implications for participants on the autism spectrum. *Review Journal of Autism and Developmental Disorders* (2014) Monica E. Carr, Dennis W. Moore, Angelika Anderson. DOI:10.1007/s40489-014-0022-9

1.1.2 Research Question Covered by the Thesis

This thesis aimed to answer the following question:

How can we effectively develop independent functioning of children and adults with Autism Spectrum Disorder (ASD)?

In addition to this overarching question, additional questions were developed and explored in the four papers. These questions were:

Paper 1:

1. Do studies report a sufficient number of baseline and intervention data points to enable the calculation of a treatment effect score?
2. Are there trends in these data suggesting this pattern is changing?

Paper 2:

1. Do estimated effect sizes calculated using PND, PAND, and NAP differ significantly from each other?
2. What benefits or limitations are evident in estimating treatment effect size using PAND, or NAP, when compared to the PND method?
3. How do calculated treatment effect scores compare with each other using available interpretation scales?

Paper 3:

1. Is self-management an effective intervention for improving target behaviours, or developing new skills, for individuals with ASD?
2. What types of skills have been developed using self-management with this population?
3. Is self-management training an evidence-based practice for individuals across the autism spectrum?

Paper 4:

1. What are the profiles of participants in goal setting research?
2. How have goal setting interventions been structured?

3. What target behaviour was goal setting applied to?
4. Who is involved in setting participant goals (self or other)?
5. Are goal setting skills maintained over time and was generalisation assessed?
6. What does the current literature suggest for applying goal setting to participants with ASD?

1.2 Theoretical Framework

The theoretical framework adopted for this thesis subscribes to a scientific approach within the field of Applied Behaviour Analysis (ABA). For professionals in this field, using a scientific approach implies a belief that behaviour can be understood in a systematic and organized manner. The three basic scientific concepts of description, prediction and control form the underlying premise of ABA (Cooper, Heron, & Heward, 2007). Description can be achieved through systematic observation of facts that relate to the observed event, and may serve as the basis for the development of a research hypothesis or specific research questions. Prediction, based upon repeated observations, may reveal a relationship between various events and may enable the prediction of the relative probability that an event may occur as the result of the presence of another observed event. Together, description and prediction may prepare an environment for systematic manipulation of a given event or variable. Control may be achieved through the demonstration of a functional relationship between events. In particular, a specific change in an event typically referred to as the *dependent variable* (DV), can be reliably achieved by specific manipulation of another event typically referred to as an *independent variable* (IV). In addition, other factors referred to as *confounding variables* (CV) should be considered unlikely to have caused the change in the dependent variable.

Historically, the world view of the field of behaviour analysis was traditionally characterized as mechanistic. According to Pepper's 1942 study (as cited in Morris, 1988) mechanism was described in terms of its root metaphor and theory of truth: as the machine and its parts, and the correspondence of theories with the facts that they predict. Pepper also originally described contextualism within the root metaphor as a historic event, and within theory of truth as successful working. Morris has described these criteria within the context of behaviour analysis to refer to effective action. For behaviour analysts, a mechanistic world view explores the cause and effect relationship between variables, whereas a contextual world view expands this relationship to incorporate the context in which it occurs. Proponents of the latter have argued that a mechanistic world view is considered insufficient for understanding behavioural change, and Morris has argued that current contemporary behaviour analytic theory is non-mechanistic (Morris, 1988).

It has been suggested that behaviour may be viewed as evolutionary in nature, hence belonging to a "becoming" ontology. In keeping with this notion, adopting a contextual ontological and epistemological position may suggest that neither response nor stimuli have any meaning by themselves. Rather, the meaning of behaviour emerges from its interacting context (Morris, 1988).

One century ago Watson was reportedly the first to have used the term "behaviourism". At that time Watson understood that introspection formed no part of the methods used by behaviourists, rather it was a purely objective experimental branch of natural science (Schneider & Morris, 1987). Some 20 years on in the 1930s, Skinner first described his own views as radical behaviourism. At this time, a main feature of radical behaviourism was the denial of consciousness. Skinner's stance has been described as being in accord with Watson's philosophical views in as much as consciousness was viewed as a non-existent entity, but in principle conscious events were amenable to scientific analysis. It

has been suggested that Skinner distinguished his views from other behaviourists not through his legitimization of private events, but rather by providing a coherent account as to how these events function as stimuli for verbal behaviour.

Radical Behaviourism underlies much of the literature that has informed this study.

1.3 History of Definitions of ASD

Diagnostic criteria, classification and prevalence estimates of Autism Spectrum Disorder have altered over time. Early versions of the American Psychiatric Association (APA) Diagnostic and Statistical Manual of Mental Disorders (DSM), DSM-I (1952) and DSM-II (1968) reflected the thinking of the time, in which diagnostic categories for ASD were not devised and psychoanalysis and psychotherapy dominated the field of psychiatry (Mayes & Horwitz, 2005). American psychiatry was organised around an environmental and behavioural model, in which symptoms were considered reflections of underlying dynamic conditions or as reactions to difficulties of life. The generally accepted belief was that mental disorders were reducible to a failure of the suffering individual to adapt to the environment (Menninger, 1963 cited in Mayes & Horwitz, 2005).

A major paradigm shift occurred within the field of psychiatry upon the adoption of a diagnostic model similar to that which was already used in medicine (Mayes & Horwitz, 2005). The revisions that were incorporated into the DSM-III (1980) brought about a symptoms based model that enabled standardised measurement of mental illness in a replicable manner. Autism was recognized as a separate category, and as a result the DSM-III legitimised claims that psychiatric professionals were treating a recognized condition (Baker, 2013; Mayes & Horwitz, 2005). This change enabled patients in the United States to obtain reimbursement for payments for treatment and services from third party insurers.

To date, no biological markers exist for ASD and a cause for the condition has yet to be identified. Diagnosis continues to be made on the basis of clinical observations as described most recently in DSM-5 (2013). Currently under DSM-5, ASD is defined in two behavioural domains: persistent impairment in reciprocal social communication and interaction, and restricted, repetitive patterns of behaviour, both present from early childhood. Sub categories including Asperger’s syndrome (AS) and Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS) that existed in DSM-IV (1994) have been eliminated from DSM-5. A description of the definitions of autism adopted over time is described in Table 1.

1.3.1 Table 1. Definitions of Autism Over Time

Kanner, L.	1943	Autistic Disturbances of Affective Contact
DSM-I	1952	No diagnostic categories. Symptoms viewed as reflections of broad underlying conditions or reactions to difficulties of life
DSM-II	1968	As above
DSM-III	1980	Autism defined as distinct from Schizophrenia, involved three domains: lack of responsiveness to others; gross impairment in communicative skills; and bizarre responses to various aspects of the environment, all developing within the first 30 months of age
DSM-III-R	1987	Autistic disorder requires 8 of 16 criteria among the three domains of social interaction, communication, and restricted interest or activities. Dropped early onset in life and provided additional category “Pervasive

		Developmental Disorder, Not Otherwise Specified (PDD-NOS) for children meeting some but not all diagnostic criteria of autistic disorder
DSM-IV	1994	Criteria for autistic disorder further refined, pervasive developmental disorders expanded to include Asperger's syndrome (AS) and Rett's syndrome (RS)
DSM-IV-TR	2000	
DSM-5	2013	Autism Spectrum Disorder (ASD) defined in two categories: persistent impairment in reciprocal social communication and interaction; plus restricted, repetitive patterns of behaviour, both present from early childhood. Sub categories including AS and PDD-NOS eliminated

1.4 Prevalence Estimates of ASD

Prevalence estimates of ASD appear to be rising rapidly. Although early estimates suggested presentation fell in the range of 4 – 6 per 10,000 children, epidemiological studies conducted by Wing in the U.K., and confirmed by Gillberg and colleagues in Sweden during the late 1980s, revealed that ASDs presented at a rate of 21 per 10,000 children (Gillberg, 1990). This estimate would equate to one in 400 children. However, a rapid increase is evident in prevalence estimates during the last decade. Using data gathered across the 14 sites included in The Autism and Developmental Disabilities Monitoring (ADDM) network, the U.S. Centers for Disease Control and Prevention (CDC) has reported consistent increases in the prevalence of ASD. In 2006 the CDC reported an average of one child in every 110 as having ASD, and by 2012 this estimate was revised to one in 88 children (CDC MMWR, 2012). Most recently, based upon data gathered from 11 sites, the estimate was revised to one in 68 children (CDC MMWR, 2014).

Within the United States it has been recently observed that the largest increases in prevalence have been among Hispanic and African American children (Pinder-Amaker, 2014). Further, when compared to CDC estimates from the United States, a significantly higher prevalence estimate was reported from a study conducted in South Korea. Kim and colleagues found that as many as one in 38 South Korean children may be diagnosed with ASD (Kim et al., 2011). However, in a global review of epidemiological studies of autism that spanned 14 countries, no evidence of an association between immigrant status or race and autism was identified (Fombonne, 2005).

High functioning autism (HFA) has been determined by some on the basis of intelligence quotients (IQs) equal to or in excess of 70 (Koyama, Tachimori, Osada, Takeda, & Kurita, 2007). However, Southall and Gast (2011) provided guidelines that distinguished HFA based on an IQ above 80, the use of functional language, and clinical reports for the Childhood Autism Rating Scale (CARS) above 30. Low-functioning or severe autism was identified on the basis of an IQ below 80, restricted communication, language or life skills, and clinical reports for CARS scores below 30.

Fombonne reported that 43.9% of children diagnosed with ASD were classified as higher functioning (Fombonne, 2005). Elsewhere in the literature, it has been suggested that as many as 50% of children diagnosed with ASD may be described as having High Functioning Autism (HFA) (Honda, Sgunuzu, Misumi, Niimi, & Ohashi, 1996; Kielenen & Moilanen, 2000).

1.5 Applied Behaviour Analysis (ABA)

To better understand an individual student and in turn provide appropriate support in an educational placement, a focus on individuality is essential. This is particularly relevant given the heterogeneity evident amongst individuals diagnosed on the autism spectrum, for

whom a combination of genetic makeup and lived experiences contribute to a unique personal profile of strengths and deficits. Although 90 percent of published literature in experimental psychology is based on group sampling and null hypothesis testing (Hubbard, Parsa, & Luthy, 1997; Loftus, 1991), Behaviour Therapy particularly ABA, is based on the intensive study of the individual (Blampied, 1999). Single-case research designs make it possible to draw scientifically valid conclusions from the investigation and treatment of individuals (Baer, Wolf, & Risley, 1968; Blampied, 1999).

The field of ABA that evolved from the earlier works of Skinner, has been defined in the literature as an applied science that focuses on the causes of socially significant behaviour change (Sigafoos & Schlosser, 2008). In his response to punitive misuse of practices ranging from shock treatment to solitary confinement, Skinner explained his original intention with the term *behavior modification* was “the management of human behaviour through contingencies of positive reinforcement specifically designed to replace the punitive techniques that are commonly observed in prisons and used by parents, teachers, employers and others” (Skinner, 1974, 813).

Despite being at the centre of ongoing controversy, the effective use of ABA interventions for individuals with an ASD has been well demonstrated in the empirical literature though misperceptions and misunderstandings of ABA are common; Sigafoos and Schlosser (2008) noted that some educational professionals wrongly think that ABA is a therapy or a treatment approach for autism.

1.6 Summary

This Chapter has provided an overview of the layout of this thesis via publication. In addition, the Chapter has described the population that is the focus of this study and has also described the research paradigm within which the study is situated. Chapter 2 will review the

literature that related to the two core areas ultimately provided in detail in the four published papers. Chapters 3 and 4 will examine in depth the issues that pertain to calculation of treatment effect scores given specific consideration to participants with a diagnosis of ASD. Chapter 5 will apply the treatment effect score argued as most appropriate in a meta-analysis of self-management intervention research conducted specifically with participants with a diagnosis of ASD. Chapter 6 will further explore goal setting intervention research beyond the inclusion criteria of a diagnosis of ASD. Chapter 7-8 will provide an integrated discussion of the four published papers, and conclude the study by highlighting the contributions to knowledge and suggesting recommendations for future research.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

The previous chapter has provided an overview of the population this study is focused on, and described the research paradigm that the study is situated in. This chapter of Literature Review will further identify the current issues and gaps that will ultimately be explored in depth in the subsequent chapters of the study. At the end of this review, this chapter will identify the areas of study that will be described in depth in Chapters 3, 4, 5 and 6.

2.2 Younger Children

In 2001 the National Academy of Sciences and the U.S. Department of Education formed a special committee to explore the issues concerning educational interventions for children with autism. In their report of the same year, the National Research Council (NRC) committee suggested that the earlier intervention begins in a child's life, the better the outcome. Particular recommendations were that services should include a minimum of 25 hours a week for 12 months a year, with the child engaged in planned and developmentally appropriate activities that target identified objectives that are unique to the child. The committee suggested that functional spontaneous communication, social instruction delivered throughout the day in various settings, cognitive development and play skills, and proactive approaches to behaviour problems be given priority (NRC, 2001).

Rather than adapting standard assessment procedures used with older children, researchers have recently developed diagnostic instruments to use with infants and toddlers. The Autism Diagnostic Observation Schedule Toddler module (ADOS-T) has been described as appropriate for use in children as young as 12 months (Luyster et al., 2009). While described by the original research group as successful in discriminating between diagnostic

groups, and yielding high sensitivity and specificity, the use of a “range of concern” rather than a cut off score has been recommended (Luyster et al., 2008). Bryson, Zwaigenbaum, McDermott, Rombough, and Brian (2008) have also developed the Autism Observation Scale for Infants (AOSI) for use with children under 2 years old, described as reliable and effective. As a result of these improved diagnostic and assessment methods early intervention has been embraced by the broader community rather than the “wait and see” attitude that was previously evident amongst many professionals who were responsible for screening and diagnosing ASD (Boyd, Odom, Humphreys, & Sam, 2010). Boyd and colleagues reported that improvements in early identification and diagnosis of ASD through the development of toddler and infant scales have driven this change, and contributed to timely access to support services.

Boyd and colleagues identified evidence-based practices and promising practices for infants and toddlers younger than 36 months diagnosed with ASD in their review of focused intervention practices used as a component of a Comprehensive Treatment Model (CTM). The research studies used in their data set were drawn from five support models: Children’s Toddler School, Project DATA (Developmentally Appropriate Treatment for Autism) for Toddlers, Early Start Denver Model, Early Social Interaction project, and the Walden Toddler Program. Eighteen components were identified and classified as either Behavioural intervention strategies or Positive Behaviour Supports (PBS). Four behavioural intervention strategies of the possible 18 components were represented in each of the five models reviewed. These components were: Prompting; Reinforcement; Naturalistic interventions; and Parent-implemented interventions (see Boyd and colleagues, 2010 for additional details). Boyd and colleagues have reported that access to high quality intervention during infancy and toddler years may lead to a reduction in severity of symptoms of ASD over the lifetime due to the malleability of the brain during this critical stage of early development.

Lovaas (1987) reported that early empirical results from behavioural interventions conducted with participants on the autism spectrum were both positive, in that it was possible to teach complex behaviours such as language and negate aggression or self-stimulatory behaviour, yet negative in that generalisation was not observed nor behaviours maintained at follow-up. Consequently, the pioneering program conducted by Lovaas and colleagues was specifically aimed at young children below the age of four, for whom it was hypothesized that it would be more likely to promote generalisation and maintenance of treatment effects, to address this limitation noted in the literature base of the era. This three year program targeted reductions in self-stimulatory and aggressive behaviours, developing compliance to verbal requests, teaching imitation and appropriate play with toys in both clinic and home settings. Subsequently, children were taught expressive and abstract language and play with peers. Settings were also expanded into the community. In the final stage of the project, appropriate and varied emotional expression, and pre-academic skills were developed, with the children placed in main-stream settings.

Several significant findings were reported from the study. Using the language of the time, 47% of the children in the experimental group achieved normal intellectual and educational functioning, 40% were described as mildly retarded and only 10% were considered profoundly retarded. However, only 2% of the control group achieved normal educational and intellectual functioning, 45% were described as mildly retarded and 53% were described as severely retarded. Lovaas and colleagues noted that findings from their study were consistent with the observation by Kanner (1943) that children on the autism spectrum may possess normal to superior intelligence.

Since the original early intervention study conducted by Lovaas (1987) reported significant gains from intensive early intervention, researchers have replicated this program and reported similar findings. In Australia, the Murdoch Early Intervention Program

(MEIP) (Birnbrauer & Leach, 1993) replicated the early intervention study originally conducted by Lovaas at the University of California, Los Angeles that ran for three years. The MEIP program was conducted in Western Australia with young children whose age ranged from 17 to 24 months at commencement of the program, and ran for an average of 21.56 months. Birnbrauer and Leach reported that substantial gains were realisable in even less ideal circumstances than those described by Lovaas, and described improvements in compliance, manageability, cooperation, independent and social play, and a reduction in self-stimulatory behaviours. In addition, the authors reported a decrease in parental stress levels as a study outcome.

2.3 Adolescents and Adults

As a result of the ability to diagnose very young children on the autism spectrum, and the widely reported success of early intervention programs developed over the last two decades, many children on the spectrum have been able to successfully complete high school. Consequently, larger numbers of adolescents are now able to enter tertiary education programs (Geller & Greenberg, 2010; Pinder-Amaker, 2014; Vanbergeijk, Klin, & Volkmar, 2008). While academically capable, many of these students may require ongoing support.

However, it has been noted that most autism-related research to date has been conducted with young children (Howlin, 2008). Howlin reported that of over 100 autism trials listed by the National Institutes of Health (NIH) in the United States, only three were specific to adults. Further, these three studies were drug trials.

A recent study conducted by Crosland and colleagues investigated research publication trends over the 15 year period 1995 – 2009, with data drawn from *Journal of Applied Behavior Analysis*, *Journal of Autism and Developmental Disorders*, and *Focus on Autism and other Developmental Disabilities*. The authors reported that research publication

trends for participants with typical cognitive functioning and typical communicative functioning increased over the period, however this increase was not considered consistent with prevalence estimates of children believed to fall at the higher functioning end of the autism spectrum for whom an increase in publications was not noted (Crosland, Clarke, & Dunlap, 2012). This finding is of concern given the rapid increase in ASD prevalence estimates.

Crosland and colleagues also reported an increase in studies for young children with ASD aged from birth through to five years from 21% to 36%. Drawing from publication rates of high-quality behavioural interventions conducted with adolescents and adults with ASD, it is believed that the volume of research for this subsection of individuals on the spectrum is low. For older children aged from six through to 11 years a decrease in publications from 61% to 48% was observed during the fifteen year period covered in the review (Crosland et al., 2012).

More recently, Miller, Vernon, Wu, and Russo (2014) conducted a systematic review of social skills interventions for adolescents, within the age range 13-17 years, and diagnosed with ASD. It was noted in their review that seven of 32 studies identified in the National Standards Project (2009) included participants of this age range, and Miller and colleagues reported that researchers are beginning to recognise the importance of interventions for adolescents with ASD.

2.4 An Important Problem

The core impairments that characterize autism - social and communicative functioning and the presence of restricted, repetitive and stereotyped interests and behaviours - are reportedly experienced to varying degrees into adulthood (Roth, Gillis, & DiGennaro Reed, 2014). Many individuals who face a life time of autism related challenges may grapple with little to no funding for support services. Research in the United Kingdom has reported

that specialist support beyond age 18 is sparse in college settings, that specialised employment services are rare, and support for independent living non-existent in much of the UK (Howlin, 2008). Howlin described adults on the spectrum as socially isolated, economically unproductive, and financially disadvantaged. In the United States, it has been reported that support systems have been slow in adapting to the needs of transition aged youth with ASD, who have largely been described as having typical dreams and aspirations (Geller & Greenberg, 2010). It is not inconceivable that the situations described are not unique to the United Kingdom or the United States, but experienced by adolescents and adults with ASD around the world.

Sigafoos and Schlosser (2008) reported that ABA based treatments are considered the most consistently effective approach for educating children. Elsewhere in the literature, a recent internet survey of treatments used by parents of children with autism reported that speech therapy was the most commonly adopted intervention, used by 70% of survey respondents (Green et al., 2006). Visual schedules (43.2%), sensory integration (38.2%), ABA (36.4%), and social stories (36.1%) also ranked amongst the five most frequently adopted treatments on a list of 108 reported treatments. Of concern, sensory integration, listed third, has not been demonstrated as evidence-based for participants with ASD, and was classified as an unestablished treatment by the National Standards Report (2009).

Elsewhere, The Association for Science in Autism Treatment has reported a long history of failed treatments and fads (Food and Drug Administration (FDA) Consumer Health Information, 2014) for individuals with ASD. In their report, the FDA warned that a number of companies may face legal action should they continue to promote false or misleading claims about products and therapies that claim to treat or cure autism. In particular, chelation therapy, hyperbaric oxygen therapy, miracle mineral solution, detoxifying clay baths, coconut kefir and other probiotic products were listed. Green and colleagues have also reported that

trends in their survey of treatments suggested that the presence or absence of an evidence base did not appear to influence whether a treatment was used often or infrequently.

In the United States, the National Autism Center's National Standards Report (2009) (NSR) noted that families, educators and service providers are presented with large volumes of often conflicting information about available treatments. Their project evaluated the effectiveness of available treatments and subsequently classified established, emerging, and unestablished treatments in the report.

As of 2008, over 1,000 peer-reviewed articles, spanning 47 years, describe scientifically robust success with individuals of all ages using treatments based on the principles of ABA (Foxx, 2008). Foxx reported that no other educational treatment approach to autism meets the standards of scientific proof that are met by ABA, nor are there any other scientifically valid treatments that produce similar treatment, educational or outcome results.

2.5 Evaluating the Evidence

Over three decades ago, Murphy and Bryan (1980) reported that both parents and professionals were increasingly demanding evidence regarding the effectiveness of services used to support developmentally disabled individuals, noting that the continued use of services whose alleged effectiveness is based on conjecture and anecdotal evidence was rapidly becoming unacceptable. At that time, Murphy and Brown suggested that the use of multiple-baseline and multiple-probe designs may provide a method to address concerns of accountability and evaluating intervention effectiveness.

In 2005, the APA Presidential Task Force on Evidence-Based Practice (EBP) developed a policy statement in which it defined evidence-based practice in psychology (EBPP), and discussed issues relating to the adoption of EBP in the field of psychology. The EBP movement was initially advocated within the medical profession, and has spread across

other health related disciplines to also include psychiatry, psychology, social work, physical and occupational therapy, and education (Kratochwill, 2007). While earlier described as evidence-based treatment or evidence-based intervention, EBP involves the identification, dissemination, and adoption of practices that are backed by scientific research.

In August, 2005, the APA Council of Representatives approved this initiative. Accordingly, EBPP has been defined as “the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences” (Goodheart et al., 2006). While the task force have noted that their definition has closely paralleled the definition of EBP adopted by the Institute of Medicine, the field of psychology has expanded this by deepening the examination of clinical expertise, and broadening understanding of patient characteristics.

Across the broader field of psychology, the task force reported that EBPP is intended to promote effective psychological practice and enhance public health by using psychological assessment, case formulation, therapeutic relationship, and interventions that are backed with empirical evidence (Goodheart et al., 2006). Given the multitude of elements that are included in EBPP, various types of research are endorsed by the APA. The term “Best research evidence” encompasses scientific results pertaining to assessment, intervention, clinical problems and patient populations which may be of relevance to laboratory, clinical or field settings.

The taskforce described EBPP as a means to enhance delivery of services that *encompasses a variety of stakeholders*. In their concluding remarks, the taskforce reported that psychological outcomes should not be limited to symptom relief and prevention of future symptomatic episodes, but should ideally also address quality of life, adaptive functioning,

the ability to make satisfying life choices, personality change and other *goals developed collaboratively between the patient and clinician* (Goodheart et al., 2006, p280).

Evidence has been defined as “research findings derived from the systematic collection of data through observation and experiment and the formulation of questions and testing of hypotheses” (Satterfield et al., 2009). Like other health services fields, agreement on how to define and grade research quality and accumulated evidence in the field of psychology has been noted as lacking (Kratochwill, 2007)). Of significance, Kratochwill indicated that research criteria are often supplemented by practice guidelines, enabling professional judgement to play a role in best-practice, both in assessment and intervention.

In a review article that has described the history and development of a model for determining an evidence-based practice (EBP), the challenge of how evidence should be defined and evaluated has been highlighted (Satterfield et al., 2009). In particular, a primary criticism in evidence-based medicine (EBM) is that evidence has been too narrowly defined, with the authors noting the significance of this in behavioural and social sciences. Satterfield and colleagues argued that in these fields the evidence base is not as extensive as in medicine, and that furthermore causality is often determined by several factors. On these grounds, Satterfield and colleagues emphasised the importance of defining evidence broadly, involving both quantitative and qualitative data.

A review of the literature suggested that quality assessment guidelines for SCD research have undergone a process of evolution since the issue was first highlighted in the 1980s. In particular, three research teams were identified that have developed quality assessment guidelines over the last sixteen years.

Chambless and Hollon (1998) drew upon the foundation work provided by the Division 12 (Clinical Psychology) Task Force on Promotion and Dissemination of Psychological Procedures (1995) and the APA Task Force on Psychological Intervention

Guidelines (1995). In their update, Chambless and Hollon evaluate treatment efficacy on the basis of demonstration of controlled research that included SCD experiments. Their proposal emphasised that replication is critical, particularly by an independent research team, and suggested a minimum requirement of two studies conducted by separate research teams. In the event that only one study supported a treatment's efficacy, or if all the research was conducted by one team, Chambless and Hollon considered the findings as promising but pending replication.

Chambless and Hollon developed a three-tiered classification scheme for SCD research. Their guidelines suggested that a treatment may be considered 'possibly efficacious' if treatment effects proved beneficial for at least three participants in a single group. Further, these authors recommended that in order to establish efficacy, three positive replications of a treatment effect be conducted by two independent research groups, involving a series of SCD studies including at least nine participants. For treatments described as efficacious and specific, identification of participant involvement in any alternative treatments during baseline phase(s) was required. With respect to visual data, the guideline suggested that treatment effects be so striking that they are readily apparent to the naked eye, however no recommendation for statistical analysis were set forth.

Subsequently, guidelines were expanded to determine whether a study met 'acceptable rigour' on the basis of replicable precision (Horner et al., 2005). In particular, Horner et al. (2005) required the inclusion of a description of participants, setting, DV, IV, and baseline that would enable replication. Their guideline specified that for the majority of design types, baseline should provide repeated measurement of the DV and establish a pattern of responding that may enable prediction of future performance. More specifically, Horner and colleagues required a demonstration of experimental effect for a minimum of three

different points in time, and also required the reporting of an Inter Observer Agreement (IOA) which met a minimal standard of 80% for each DV.

Most recently the What Works Clearinghouse (WWC) Procedures and Standards Handbook has been released by the US Department of Education Institute of Education Sciences (IES) expert panel (Kratochwill et al., 2010; Kratochwill et al., 2012; Kratochwill et al., 2013). The WWC was formed under the Education Sciences Reform Act (2002) with a focus on developing a model for education research that is based on high experimental rigour, in particular meta-analyses, in addition to addressing concerns of EBP and empirically-supported treatment that may enable federal and state governments to invest in educational, clinical and social practices that are scientifically valid (Horner, Swaminathan, Sugai, & Smolkowski, 2012).

The WWC assessment guidelines describe quality assessment procedures for both group designs (randomized controlled trial, quasi-experimental design, and regression discontinuity design) and SCD research. Under their guidelines, unpublished studies and published studies with a publication date not exceeding 20 years may be eligible for inclusion, and studies should present a primary analysis of the effect of an intervention.

Three versions of the WWC SCD Technical Documentation have been released since 2010: Version 1.0 (Kratochwill et al., 2010), Version 2.1 (Kratochwill et al., 2012), and Version 3.0 (Kratochwill et al., 2013). It has been noted in the literature that these guidelines are intended to guide a conversation across the SCD research community to determine professional consensus on the adequacy of design type, guidelines for conducting visual analysis, options for incorporating statistical analysis for interpreting SCD data and protocols for conducting meta-analyses of SCD literature (Horner & Kratochwill, 2012).

The SCD guidelines describe a procedure in which the design is initially evaluated to determine whether it meets evidence standards, meets evidence standards with reservations,

or does not meet evidence standards. This decision is based upon study rating determinants that the panel have described for systematic manipulation of the IV, adequate IOA for each outcome measure gathered for a minimum of 20% of data points, and adequate attempts to demonstrate effects over time and with sufficient data points for each phase as described in detail for reversal/withdrawal, multiple baseline and alternating treatment designs (see Procedures and Standards Handbook Version 3.0, page E.3 for most current data requirements).

All three versions of the guidelines then recommend conducting visual analysis of each outcome variable to determine strong evidence, moderate evidence or no evidence. However, versions 1.0 and 2.1 suggested also conducting an effect-size estimation calculation. In these versions, Kratochwill and colleagues reported that regression based estimates may be preferable to nonparametric based calculations, and suggested that if a nonparametric calculation were adopted, PND should be avoided. In addition these earlier versions of the guidelines suggested conducting a sensitivity analysis using several calculation methods to estimate effect size.

However, in the most recent version 3.0 Kratochwill and colleagues (2013) have cautioned against the use of an effect size calculation altogether until a greater consensus is achieved across the SCD research field. In the interim, the panel has recommended conducting visual analysis to determine the strength of treatment effect, and have described a process to conduct such analysis.

In summary, the WWC guidelines recommend a practice may be described as evidence-based if studies have met the minimum 5-3-20 rule:

- a) At least five SCD studies document experimental control;
- b) The five studies were drawn from at least three different research teams/locations, and;
- c) The five studies document effects for at least 20 different participants.

Separately, the NSR (2009) was developed to provide information to parents, educators and clinicians regarding the strength of evidence of educational and behavioural treatments for individuals below the age of 22 years. Participants were included if they were diagnosed with Autistic Disorder (AD), AS, or PDD-NOS. Studies for individuals diagnosed with Rett's Disorder (RD) and Childhood Disintegrative Disorder (CDD) were not included, nor were individuals identified as "At Risk" for ASD. Their report was based on literature published between 1957 through to 2007, and was limited to studies that included quantitative data in order to facilitate visual analysis of treatment effects. With the exception of curative diets, biomedical studies including medical trials, nutrition supplement studies, complementary and alternative medical interventions were excluded.

A scientific merit rating scale (SMRS) was developed to determine the extent to which an intervention was effective, and studies were evaluated on the basis of research design, measurement of the dependent variable, measurement of independent variable, participant ascertainment (reliability of participant diagnostic measures), and generalisation. Assessment criteria were devised for both group research and SCD research designs (See NSR Table 1, SMRS, page 18-22 for complete detail). In addition, treatment effects ratings criteria were developed for group and SCD research. Separate criteria were specified for alternating treatment designs. Intervention effects were classified as beneficial, ineffective, adverse, or unknown (See NSR Table 2, Treatment Effects Rating, page 26 for complete detail). The NSR (2009) identified 11 established treatments, 21 emerging treatments, and five unestablished treatments.

The three frameworks developed by Chambless and Hollon (1998), National Autism Center (2009) and WWC (2010-2013) were critically reviewed given consideration to the unique needs of individuals on the autism spectrum. Three important limitations were apparent. First, participant age ranges varied amongst the versions. The earlier framework

developed by Chambless and Hollon primarily focused on interventions conducted with adults. While the WWC guideline has not included upper age limits, the NSR (2009) capped research to not include participants exceeding 22 years old. Although individuals with autism may require support services across the life time, arguably research from early intervention and with school aged children is also important for inclusion when informing evidence-based best practice.

Second, unlike medical research, much of the seminal literature in the field of ABA was conducted outside a 20 year time frame, and these early findings remain relevant. The framework developed by Kratochwill et al. (2010), that has further developed the 5-3-20 guidelines developed by Horner et al., (2005), has suggested limiting inclusion criteria to studies published within the last 20 years. The panel have argued that restricting studies to within this 20 year time frame avoids inclusion of research that was previously conducted with populations or within contexts that may be considered very different to today. However, behaviour therapy has elsewhere been described as a treatment based on research rather than one deduced from theory, that produces behavioural change through environmental manipulations (Lovaas, Koegel, Simmons, & Stevens Long, 1973). Lovaas and colleagues reported that the methodology enables us to contribute in a cumulative manner to psychological treatment. To not limit the body of research to publications within the last 20 years is in direct contrast with the WWC inclusion criteria. In an argument more specific to individuals with ASD, that favours retention of early studies, research should not be omitted on the grounds of publication date as the core characteristics that define this population remain unaltered over time.

Third, debate currently surrounds a preferred method for calculating treatment effect size, and the use of visual analysis is recommended. Importantly, the APA Taskforce on Statistical Inference (1999) have considered treatment effect scores a requirement for

research publication (Leland Wilkinson and the Taskforce on Statistical Inference, 1999).

While it is generally agreed in the literature that establishing the best method to measure strength of treatment effect in SCD is a complex issue (Horner & Kratochwill, 2012; Shadish, 2014; Scruggs & Mastropieri, 2013; Parker et al., 2005), it may be argued that for the WWC to revert to a recommendation to evaluate strength of treatment effect on the basis of visual analysis is at odds with the earlier APA recommendation. As such, adoption of some quantitative measure of treatment effect is arguably necessary now.

2.5.1 Measuring Treatment Effect

It has been acknowledged that support systems are currently sparse for post-high school transition aged students and adults on the spectrum. Parents or older siblings may take on support roles in the absence of the school system or community supports that are available to many younger children. It is therefore important that information on the effectiveness of interventions is accessible and understandable to a broad array of stakeholders. Drawing from the broader educational psychology literature, Parker and colleagues (2005) have reported that the application of meta-analysis to SCD research has highlighted the need for valid, objective measures of treatment effects that can be communicated beyond clinical contexts (Parker et al., 2005).

Maggin, O’Keeffe, and Johnson (2011) conducted a systematic review of meta-analyses with SCD research involving students classified as disabled. In their review of publications during the years 1985-2009 a rapid acceleration in rate of publications over the last five years was identified, likely attributed to heightened interest in evidence-based practices. In particular, Maggin and colleagues reported that the Percentage of Nonoverlapping Data (PND) (Scruggs, Mastropieri, & Casto, 1987) effect size calculation was the most frequently adopted method, applied in 47 of the 84 (55%) effect sizes that were

included in the review. The Standardized Mean Difference (SMD) was the second most frequently used measure, applied in 16 of 84 (19%) of the studies.

Two major issues have been identified in the ongoing debate that surrounds adoption of a preferred method for calculating treatment effect. Firstly, an agreement is lacking on which calculation method to adopt. Secondly, it has been argued that a treatment effect score should be compatible with that reported for group design research, so that both may be included in meta-analyses (Horner & Kratochwill, 2012).

Brossart, Parker, Olson, and Mahadevan (2006) have identified several advantages associated with the use of an effect size calculation, when compared to visual analysis and statistical significance testing, claiming such a measure is the obvious choice for summarizing single-case study effects. Brossart and colleagues reported that effect sizes provide an index of the strength of association between intervention and outcome, hence providing the basis for explanation, control and prediction attributed to intervention. Effect sizes, which are not affected by sample size, support decisions to make incremental changes rather than simply continue or terminate decisions, by providing a continuous index of treatment success.

By comparison, Parker, Vannest, and Brown (2009) indicated significant limitations of effect size calculations, but concluded that they provide a useful supplement to visual analysis. In particular, Parker and colleagues argued that an effect size alone does not adequately distinguish between improvements in trend, in mean or median level, or both.

The widely adopted PND score has been found to be strongly correlated with expert ratings of treatment effectiveness, and further described as advantageous given that the procedure is conceptually meaningful (Scruggs & Mastropieri, 1998). One of the main criticisms of PND frequently mentioned is insufficient sensitivity to changes in slope.

However, Scruggs and Mastropieri (1998) have argued that the number of baseline observations typically found in single-subject research is inadequate to calculate meaningful trends.

While the first two versions of the WWC SCD guidelines suggested calculating a treatment effect score for studies that provide strong or moderate evidence according to their protocols for the various types of SCDs, no particular method was recommended (Kratochwill et al., 2010; Kratochwill et al., 2012). At that time, the WWC suggested that the use of a regression based calculation of treatment effect may be advantageous, arguing that many researchers may be familiar with these techniques, that the methods can account for any trend in the data, and that they can be used to calculate an effect size from a single case (unlike multilevel models that require several cases). However, earlier in the literature Scruggs and Mastropieri have reported that alternate metrics may produce effect sizes that are non-meaningful, and that regression based procedures for synthesis of single-case research will almost certainly not be accurate given limitations in the number of observations, and in failure to meet the assumption of equal time intervals across data points (Scruggs & Mastropieri, 1998).

The earlier two versions of the WWC SCD guidelines suggested that a nonparametric measure - PND, PAND or Percent Exceeding the Median (PEM) - may be adequate as an approximate measure of size of treatment effect, but emphasized that any of these nonparametric methods fail to address trend. At that time, the panel suggested that if a nonparametric method were adopted a sensitivity analysis of several methods be conducted (Kratochwill et al., 2010; Kratochwill et al., 2012).

The current version of the WWC SCD guidelines has suggested that strength of treatment effect be determined using visual analysis until a general consensus is achieved

across the field of educational psychology (Kratochwill et al., 2013). Shadish (2014) has reported that ongoing research on calculating treatment effect has seen great progress over the last decade, and noted that the field may begin to use a wider array of statistics software more often some of which are familiar to most SCD researchers such as SPSS and SAS, and other less common such as R and WinBUGS.

Shadish (2014) presented five articles published by several research groups, including members of the WWC panel, that have explored recently developed methods of treatment effect calculation in an effort to understand whether varying approaches resulted in a similar outcome report of treatment effect. All studies used the same data set, extracted from published interventions on the effects of response cards on disruptive behaviour and academic responding during math lessons for four year olds (Lambert, Cartledge, Heward, & Lo, 2006). Calculation methods that were explored in these five articles included SMD, multilevel modelling, Bayesian statistics, and generalized additive models.

Results from four of the five articles reported a reduction of disruptive behaviour that ranged from 5.03 to 5.70 intervals. The fifth article described a modest outlying result, and reported 6.70 intervals of disruptive behaviour. Shadish described these five reported outcomes as reasonably consistent with each other, and with visual inspection of the original graphed data. In conclusion, Shadish emphasized several issues that may be of particular relevance to stakeholders interested in evaluating treatment effectiveness for participants on the autism spectrum. Shadish suggested that the SCD research community may in fact be better described as a group of SCD communities with varying opinions on the application of statistics to research. Further, it was asserted that the issue may not so much be whether researchers *will* eventually derive an agreed upon improved statistical approach for evaluating treatment effect, but rather whether *SCD researchers will adopt new methods*. Arguing this possibility, it was pointed out that such a change in the field would require a

paradigm shift in SCD research. Shadish noted that such a major change has historically been rare and that any future change will depend upon many competing variables such as influences from external sources, changes in training and publication practise, and recognition of SCD research misconceptions about the potential value of statistics.

While currently a consensus across the field regarding the most appropriate method of evaluating the effectiveness of intervention research and conducting meta-analyses has not yet been established, the ASD community of stakeholders requires information describing evidence-based best practise *immediately*. It is important to identify potential treatments that may in fact cause harm, and in the best interests of all parties to avoid selecting treatments that may be ineffective. Accordingly, research has been conducted for this thesis that has specifically examined the issues surrounding the calculation of treatment effects for participants that have a diagnosis of ASD, in order to determine what may be the most appropriate methodological approach to currently adopt while additional research efforts are pursued by the academic community that may ultimately yield a more sophisticated statistical approach.

2.6 Summary

This chapter has identified the ongoing debate surrounding a preferred method to calculate treatment effect scores in SCD research. This issue was noted to be of central concern when conducting a quality assessment of intervention research, and of utmost significance when identifying evidence based practice. The following four chapters will explore these issues in depth. Chapters 3 and 4 will examine SCD data collected for participants with ASD and identify a preferred treatment effect calculation method. Chapter 5 will apply the suggested calculation to self-management intervention research and conduct a quality assessment of the literature to identify evidence based practice. Chapter 6 will

explore additional literature on goal setting, drawing from interventions conducted with participants either with or without a diagnosis of ASD to identify what is known about this technique. The findings will be used to inform goal setting intervention research specifically for the ASD population.

CHAPTER 3: PUBLISHED PAPER 1

How should we determine treatment effectiveness with single-case design research for participants with autism spectrum disorder? *Review Journal of Autism and Developmental Disorders* (2014) Monica E. Carr, Angelika Anderson, Dennis W. Moore, William H. Evans.
DOI:10.1001/s40489-014-0030-9

Declaration for Thesis Chapter Three

Declaration by candidate

In the case of Chapter 3, the nature and extent of my contribution to the work was the following:


Nature of contribution	Extent of contribution (%)
I have conceptualized this research project independently, and independently developed the data set. I have initially graphed and analysed data independently, and have further refined the analysis and expanded the data set under the guidance of Prof Dennis Moore. I have prepared the first draft of the manuscript independently, and further developed the manuscript under the guidance of Prof Dennis Moore. Final refinement and review of the manuscript have been supported by the guidance of Dr Angelika Anderson, and Prof William Evans.	90

The following co-authors contributed to the work. If co-authors are students at Monash University, the extent of their contribution in percentage terms must be stated:

Name	Nature of contribution	Extent of contribution (%) for student co-authors only
Angelika Anderson	Guidance and collaboration for final manuscript prior to submission to peer review	4
Dennis W. Moore	Guidance and collaboration of data analysis, expansion of data set, and review of manuscript throughout duration of project	4
William H. Evans	Guidance and collaboration for final manuscript prior to submission to peer review	2

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the candidate's and co-authors' contributions to this work*.

Candidate's
Signature

 Date
21 April, 2015

Main
Supervisor's
Signature

 Date
21 April 2015

*Note: Where the responsible author is not the candidate's main supervisor, the main supervisor should consult with the responsible author to agree on the respective contributions of the authors.

How Should We Determine Treatment Effectiveness with Single-Case Design Research for Participants with Autism Spectrum Disorder?

**Monica E. Carr, Angelika Anderson,
Dennis W. Moore & William H. Evans**

**Review Journal of Autism and
Developmental Disorders**

ISSN 2195-7177

Volume 2

Number 1

Rev J Autism Dev Disord (2015) 2:8-18
DOI 10.1007/s40489-014-0030-9



Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

How Should We Determine Treatment Effectiveness with Single-Case Design Research for Participants with Autism Spectrum Disorder?

Monica E. Carr · Angelika Anderson ·
Dennis W. Moore · William H. Evans

Received: 30 June 2014 / Accepted: 1 July 2014 / Published online: 19 July 2014
© Springer Science+Business Media New York 2014

Abstract Guidelines to inform research evidence standards have acknowledged that there is currently no agreed-upon method for treatment effect size estimation in single-case research. This study has examined the application of treatment effect size calculations to supplement visual analysis in single-case research designs (SCD) for participants with autism spectrum disorder (ASD). Ethical considerations for researchers regarding the collection of baseline data in light of behaviors often associated with ASD are discussed. The adequacy of the volume of data points from baseline and treatment phases was explored, and the conclusion had drawn that the majority of studies were not suitable for regression calculations. The median length of total data series was also explored, and the suitability of three nonparametric hand calculations, percentage of nonoverlap (PND), percentage of all nonoverlapping data (PAND), and nonoverlap of all pairs (NAP), is discussed in this light.

Keywords Autism · Treatment effect · Regression · Nonparametric

This study was completed in partial fulfillment of the requirements for a PhD for the first author.

M. E. Carr (✉) · A. Anderson · D. W. Moore
Faculty of Education, Monash University, Building 6, Clayton
Campus, Clayton, VIC 3800, Australia

A. Anderson

D. W. Moore

W. H. Evans
School of Education, University of West Florida, Building 85,
Pensacola, FL 32514, USA

Introduction

Heterogeneity presents a unique challenge within the field of autism research, as individuals with Autism Spectrum Disorder (ASD) exhibit significant variability in the kind and extent of symptomatology. Research conducted in 2006 has shown that parents use a wide range of treatments with their children, with a greater number of treatments being used for younger children, and for children with greater severity of symptoms (Green et al., 2006). Green and colleagues also reported that the most commonly utilized treatments included those without empirical evidence.

Applied Behavior Analysis (ABA) is an applied science that focuses on socially significant behavior change (Baer, Wolf, & Risley, 1968; Sigafoos & Schlosser, 2008). While most published psychological research is based on between group research designs, ABA typically examines behavior at the level of the individual and generally utilizes single-case research designs (SCD), thus permitting a scientifically valid conclusion to be drawn from the intensive investigation of an individual (Blampied, 1999). Interventions based on such research have been used extensively in working with participants with ASD since the early 1980s.

Systematic reviews and meta-analyses of SCD literature are becoming increasingly important to a variety of stakeholders and have been conducted within academic literature, by government agencies, and health service providers to address the need for evidence based practice guidelines as well as to inform decisions at a policy level. Across the broader field of healthcare, the PRISMA Statement (2009) sets forth a checklist of 27 items that should be addressed in systematic reviews or meta-analyses of literature (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009). The PRISMA Statement (2009) has been used to guide reviews that are ultimately read by clinicians to inform practice, granting agencies to fund future research, and other stakeholders. Such reviews may include between-group design

research and SCD research. It has been acknowledged that one limitation of meta-analytic research is that historically SCD research has often been omitted (Allison & Gorman, 1993).

Results from single-case meta-analyses have recently been used by health insurance providers both to support and deny the necessity for intensive behavioral intervention for individuals with ASD (Campbell, 2013). Campbell reported that in 2011, CIGNA companies concluded that behavioral intervention is an effective therapy, while in the same year, United Healthcare Services, Inc. used similar evidence to justify their policy that claims the same treatment is not a medical necessity. Separately, the National Autism Center's National Standards Report (2009) has assessed the existing SCD published peer reviewed literature base for participants under 22 years of age diagnosed with ASD. The report has categorized 11 ABA-based interventions as established, 21 as emerging, and five as unestablished treatments. Established treatments were antecedent package, behavioral package, comprehensive behavioral treatment for young children, joint attention intervention, modeling, naturalistic teaching strategies, peer training, pivotal response treatment, schedules, self-management, and story-based intervention package. Emerging treatments were augmentive and alternative communication devices, cognitive behavioral intervention package, developmental relationship-based treatments, exercise, exposure package, imitation-based interaction, initiation training, language training (production), massage/touch therapy, multi-component package, music therapy, peer-mediated instructional arrangements, the picture exchange communication system, reductive package, scripting, sign instruction, social communication intervention, social skills package, structured teaching, technology-based treatment, and theory of mind training. Unestablished treatments were academic interventions, auditory integration training, facilitated communication, gluten- and casein-free diets, and sensory integrative packages. Their report also highlighted the importance of using meta-analytic procedures to identify ineffective or harmful treatments, although in their 2009 review, no studies were identified that met criteria. At the time of writing, the National Association of Insurance Commissioners and Top Health Insurance (2013) ranked United Healthcare as the number one insurer in the USA, providing services to an estimated 70 million Americans. Given the current ASD prevalence estimate of one in 88 (CDC MMWR, 2012), approximately 795,000 individuals are affected by such policy decisions in the USA alone.

Variations in literature synthesis, such as described above, have wide reaching implications for individuals on the autism spectrum and may impact whether or not an individual is able to access support services. However, despite interest across the broader educational field regarding how best to calculate and report strength of treatment effects, an agreement on how to measure and interpret the strength of treatment effects with SCD research studies has yet to be achieved.

SCD researchers in the field of ABA have traditionally relied on visual analysis as the principal method of determining intervention effects (Kratochwill & Levin, 2014; Matyas & Greenwood, 1990; Shadish, 2014). Visual analysis can be used to document experimental control and determine the overall effectiveness of an intervention by assessing all conditions within a design, with graphical inspection involving the evaluation of time series data in terms of systematic changes in level, trend, and variability, both within and across intervention phases (Horner et al., 2005). Historically, visual methods have been favored over statistical approaches on the basis that the former is less likely to report false positive treatment outcomes (Shadish, 2014). While visual analysis has wide appeal, it is not without criticism. In particular, unreliability of judgment across raters has been frequently reported (Campbell, 2013; Parker & Brossart, 2003; Scruggs, Mastropieri, & Casto, 1987) though it has also been noted that critics have seldom addressed consistency in visual analysis beyond two phases (Horner, Swaminathan, Sugai, & Smolkowski, 2012). Horner and colleagues have argued for the continued use of visual analysis in the absence of agreement on a statistical measure to determine treatment effect.

However, while much support is acknowledged for the continued use of visual analysis, we argue that such an approach in literature synthesis may not adequately address several important and topical issues in ASD research and treatment. Firstly, the American Psychological Association (APA) Taskforce on Statistical Inference (1999) argued in support of the earlier APA (1994) publication manual's suggestion to include a treatment effect size in research reports, claiming that a treatment effect size permits the evaluation of the stability of findings across samples and is important to future meta-analyses. At this time, the earlier guideline was formalized as a requirement for research publications (Leland Wilkinson and the Taskforce on Statistical Inference, 1999). While some (APA Taskforce on Statistical Inference, 1999) have argued that, with improvements in state-of-the-art statistical analysis software, statistics are commonly reported without an understanding of computational methods or an understanding of what the statistics mean, Kratochwill and Levin (2014) recently highlighted that there are a growing number of nonoverlap methods that can be hand calculated, which may be advantageous as a supplement to visual analysis. Arguments for the retention of simple calculation methods such as percentage of nonoverlap (PND) have also been reported (Scruggs & Mastropieri, 2013).

Secondly, the evidence based best practice movement across the broader field of educational psychology highlights the importance of calculating a treatment effect score when evaluating and synthesizing literature. A treatment effect score is considered essential to informing an evidence base. The U.S. Department of Education, What Works Clearinghouse (WWC) has produced a procedures and standards handbook

to assess the quality of both group and SCD studies and developed protocols to evaluate their effectiveness in order to establish a scientific evidence base for educational research (Kratochwill et al., 2010). The WWC SCD Pilot Version 1.0 guidelines suggested a preference for the use of regression-based calculations, advising specifically against the adoption of the PND calculation. Elsewhere in the earlier literature, researchers have stated a preference for regression-based approach (Allison & Gorman, 1993; Parker & Brossart, 2003). By contrast, the most recent update to the WWC (2013) Version 3.0 has reverted to recommending visual analysis as the primary procedure used in determining the strength of a treatment effect (Kratochwill et al. 2013). In this current version, the panel has predicted that at some future point, when the field has achieved greater consensus about appropriate quantitative techniques, new standards for effect demonstration will be developed. However, at the time of writing, the WWC guidelines do not specify a metric to use when calculating a treatment effect size.

SCDs are unique in that significant design decisions regarding length of baseline data collection and when to implement or withdraw treatment are not determined in advance, depending rather on the participant data that are collected. Typically, phase changes are made once the data within a phase are considered stable, as is characterized by the absence of slope and no more than a small level of performance variability within the phase (Kazdin, 1978). Ethical consideration of a participant's circumstances and of the behavior under investigation may also contribute to determining how many data points are collected. If a participant has a severe skill deficit, or if a behavior is harmful either to the self or others, it may not be socially valid or ethically acceptable to prolong baseline data collection. Such scenarios are common in studies involving individuals with ASD. Even in cases where the intervention does not directly target a reduction in problem behavior, challenging behavior may still be present and an issue to be considered.

Debate regarding the most appropriate methodological approach to interpreting SCD research dates back as far as the early 1970s (Kratochwill & Brody, 1978). Effect size may be calculated using regression-based estimators, standardized mean difference, or nonparametric methods. Of these, the most extensively adopted method is the nonparametric calculation, PND developed by Scruggs, Mastropieri, and Casto (1987). However, PND has been criticized for misrepresenting treatment effects, being insufficiently sensitive to changes in slope, producing an invalid outcome in instances in which outlier data points in baseline obscure true intervention effects, being ineffective as a discriminator for powerful treatment effects, and that the number of baseline observations may in itself distort outcomes (Scruggs & Mastropieri, 1998).

A variety of statistical approaches for use in SCD meta-analyses are currently under development including

procedures for modeling trend, determining and estimate of treatment effect size, and investigating statistical methods to improve estimates for small data samples (Shadish, 2014). In a recent special series of articles exploring emerging approaches to calculating treatment effect, it has been suggested that what exists is a group of communities and that perspectives of researchers from various sub-communities may differ greatly on the role of statistical analysis in research interpretation and reporting (Shadish, 2014). In light of this claim, we argue the significance of examining SCD data specifically in the context of participants with ASD.

Data Requirements for Treatment Effect Calculations

A review of existing literature was conducted to identify minimum data requirements for use in both regression-based and nonparametric approaches to calculating a treatment effect score. Parker et al. (2005) noted a minimum of six data points per phase, and at least 14 data points in a phase A and B comparison are required for a regression-based effect size calculation. In their research based upon a convenience sample of 77 published AB datasets, Parker and colleagues reported a median number of data points per graph of 23 (counting only A and B phases), with a median length of 10 for phase A and 11 for phase B for their sample.

In addition, our search identified three nonparametric approaches that can be calculated by hand: PND (Scruggs et al. 1987), percentage of all nonoverlapping data (PAND) (Parker et al. 2007), and nonoverlap of all pairs (NAP) (Parker & Vannest, 2009). PND was developed specifically to supplement visual analysis. While PND does not specify a minimum number of data points, outlier ceiling or floor effects in baseline data can result in the calculation of a zero score. Baseline variability may mean that a PND score should not be calculated, and Scruggs and Mastropieri (2013) have reiterated their original advice against the calculation of an effect score in those cases where the result would be inconsistent with visual examination.

PAND presents an approach designed to address the limitation of rogue outlier baseline data present in the PND calculation. For the nonparametric calculation PAND, Parker et al. (2007) reported that a minimum threshold of 20 data points in total is necessary. Parker and colleagues demonstrated the suitability of PAND for an initial AB comparison using a dataset comprised of multiple baseline design samples in which 60 to 80 data points were typical.

NAP offers a calculation that utilizes all data points in a pairwise comparison. Parker and Vannest (2009) examined the performance of NAP using 200 AB contrasts and reported that the median length of a full data series in their sample was 18 data points. In particular, phase A had a median of 8 data points, and phase B had a median of 9 data points.

Purpose of Study

A preferred method to evaluate SCD research for participants with ASD has yet to be agreed upon by experts across the field, and contradictory recommendations that suggest either calculating a treatment effect score, or conducting visual analysis, have been made by different advisory panels within the broader educational psychology community. Given the apparent significant discrepancies in interpretation of meta-analytic reports on ABA based treatment research for participants with ASD conducted by leading US healthcare policy makers, we argue that it is important to the ASD research community to further examine SCD data with a view to determining how best *now* to calculate treatment effects while alternate improved procedures are being developed. We argue that the ASD community currently requires a suitable method to evaluate SCD intervention research to inform an evidence base, policy guidelines, and educational and clinical practice in this interim.

Accordingly, the purpose of this research project was to gauge the feasibility of calculating a treatment effect score using SCD data specifically in the context of participants with ASD. In their literature review, The National Standards Report (2009) has classified treatments as established, emerging, or unestablished. To compare the data collection trends of researchers working with participants on the autism spectrum, we have selected self-management interventions as an example of an established treatment, and physical activity as an example of an emerging treatment. The following research questions were developed:

1. Do studies report a sufficient number of baseline and intervention data points to enable the calculation of a treatment effect score?
2. Are there trends in these data suggesting this pattern is changing?

Method

Locating Studies

Studies were located by conducting a systematic search of peer-reviewed literature published prior to November 2013. Both the PsycINFO and ERIC databases were queried using the keywords “autism*” and “Asperger’s syndrome.” For self-management interventions, the following terms were queried: “self-management,” “self-regulation,” “self-regulate,” “self-monitoring,” “self-recording,” “self-reinforcement,” “self-evaluation,” “self-advocacy,” “self-observation,” “self-instruction,” “empowerment,” “self-determination,” and “self-control.” For physical activity interventions, the following

terms were queried: “physical activity,” “exercise,” and “fitness.” In addition, a hand search of the reference lists of existing systematic reviews on both self-management and exercise was undertaken.

The abstract of each article was examined to determine whether an article met inclusion criteria for further review, and the original article was retrieved and reviewed when necessary. No age limits were placed upon participants. Inclusion criteria required that:

1. Participants had an existing diagnosis of ASD or AS. In instances in which several participants with various conditions were included in a single article, only participants with either an ASD or AS were included for further review.
2. The study utilized a single subject research design such as a multiple baseline, reversal, changing criterion, or alternating treatment design.
3. The study presented data from each phase in graphical format for each participant thereby enabling the calculation of a treatment effect.
4. Components of self-management or exercise were included throughout the intervention.
5. Articles were published in English in a peer-reviewed journal.

This search procedure identified 38 articles that utilized SCDs in self-management interventions and a further eight articles that utilized SCDs for exercise interventions. Two studies appear in both self-management and exercise searches, in interventions that targeted participation in physical activity that also used self-management procedures (Todd & Reid, 2006; Todd, Reid, & Butler-Kisber, 2010).

Data Requirements for Treatment Effect Calculations

1. Regression-based approach: A minimum of six data points per baseline or treatment phase and at least 14 data points in a phase A and B comparison
2. PND: No minimum number of data points required; however, baseline stability must be evident
3. PAND: A minimum of 20 data points across baseline and treatment phases required
4. NAP: No minimum number of data points specified

Reliability of Data

Inter-coder agreement was calculated with the first and one of the co-authors separately coding each study before comparing results. Initial trial coding was performed using three studies to ensure consistency between assessors.

A 30 % random sample of abstracts from all search results was reviewed independently, to determine the reliability of the article selection process. Inter-coder agreement was determined by dividing the number of agreements by the total number of agreements plus disagreements and multiplying by 100. Inter-coder agreement was 97 % for self-management interventions and 100 % for exercise interventions.

Once the data set was developed, the accuracy of the data point count procedure was checked. The first author randomly selected a 50 % sample of included studies, and a co-author independently counted data points in baseline, treatment, and any subsequent phases. These results were then compared to the first author's counts. Inter-coder assessment was 98 % for self-management interventions and 100 % for exercise interventions.

Results

Thirty eight self-management intervention articles included in the data set reported treatment data from a variety of behaviors and settings for 102 participants. Given that many treatments were repeated across either behaviors or settings, a total of 215 data series were included in these graphs. A further eight exercise intervention articles reported treatment data for an additional 20 participants, and the corresponding graphs included 43 data series.

Hand counts of the number of data points reported in each baseline and treatment phase were conducted for each of the data series included in the graphs. These tallies were recorded manually into an Excel spreadsheet and subsequently compared to the advised minimal baseline and treatment phase data requirements for regression-based procedures, PND, PAND, and NAP to determine the feasibility of using these procedures with these data sets. Table 1 provides a summary of the results of each comparison.

The feasibility of the application of regression-based calculations was determined via the adherence to a minimum of six data points in the first baseline phase, six data points in the first treatment phase, and a minimum total of 14 data points across the first AB comparison. With this threshold, 97 of the 215 self-management data series (45.1 %) provided sufficient data. This is illustrated graphically in Fig. 1. In order to calculate a treatment effect for an entire study, each individual data series should meet the minimum data threshold. Given this constraint, only nine self-management articles (23.7 %) had sufficient data for a regression-based calculation. When considered on a participant basis, we identified 36 individuals (35.3 %) for whom a sufficient volume of data was reported for each data series to permit a regression-based treatment effect calculation per participant.

For exercise interventions, 11 of the 43 data series (25.6 %) met the minimum total of data points across the first AB

comparison. This is illustrated graphically in Fig. 2. When the exercise intervention articles were viewed in their entirety, a single article (12.5 %) included sufficient data to enable the calculation of treatment effect for the entire study. When considered on a participant basis, we identified four individuals (20.0 %) for whom sufficient data is provided to enable a regression-based treatment effect calculation per participant.

Within the self-management studies, ceiling or floor data points in baseline occurred in three data series that resulted in a 0 % PND calculation. Variability in baseline data for one study that included two participants indicated that a PND should not be coded. Accordingly, data from 34 of the 38 studies (89.5 %) appear sufficient for a PND calculation. When viewed on a participant level, 96 of 102 (94.1 %) of participant data appear sufficient for a PND calculation.

For the exercise interventions, floor data points in baseline occurred in one study that reported one data series for a single participant. For this study, a 0 % PND would be calculated. Overall, seven of the eight studies (87.5 %) report sufficient data for a PND calculation. At a participant level, data sufficient for a PND calculation is available for 19 of 20 (95.0 %) participants.

The calculation of PAND requires a minimum of 20 data points per data series across baseline and treatment phases. Using this guideline, 117 self-management data series (54.4 %) reported a sufficient number of data points. When each article was considered overall, 22 self-management articles (57.9 %) included data series of an adequate length for the application of PAND. Examination of the self-management data on a per participant basis revealed that PAND was appropriate to apply to 53 individuals (52.0 %).

For exercise interventions, 26 data series (60.5 %) reported a sufficient number of data points. Viewed overall by article, five exercise articles (62.5 %) included data series of an adequate length to apply the PAND calculation. Examination of the exercise data on a per participant basis revealed that PAND was appropriate to apply to 13 individuals (65.0 %).

Although no minimum threshold was reported for the application of the NAP calculation, Parker and Vannest (2009) reported the median length of the full data series of 200 selected AB contrasts was 18 data points. By way of comparison, the self-management data set of 215 data series had a median length of 13.75 data points for the first AB comparison and a median length of 25.5 data points across the full data series. For exercise interventions, the 43 data series had a median length of 13 data points for the first AB comparison and 25 data points across the full data series.

The length of the data series including baseline and treatment data that was reported in each self-management and exercise article was plotted over time and a line of best fit was calculated for the data. A split-middle line of progress (Cooper, Heron, & Heward 2007) was plotted and indicated a declining trend over time in the number of data points collected (see Fig. 3).

Table 1 Feasibility of treatment effect score calculation for studies

Authors	No. of data series graphs	No. of participants	Sufficient data for regression	Sufficient data for PND	Sufficient data for PAND	Sufficient data for NAP
Self-management studies, <i>n</i> =38						
Koegel and Koegel (1990)	7	4	No	Yes	No	Yes
Sainato et al. (1990)	3	3	No	Yes	Yes	Yes
Koegel et al. (1992)	14	4	No	Yes ^a	Yes	Yes
Stahmer and Schreibman (1992)	6	3	Yes	Yes ^b	Yes	Yes
Koegel and Frea (1993)	6	2	Yes	No ^c	Yes	Yes
Pierce and Schreibman (1994)	18	3	No	Yes	No	Yes
Strain, Kohler, Storey, & Danko (1994)	10	3	Yes	Yes	Yes	Yes
Newman et al. (1995)	3	3	No	Yes	Yes	Yes
Strain and Danko (1995)	3	3	No	Yes	Yes	Yes
Dixon et al. (1995)	2	1	No	Yes	No	Yes
Newman et al. (1996)	3	3	No	Yes	Yes	Yes
Shearer et al. (1996)	3	3	No	Yes ^d	No	Yes
Kern et al. (1997)	3	1	No	Yes	No	Yes
Callahan and Rademacher (1999)	2	1	Yes	Yes	Yes	Yes
Reinecke et al. (1999)	3	3	Yes	Yes	Yes	Yes
Mancina et al. (2000)	6	1	Yes	Yes	Yes	Yes
Newman et al. (2000)	3	3	Yes	Yes	Yes	Yes
Shabani et al. (2001)	2	1	No	Yes	No	Yes
Morrison et al. (2001)	16	4	Yes	Yes	Yes	Yes
Mithaug and Mithaug (2003)	2	2	No	Yes	Yes	Yes
Wehmeyer et al. (2003)	2	1	No	Yes	Yes	Yes
Apple et al. (2005)	8	3	No	Yes	No	Yes
Newman and Ten Eyck (2005)	3	3	No	Yes	Yes	Yes
Ganz and Sigafoos (2005)	1	1	No	Yes	Yes	Yes
Agran et al. (2005)	2	2	No	Yes	Yes	Yes
Todd and Reid (2006)	3	3	No	Yes	Yes	Yes
Delano (2007)	6	3	No	Yes	No	Yes
Loftin et al. (2008)	3	3	No	Yes	No	Yes
Palmen et al. (2008)	9	9	No	Yes	No	Yes
Dorminy et al. (2009)	8	4	No	Yes	No	Yes
Soares et al. (2009)	1	1	Yes	Yes	Yes	Yes
Deitchman et al. (2010)	2	3	No	Yes	No	Yes
Asaro-Saddler and Saddler (2010)	9	3	No	Yes	No	Yes
Cihak et al. (2010)	18	3	No	Yes	No	Yes
Holifield et al. (2010)	8	2	No	Yes	No	Yes
Todd et al. (2010)	3	3	No	Yes	Yes	Yes
Parker and Kamps (2010)	12	2	No	Yes	No	Yes
Shogren et al. (2011)	2	2	No	Yes	Yes	Yes
	209	102	9	34	22	38
Percentage of studies with sufficient data points			23.7 %	89.5 %	57.9 %	100.0 %
Exercise studies, <i>n</i> =8						
Kern et al. (1982)	6	4	No	Yes	No	Yes
Allison et al. (1991)	1	1	Yes	Yes ^e	Yes	Yes
Celiberti et al. (1997)	3	1	No	Yes	No	Yes
Todd and Reid (2006)	3	3	No	Yes	Yes	Yes
Nicholson et al. (2011)	12	4	No	Yes	Yes	Yes
Todd et al. (2010)	3	3	No	Yes	Yes	Yes

Table 1 (continued)

Authors	No. of data series graphs	No. of participants	Sufficient data for regression	Sufficient data for PND	Sufficient data for PAND	Sufficient data for NAP
Schmidt et al. (2013)	6	1	No	Yes	Yes	Yes
Yanardag et al. (2013)	9	3	No	Yes	No	Yes
	43	20	1	7	5	8
Percentage of studies with sufficient data points			12.5 %	87.5 %	62.5 %	100.0 %

^a Ceiling data point in baseline for Howard community setting assumed normal baseline variability, PND coded as 0 %

^b Target behavior data appropriate for PND on all participants for appropriate play data. Additional measurements taken for self-stimulatory behavior given detrimental effects on appropriate play; Bruce and Justin PND coded as 0 %

^c Variability in baseline data, unable to code PND

^d Ceiling data points in baseline for Child One engagement task assumed in line with baseline variability; PND coded as 0 %

^e Floor data points in baseline; PND coded as 0 %

A line of best fit was also plotted using Excel and produced a trend line that closely paralleled that calculated using the quarter-intersect and split-middle line of progress methods, hence confirming this declining trend. The Excel calculation derived the equation $y = -0.1339x + 49.558$ for this trend line. The graph illustrates that older research papers reported a greater volume of observational data than more recent articles.

Discussion

Our results show that for 64 of the 102 participants included in the self-management interventions, and 16 of the 20

participants included in the exercise interventions, the number of data points in the first AB phase comparison was below the required minimum for regression-based estimates of treatment effect sizes.

While not always described as the target behavior of an intervention, a variety of problem behaviors or unacceptable performances were described for these participants. Examples of these behaviors included physical aggression towards others, elopement for the purpose of engaging in ritualistic behaviors, self-injury, non-compliance, loud screaming, psychotic speech, inappropriate touching or hugging, destruction of property, head banging, placing non-edibles in mouth, tantrums, threats towards others, prolonged crying or body

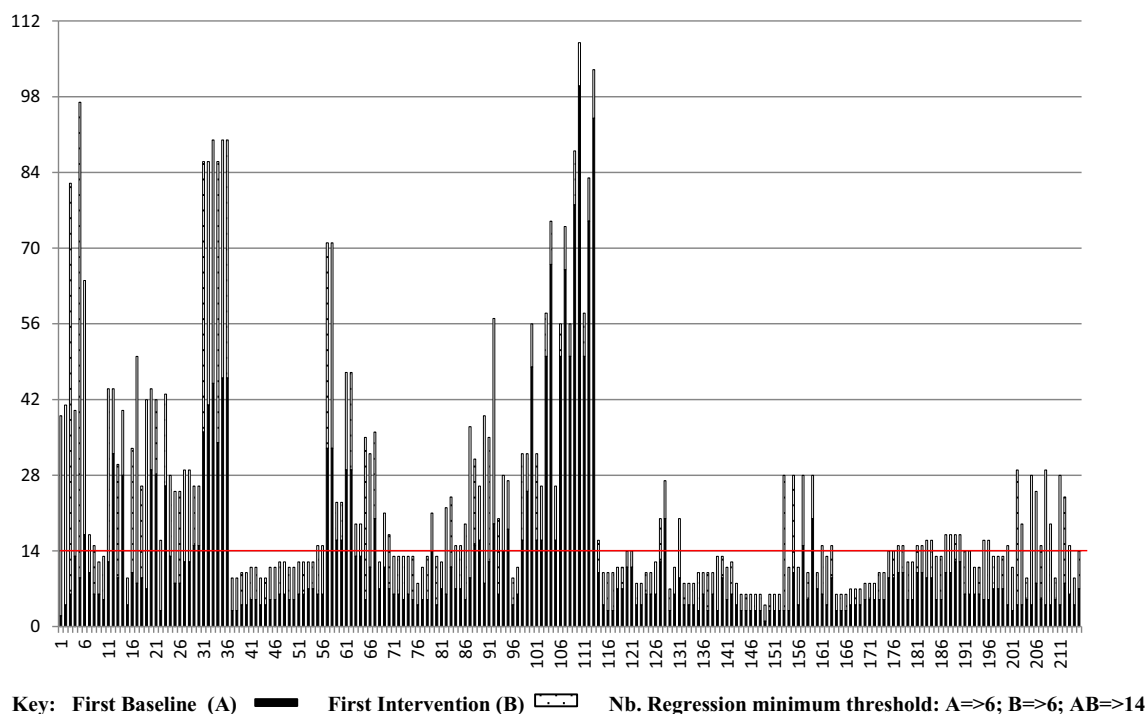


Fig. 1 Self-management data points

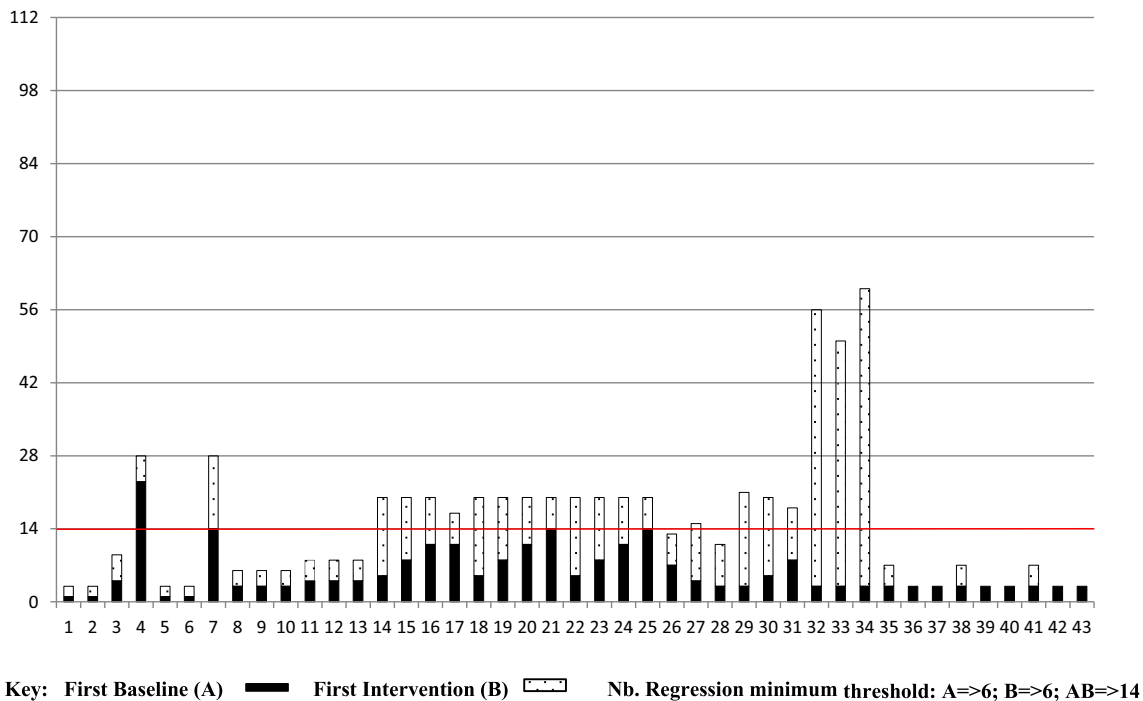


Fig. 2 Exercise data points

rocking, and social withdrawal. Unacceptable levels of classroom engagement or academic performance were also described. In some instances, collateral reduction in these challenging behaviors was recorded or observed in interventions that targeted the development of a skill.

Problem behaviors such as those described present an ethical dilemma to researchers as the collection of lengthy

baseline data for these participants might be considered unacceptable practice. Significantly, the data series collected in both established and emerging treatments for participants with autism are often limited in length. The nature of the behaviors frequently under investigation with this population is at odds with the ability to apply a regression-based calculation to determine the strength of treatment effects. Furthermore, for

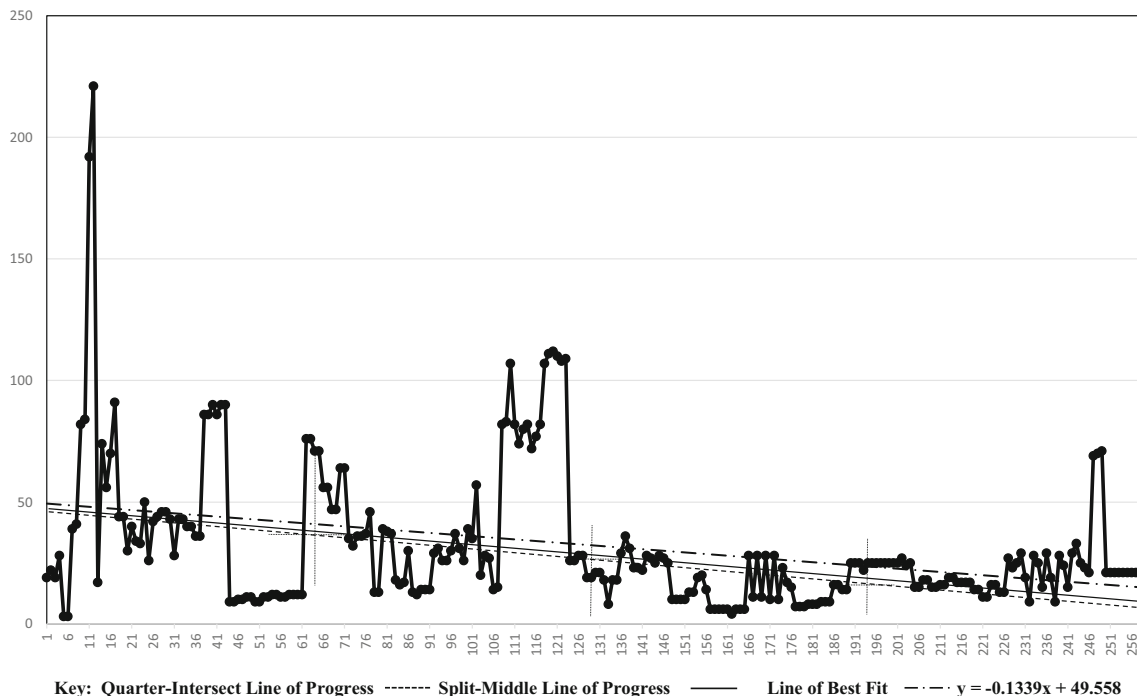


Fig. 3 Total data points collected over time for self-management and exercise interventions

the studies included in this dataset, we have identified a declining trend in the volume of data points that are collected in baseline and treatment phases. This downward trend, in addition to the nature of behaviors often described among participants on the autism spectrum, suggests that it is unlikely that regression-based calculation methods will be appropriate in future research. Alternative methods that are appropriate where shorter data sets are the norm appear advantageous.

Previous criticism of PND has included claims that outlying data in baseline phases result in calculations that do not accurately reflect the success of an original study. However, examination of both self-management and exercise extant data sets reveals that there are relatively few occasions in which a 0 % PND is calculated or that baseline variability results in an inability to code a study. Our findings support recent claims by Scruggs and Mastropieri (2013) that PND continues to produce treatment effect scores that accurately represent the findings described by original authors.

By contrast, PAND appears appropriate for slightly less than two thirds of the studies included in this review. While the literature has suggested merit in adopting this calculation as an alternative to PND, this benefit appears to be offset by the reduced utility of the method as a result of insufficient data collection for participants with ASD.

NAP is not constrained by the volume of data points collected in intervention and consequently is appropriate to apply to all studies included in both self-management and exercise SCD research. However, as the calculation is based on a comparison of all pairs contained within baseline and treatment data, a somewhat more cumbersome hand calculation is required than that of PND or PAND.

Conclusion

We have explored the volume of SCD data collected in both an established and an emerging treatment for participants on the autism spectrum. Relatively short data series were frequently reported, and results of this study suggest a declining trend in the length of data series reported over time, with older studies including a greater volume of data when compared to more recent studies. Behavioral challenges were described for many of the participants; consequently, the collection of additional data, particularly extended baseline phases, may pose an ethical dilemma to researchers, clinicians, or other stakeholders. It appears unlikely that data yielded from such applied research in the future will provide the longer data series necessary for more complex treatment effect calculations. Accordingly, while a regression-based calculation is arguably more accurate, the nature of the data examined in this study appears at odds with this method. A nonparametric approach may be preferable in the calculation of effect sizes in research involving participants with ASD.

The feasibility of three nonparametric hand calculations was explored. Both PAND and NAP are considered by many to offer a potentially superior calculation to PND. However, our findings suggest that PAND cannot always be calculated given the volume of data points that are typically reported. Excluding studies with limited baseline data points from effect size calculations may distort meta-analysis findings. While the NAP procedure is unrestricted in this regard, the reality of conducting this more cumbersome calculation may present a barrier to many stakeholders to the adoption of this method in terms of the time involved, the increased chance of calculation error, and of erroneous interpretation of the treatment effect score that is produced. The simpler PND calculation appears appropriate for the majority of data sets included in this study, and relatively few instances in which outlier ceiling or floor data points in baseline data obscure the true treatment effects were identified.

Further, while acknowledging the inability of PND to differentiate between demonstrably powerful treatment effects, we observe that baseline variability systematically reduces the product of a PND calculation, thereby making the calculation inherently more conservative as a measure of strength of treatment effects. A calculation that is complementary to visual examination and does not require extensive additional training is highly desirable. The continued use of a PND calculation is arguably advantageous given that it has been widely applied and hence facilitates comparison with previous research findings. The PND calculation is also advantageous in that it can be calculated and interpreted with limited additional training, by a variety of stakeholders including clinicians, teachers, and parents. In light of the findings from this study, the continued use of PND appears justified though further research leading to the development of more robust and, at the same time, sensitive procedures is clearly warranted.

References

*Studies included in the review

- *Agran, M., Alper, S., Cavin, M., Sinclair, T., Wehmeyer, M., & Hughes, C. (2005). Using self-monitoring to increase following-direction skills of students with moderate to severe disabilities in general education. *Education and Training in Developmental Disabilities*, 40, 3–13
- *Allison, D. B., Basile, V. C., & MacDonald, R. B. (1991). Brief report: Comparative effects of antecedent exercise and lorazepam on the aggressive behavior of an autistic man. *Journal of Autism and Developmental Disorders*. doi:10.1007/bf02207001
- Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: the case of the single case. *Behaviour Research and Therapy*. doi:10.1016/0005-7967(93)90115-B.

- *Apple, A. L., Billingsley, F., & Schwartz, I. S. (2005). Effects of video modeling alone and with self-management on compliment-giving behaviors of children with high-functioning ASD. *Journal of Positive Behavior Interventions*. doi:10.1177/10983007050070010401
- *Asaro-Saddler, K., & Saddler, B. (2010). Planning instruction and self-regulation training: Effects on writers with autism spectrum disorders. *Exceptional Children*. doi:10.1177/1053451208330895
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*. doi:10.1901/jaba.1968.1-91.
- Blampied, N. M. (1999). A legacy neglected: Restating the case for single-case research in cognitive-behaviour therapy. *Behaviour Change*. doi:10.1375/behc.16.2.89.
- *Callahan, K., & Rademacher, J. A. (1999). Using self-management strategies to increase the on-task behavior of a student with autism. *Journal of Positive Behavior Interventions*. doi:10.1177/109830079900100206
- Campbell, J. M. (2013). Commentary on PND at 25. *Remedial and Special Education*. doi:10.1177/0741932512454725.
- CDC MMWR. (2012). Prevalence of autism spectrum disorders—autism and developmental disabilities monitoring network, 14 sites, United States, 2008. *Centers for Disease Control and Prevention*, 61(3), 1–19.
- *Celiberti, D. A., Bobo, H. E., Kelly, K. S., Harris, S. L., & Handleman, J. S. (1997). The differential and temporal effects of antecedent exercise on the self-stimulatory behavior of a child with autism. *Research in Developmental Disabilities*. doi:10.1016/s0891-4222(96)00032-7
- *Cihak, D. F., Wright, R., & Ayres, K. M. (2010). Use of self-modeling static-picture prompts via a handheld computer to facilitate self-monitoring in the general education classroom. *Education and Training in Autism and Developmental Disabilities*, 45, 136–149
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River: Pearson Merrill Prentice Hall.
- *Deitchman, C., Reeve, S. A., Reeve, K. F., & Progar, P. R. (2010). Incorporating video feedback into self-management training to promote generalization of social initiations by children with autism. *Education and Treatment of Children*. doi:10.1353/etc.0.0102
- *Delano, M. E. (2007). Improving written language performance of adolescents with Asperger syndrome. *Journal of Applied Behavior Analysis*. doi:10.1901/jaba.2007.50-06
- *Dixon, R. S., Moore, D. W., Hartnett, N., Howard, R., & Petrie, K. (1995). Reducing inappropriate questioning behaviour in an adolescent with autism: A case study. *Behaviour Change*, 12, 163–166
- *Dorminy, K. P., Luscre, D., & Gast, D. L. (2009). Teaching organizational skills to children with high functioning autism and Asperger's syndrome. *Education and Training in Developmental Disabilities*, 44, 538–550
- *Ganz, J. B., & Sigafoos, J. (2005). Self-monitoring: Are young adults with MR and autism able to utilize cognitive strategies independently? *Education and Training in Developmental Disabilities*, 40 (1), 24–33
- Green, V. A., Pituch, K., Itchon, J., Choi, A., O'Reilly, M., & Sigafoos, J. (2006). Internet survey of treatments used by parents of children with autism. *Research in Developmental Disabilities*. doi:10.1016/j.ridd.2004.12.002.
- *Holifield, C., Goodman, J., Hazekorn, M., & Heflin, L. J. (2010). Using self-monitoring to increase attending to task and academic accuracy in children with autism. *Focus on Autism and Other Developmental Disabilities*. doi:10.1177/1088357610380137
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S. L., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*. doi:10.1177/001440290507100203.
- 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*. doi:10.1353/etc.2012.0011.
- Kazdin, A. E. (1978). Methodological and interpretive problems of single-case experimental designs. *Journal of Consulting and Clinical Psychology*. doi:10.1037/0022-066X.46.4.629.
- *Kern, L., Koegel, R. L., Dyer, K., Blew, P. A., & Fenton, L. R. (1982). The effects of physical exercise on self-stimulation and appropriate responding in autistic children. *Journal of Autism and Developmental Disorders*. doi:10.1007/bf01538327
- *Kern, L., Marder, T. J., Boyajian, A. E., Elliot, C. M., & McElhattan, D. (1997). Augmenting the independence of self-management procedures by teaching self-initiation across settings and activities. *School Psychology Quarterly*. doi:10.1037/h0088944
- *Koegel, L. K., Koegel, R. L., Hurley, C., & Frea, W. D. (1992). Improving social skills and disruptive behavior in children with autism through self-management. *Journal of Applied Behavior Analysis*. doi:10.1901/jaba.1992.25-341
- *Koegel, R. L., & Frea, W. D. (1993). Treatment of social behavior in autism through the modification of pivotal social skills. *Journal of Applied Behavior Analysis*. doi:10.1901/jaba.1993.26-369
- *Koegel, R. L., Koegel, L. (1990). Extended reductions in stereotypic behavior of students with autism through a self-management package. *Journal of Applied Behavior Analysis*. doi:10.1901/jaba.1990.23-119
- Kratochwill, T. R., & Brody, G. H. (1978). Single subject designs: A perspective on the controversy over employing statistical inference and implications for research and training in behavior modification. *Behavior Modification*. doi:10.1177/014544557823001.
- Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). What Works Clearinghouse Procedures and Standards Handbook (Version 3.0).
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). WWC Single-case designs technical documentation Version 1.0
- Kratochwill, T. R., & Levin, J. R. (2014). Meta- and statistical analysis of single-case intervention research data: Quantitative gifts and a wish list. *Journal of School Psychology*. doi:10.1016/j.jsp.2014.01.003.
- Leland Wilkinson and the Taskforce on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*. doi:10.1037/0003-066X.54.8.594.
- *Loflin, R. L., Odom, S. L., & Lantz, J. F. (2008). Social interaction and repetitive motor behaviors. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-007-0499-5
- *Mancina, C., Tankersley, M., Kamps, D., Kravits, T., & Parrett, J. (2000). Brief report: Reduction of inappropriate vocalizations for a child with autism using a self-management treatment program. *Journal of Autism and Developmental Disorders*. doi:10.1023/a:1005695512163
- Matyas, T., & Greenwood, K. (1990). Visual analysis of single-case time series: Effects of variability, serial dependence, and magnitude of intervention effects. *Journal of Applied Behavior Analysis*. doi:10.1901/jaba.1990.23-341.
- *Mithaug, D. K., & Mithaug, D. E. (2003). Effects of teacher-directed versus student-directed instruction on self-management of young children with disabilities. *Journal of Applied Behavior Analysis*. doi:10.1901/jaba.2003.36-133
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*. doi:10.1371/journal.pmed.1000097.
- *Morrison, L., Kamps, D., Garcia, J., & Parker, D. (2001). Peer mediation and monitoring strategies to improve initiations and social skills for students with autism. *Journal of Positive Behavior Interventions*. doi:10.1177/109830070100300405

- National Standards Report. (2009). *National Autism Center*, Randolph MA.
- National Association of Insurance Commissioners, Top Health Insurance Companies (2013). Retrieved April 8, 2014, from <http://health.usnews.com/health-news/health-insurance/articles/2013/12/16/top-health-insurance-companies>.
- *Newman, B., Buffington, D., & Hemmes, N. (1996). Self-reinforcement used to increase the appropriate conversation of autistic teenagers. *Education and Training in Mental Retardation and Developmental Disabilities*, 31, 304–309
- *Newman, B., Buffington, D., O' Grady, M., McDonald, M., Poulson, C., & Hemmes, N. (1995). Self-management of schedule following in three teenagers with autism. *Behavioral Disorders*, 20, 190–196
- *Newman, B., Reinecke, D. R., & Meinberg, D. L. (2000). Self-management of varied responding in three students with autism. *Behavioral Interventions*. doi:10.1002/(SICI)1099-078X(200004/06)15:2<145::AID-BIN50>3.0.CO;2-3
- *Newman, B., & Ten Eyck, P. (2005). Self-management of initiations by students diagnosed with autism. *The Analysis of Verbal Behavior*, 21, 117–122
- *Nicholson, H., Kehle, T. J., Bray, M. A., & Heest, J. V. A. N. (2011). The effects of antecedent physical activity on the academic engagement of children with autism spectrum disorder. *Psychology in the Schools*. doi:10.1002/pits
- *Palmen, A., Didden, R., & Arts, M. (2008). Improving question asking in high-functioning adolescents with autism spectrum disorders: Effectiveness of small-group training. *Autism: The International Journal of Research and Practice*. doi:10.1177/1362361307085265
- *Parker, D., & Kamps, D. (2010). Effects of task analysis and self-monitoring for children with autism in multiple social settings. *Focus on Autism and Other Developmental Disabilities*. doi:10.1177/1088357610376945
- Parker, R. I., & Brossart, D. F. (2003). Evaluating single-case research data: A comparison of seven statistical methods. *Behavior Therapy*. doi:10.1016/s0005-7894(03)80013-8.
- Parker, R. I., Brossart, D. F., Vannest, K. J., Long, J. R., De-Alba, R. G., Baugh, F. G., & Sullivan, J. R. (2005). Effect sizes in single case research: How large is large? *School Psychology Quarterly*, 34, 116–132.
- Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND): An alternative to PND. *The Journal of Special Education*. doi:10.1177/00224669070400040101.
- Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*. doi:10.1016/j.beth.2008.10.006.
- *Pierce, K. L., & Schreibman, L. (1994). Teaching daily living skills to children with autism in unsupervised settings through pictorial self-management. *Journal of Applied Behavior Analysis*. doi:10.1901/jaba.1994.27-471
- *Reinecke, D. R., Newman, B., & Meinberg, D. L. (1999). Self-management of sharing in three preschoolers with autism. *Education and Training in Mental Retardation and Developmental Disabilities*, 34, 312–317
- *Sainato, D. M., Strain, P. S., Lefebvre, D., & Rapp, N. (1990). Effects of self-evaluation on the independent work skills of preschool children with disabilities. *Exceptional Children*, 56, 540–549
- *Schmidt, J. D., Luiselli, J. K., Rue, H., & Whalley, K. (2013). Graduated exposure and positive reinforcement to overcome setting and activity avoidance in an adolescent with autism. *Behavior Modification*. doi:10.1177/0145445512456547
- Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: Issues and applications. *Behavior Modification*. doi:10.1177/01454455980223001.
- Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25: Past, present, and future trends in summarizing single-subject research. *Remedial and Special Education*. doi:10.1177/0741932512440730.
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education*. doi:10.1177/074193258700800206.
- *Shabani, D. B., Wilder, D. A., & Flood, W. A. (2001). Reducing stereotypic behavior through discrimination training, differential reinforcement of other behavior, and self-monitoring. *Behavioral Interventions*. doi:10.1002/bin.096
- Shadish, W. R. (2014). Analysis and meta-analysis of single-case designs: An introduction. *Journal of School Psychology*. doi:10.1016/j.jsp.2013.11.009.
- *Shearer, D. D., Kohler, F. W., Buchan, K. A., & McCullough, K. M. (1996). Promoting independent interactions between preschoolers with autism and their nondisabled peers: An analysis of self-monitoring. *Early Education & Development*. doi:10.1027/s15566935eed0703_1
- *Shogren, K., Lang, R., Machalicek, W., Rispoli, M., & O'Reilly, M. (2011). Self- versus teacher management of behavior for elementary school students with Asperger syndrome: Impact on classroom behavior. *Journal of Positive Behavior Interventions*. doi:10.1177/1098300710384508
- Sigafoos, J., & Schlosser, R. (2008). Applied behavior analysis is NOT an autism therapy. *Evidence-Based Communication Assessment and Intervention*. doi:10.1080/17489530802640391.
- *Soares, D., Vannest, K., & Harrison, J. (2009). Computer aided self-monitoring to increase academic production and reduce self-injurious behavior in a child with autism. *Behavioral Interventions*. doi:10.1002/bin.283
- *Stahmer, A. C., & Schreibman, L. (1992). Teaching children with autism appropriate play in unsupervised environments using a self-management treatment package. *Journal of Applied Behavior Analysis*. doi:10.1901/jaba.1992.25-447
- *Strain, P. S., & Danko, C. D. (1995). Caregivers' encouragement of positive interaction between preschoolers with autism and their siblings. *Journal of Emotional and Behavioral Disorders*. doi:10.1177/106342669500300101
- *Strain, P. S., Kohler, F. W., Storey, K., & Danko, C. D. (1994). Teaching preschoolers with autism to self-monitor their social interactions: An analysis of results in home and school settings. *Journal of Emotional and Behavioral Disorders*. doi:10.1177/106342669400200202
- *Todd, T., & Reid, G. (2006). Increasing physical activity in individuals with autism. *Focus on Autism and Other Developmental Disabilities*. doi:10.1177/10883576060210030501
- *Todd, T., Reid, G., & Butler-Kisber, L. (2010). Cycling for students with ASD: Self-regulation promotes sustained physical activity. *Adapted Physical Activity Quarterly*, 27, 226–41
- *Wehmeyer, M. L., Yeager, D., Bolding, N., Agran, M., & Hughes, C. (2003). The effects of self-regulation strategies on goal attainment for students with developmental disabilities in general education classrooms. *Journal of Developmental and Physical Disabilities*. doi:10.1023/a:1021408405270
- *Yanardag, M., Akmanoglu, N., & Yilmaz, I. (2013). The effectiveness of video prompting on teaching aquatic play skills for children with autism. *Disability and rehabilitation*. doi:10.3109/09638288.2012.687030

CHAPTER 4: PUBLISHED PAPER 2

A sensitivity analysis of three nonparametric treatment effect scores for single-case research for participants with autism. *Review Journal of Autism and Developmental Disorders* (2014)

Monica E. Carr. DOI:10.1007/s40489-014-0037-2

A Sensitivity Analysis of Three Nonparametric Treatment Effect Scores for Single-Case Research for Participants with Autism

Monica E. Carr

Review Journal of Autism and Developmental Disorders

ISSN 2195-7177

Rev J Autism Dev Disord
DOI 10.1007/s40489-014-0037-2



Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

A Sensitivity Analysis of Three Nonparametric Treatment Effect Scores for Single-Case Research for Participants with Autism

Monica E. Carr

Received: 14 September 2014 / Accepted: 15 October 2014
© Springer Science+Business Media New York 2014

Abstract The widely used percentage of nonoverlapping data (PND) treatment effect calculation was compared to more recently developed methods which, it has been argued, better account for outlying variables and trend in single-case design (SCD) intervention studies. Percentage of all nonoverlapping data (PAND) and nonoverlap of all pairs (NAP) were selected for comparison as both are amenable to hand calculation, making them widely accessible to clinicians and teachers as well as researchers. A data set was developed through a systematic search of peer-reviewed literature on self-management interventions conducted with participants with autism spectrum disorder (ASD). Treatment effect sizes derived from each method were compared for studies that provided sufficient data. Results indicated that PND provided a conservative measure of strength of treatment effect when compared to PAND and NAP scores. Interpretation scales for treatment effect scores derived from each method were reviewed. Implications for selecting a calculation method for participants with ASD are discussed.

Keywords Autism · Treatment effect · Nonparametric · Hand calculation · PND

Introduction

In 1994, the American Psychological Association (APA) publication manual encouraged the inclusion of a treatment effect size in research reports. Since that time, the APA Taskforce on Statistical Inference (1999) argued that a treatment effect size permits the evaluation of the stability of findings across samples and is important to future meta-analyses. Subsequently, in

some quarters, treatment effect scores have been considered a requirement for research publication (Leland Wilkinson and the Taskforce on Statistical Inference 1999).

The initial report developed by the APA Taskforce on Statistical Inference (1996) warned that, with advances in state-of-the-art statistical analysis software, statistics are commonly reported without understanding of the computational methods or necessarily even an understanding of what the statistics mean. Parker and colleagues have also argued that treatment effect calculations reported in meta-analysis of single-case design (SCD) research should be interpretable by various different stakeholder groups (Parker et al. 2005), a point well illustrated with autism intervention-related research where parents, educators and policy makers as well as clinicians and researchers all need to understand reports on treatment effects.

Given the increasing demand to develop an evidence base in educational psychology, quality assessment guidelines have been developed by the US Department of Education What Works Clearinghouse (WWC) (Kratochwill et al. 2010; Kratochwill et al. 2013a; Kratochwill et al. 2013b). Various methods for determining treatment effects have been proposed for use in meta-analyses of SCD research, although the merits of these different computation methods remain a matter of debate (Horner and Kratochwill 2012; Horner et al. 2012; Kratochwill et al. 2013b; Scruggs and Mastropieri 2013). Unlike group research designs, a generally accepted method for the calculation of treatment effect size for SCD research has yet to be established. The initial version of the WWC SCD guidelines indicated a preference for regression based procedures for calculating effect sizes though the WWC panel subsequently suggested conducting a sensitivity analysis of treatment effect scores using several indices. Most recently, the WWC has moved away from the use of a treatment effect score and reverted to visual analyses until a general consensus on the most appropriate method has been reached (Kratochwill et al. 2013b).

M. E. Carr (✉)
Monash University, Melbourne, VIC 3800, Australia

There is an extensive body of literature examining approaches to the evaluation of the strength of treatment effects for SCD research. Shadish (2014) has reported that a number of new methods are currently in development including standardised mean difference approaches, multilevel analysis and Bayesian generalised least squares procedures. In a recent review of SCD research conducted with students with a broad array of disabilities, Maggin et al. (2011a) reported percentage of nonoverlapping data (PND) as the most frequently used treatment effect score appearing in 55 % of the 84 effect sizes garnered from 68 literature syntheses. Maggin and colleagues also reported that nearly 40 % of authors did not describe the method they used for comparing data from various phases within studies when estimating effect sizes. Of the studies that did include these details, several approaches were described. These included taking an arithmetic mean across all baseline and treatment phases, selecting only certain phases (i.e., A1B2) and consolidating baselines and treatment phases (i.e., A1A2B1B2).

The PND (Scruggs et al. 1987) was developed as a method to synthesise SCD literature which could be easily calculated and readily as well as meaningfully interpreted. Several positive features of PND have been described and include ease of calculation from graphical rather than raw data, high degree of inter-rater reliability, applicability to any SCD design type and ease of interpretation (Campbell 2013; Parker et al. 2007). The continued utility of the procedure has recently been argued by the original developers (Scruggs and Mastropieri 2013). However, PND is not without its critics, and major limitations of the procedure include the following: (i) PND requires its own interpretation guidelines as it does not correlate to an accepted effect size; (ii) it lacks sensitivity in discriminating treatment effectiveness as the calculated score approaches 100 %; (iii) PND is reliant on a single extreme data point in baseline, and all other baseline data are excluded from the calculation; and (iv) as PND has no known sampling distribution, confidence intervals cannot be calculated (Parker et al. 2007).

All current alternative procedures have their own limitations however. Maggin et al. (2011b) compared 11 commonly used effect size measures including three parametric methods: interrupted time series analysis procedure (ITSACORR), piecewise regression, hierarchical linear modelling; and seven nonparametric methods: PND, percentage of all nonoverlapping data (PAND), percent of zero data points (PZD), pairwise overlap squared (PDO²), percentage exceeding mean (PEM), percentage exceeding mean trend line (PEM-T), improvement rate difference (IRD) and the standardised mean difference. Of the nonparametric methods that were assessed, PAND received the most favourable assessment.

PAND (Parker et al. 2007) has been presented as an alternative to PND, the developers recommending it for documentation and accountability purposes in schools and clinics, in

addition to applications in meta-analyses and academic research reports. Their method was illustrated with sample data that typically contained between 60 and 80 data points, and the authors noted that the method was not well suited for data series that contained fewer than 20–25 data points. Parker et al. (2011) reported that PAND has been adopted in two meta-analyses along with a *phi* correlation coefficient, which is analogous to an R^2 score, that is frequently reported in large N studies (Parker et al. 2011).

Recently, Parker and Vannest (2009) developed the nonoverlap of all pairs (NAP) procedure, suggesting that this method offers an improvement on both PND and PAND. Several anticipated advantages have been suggested by these researchers, notably that the calculation uses all data points and as such should yield a more representative treatment effect score. Unlike other nonparametric indices, NAP is not based on means or medians, and it has been suggested that the calculation should relate more closely to the regression term R^2 . Importantly for stakeholders within the autism spectrum disorder (ASD) community, NAP can be calculated by hand. NAP was omitted from the effect size comparison conducted by Maggin and colleagues (2011b) however, Parker and colleagues (2011) reported that NAP has also been used in several recent meta-analyses.

Although greater consensus is evident between researchers regarding how to calculate treatment effects in group design research (Kratonchwill et al. 2013b), it has been noted that the interpretation of these treatment effect scores can also be problematic (Brossart et al. 2006). Brossart and colleagues observed that a basis for comparison of treatment effect sizes obtained using different calculations is lacking in the literature, arguably making this task challenging for clinicians and other stakeholders. These researchers noted that simple methods tend to yield different effect sizes than regression-based methods and that even ballpark interpretation guidelines for R^2 —e.g., “large” ($R^2=.25$), medium ($R^2=.09$) and small ($R^2=.01$)—drawn from large N group research in social science vary depending on the field of investigation.

Guidelines for interpretation of derived scores have been clearly defined for PND. However, other than the original developers describing a *phi* correlation coefficient for PAND, based upon a Pearson R 2×2 contingency table (Parker et al. 2007), it appears that an interpretation scale for PAND is not available in the literature to date. Parker and Vannest (2009) did provide a tentative interpretation scale for NAP, analogous to that used in PND, based on a process of expert judgements of 200 data sets.

In a recent exploration of the characteristics of SCD data for participants with ASD, Carr et al. (2014) reported a declining trend over time in the volume of data gathered in both baseline and treatment phases in SCD studies in both an *established treatment* (self-management) and an *emerging treatment* field (physical exercise) as classified by the

National Standards Report (2009). Only 23.7 % of the studies included in the review reported a sufficient volume of data for a regression-based calculation. Carr and colleagues also explored the applicability of three readily hand-calculated non-parametric procedures for calculating effect sizes. PND was selected because of the frequency with which the procedure is reported and both PAND and NAP because they were developed to address limitations evident in PND. The authors concluded that a NAP calculation, which is not restricted by either volume of data points or the presence of ceiling or floor points in baseline, appeared appropriate for all studies that were included in the review and that PND was applicable with 90 % of the studies sampled. Conversely, PAND, which can only be applied when a minimum of 20 data points are presented, appeared applicable for only 54 % of the studies.

The purpose of this current study was to conduct a sensitivity analysis on treatment effect scores for use by the variety of stakeholders working with the ASD community. Accordingly, a primary requirement of the procedures included was the ability to perform all calculations by hand. PND has been selected as the basis of comparison in the sensitivity analysis, as it has been widely adopted in published literature. Burns et al. (2012) have recently recommended that additional research on new overlap approaches, particularly PAND and NAP, is warranted. The literature review for this current study has also found support for the suggestion made by Burns and colleagues, with NAP identified on the basis of anticipated improvements and PAND on the basis of prior favourable review. The current study sought to explore the advantages and limitations of PND, PAND and NAP. In addition, it was noted in the literature that previous research on newer calculations has been limited to AB designs (Brossart et al. 2006; Parker and Vannest 2009). As such, this study has included data from all phases.

The following research questions were investigated:

1. Do estimated effect sizes calculated using PND, PAND and NAP differ significantly from each other?
2. What benefits or limitations are evident in estimating treatment effect size using PAND, or NAP, when compared to the PND method?
3. How do calculated treatment effect scores compare with each other using available interpretation scales?

Method

Data Set Creation

Studies were located by conducting a systematic search of peer-reviewed literature prior to November 2013. Both PsycINFO and ERIC databases were queried using the

keywords “autism*” and “Asperger’s syndrome” which were combined with the following terms typically associated with self-management: “self-management”, “self-regulation”, “self-regulate”, “self-monitoring”, “self-recording”, “self-reinforcement”, “self-evaluation”, “self-advocacy”, “self-observation”, “self-instruction”, “empowerment”, “self-determination” and “self-control”. In addition, a hand search of the reference lists of existing systematic reviews of self-management studies was undertaken.

The abstract of each article was examined to determine whether the article was likely to meet inclusion criteria for further review. The original article was retrieved and reviewed when further clarification appeared necessary. No age limits were placed upon participants. Inclusion criteria required that

1. Participants had an existing diagnosis of ASD or AS (for studies that included participants with differing conditions, only participants with ASD or AS were included for further review).
2. The study utilized a single-subject research design such as a multiple baseline, reversal, changing criteria or alternating treatment design.
3. Data for each phase and for each participant was presented in graphical format thus enabling calculation of PND, PAND and/or NAP.
4. Components of self-management were included throughout the intervention.
5. Articles were published in an English language peer-reviewed journal.

This search procedure identified 38 articles that were included for further review.

Calculating Treatment Effect

A treatment effect score was calculated for each participant included in each study as described for the following three methods.

PND (Scruggs et al. 1987) was calculated by counting the number of treatment data points that exceed the most extreme baseline data point, in the expected direction determined by whether an increase or decrease in target behaviour was desired. This number was then divided by the total number of treatment phase data points. Scruggs and colleagues have advised against coding a study when baseline stability has not been established and additionally noted that for cases including ceiling or floor baseline data points that yield a 0 % PND, the variation between treatment effect score and original research findings should be described.

PAND (Parker et al. 2007) was calculated by determining the minimum number of data points that need to be removed from either the baseline and/or treatment phases to eliminate all overlap. The number of remaining data points was then

divided by the total number of data points across baseline and treatment phases. This number represents the overlap, which is then subtracted from 100 to derive the nonoverlap and finally multiplied by 100 to express this value as a percentage.

NAP (Parker and Vannest 2009) was calculated by counting all nonoverlapping pairs. Often, this is achieved most quickly by counting overlapping pairs and subtracting from the total possible pairs to obtain the nonoverlap count. The total possible pairs are determined by multiplying the number of data points in the baseline phase with the number of data points in the treatment phase. Scores are assigned for each pairwise comparison and totalled. An overlap counts as one point, a tie counts as a half point and a nonoverlap receives a zero. Each overlap score is summed, and the total subtracted from the total possible pairs. The result is in turn divided by the total possible pairs and then multiplied by 100 to derive the percentage of all nonoverlapping pair treatment effect score.

Various approaches to determining treatment effect scores beyond an initial AB phase comparison were identified in previous research of treatment effect calculation methods. Skiba et al. (1985) argued in favour of an effect size based solely on the first AB phase comparison, claiming that treatment effects beyond the first treatment tested may be confounded with multiple treatment interference or that failure to revert to baseline levels in subsequent baseline phases may be attributed to lack of experimental control or powerful treatment effects. Other approaches were based on a combination of comparable phases prior to calculation of an arithmetic mean (Scruggs et al. 1987) and a comparison of first A with last B phase (Allison and Gorman 1993). The methodology adopted by Scruggs and colleagues was selected as preferable, based on consistency with the widely published PND metric.

Interpretation of Treatment Effect Scores

The scales provided by the respective original authors of each method have been adopted to interpret treatment effect scores and are summarised in Table 1. Scruggs and colleagues have suggested the following ranges for the interpretation of PND scores: 0–50 % ineffective, 50–70 % questionable, 70–90 % effective and 90 % or greater very effective (Scruggs and Mastropieri 1998). Parker and colleagues (2007) presented

PAND as an alternative to PND; however, their original paper does not describe an interpretation scale analogous to that of PND. While a *phi* correlation coefficient can be derived using a 2×2 table of proportions, an interpretation scale for the output of this computation has not been described by the developers. Consequently, an interpretation scale for PAND or *phi* has been omitted from Table 1. In their more recent research that compared treatment effect scores with expert visual judgements made on 200 published AB phase comparisons, Parker and colleagues have proposed the following tentative ranges for the interpretation of NAP: 0–65 % weak effect, 66–92 % medium effect and 93–100 % strong effect (Parker and Vannest 2009).

Inter-Observer Agreement

Reliability of computations was verified by conducting inter-observer checks, and an initial trial coding was performed using four randomly selected studies. Both the author and a Senior Professor within the Faculty of Education separately hand counted data points for each phase of each data series and recorded tallies on a coding sheet. PND, PAND and NAP were then calculated independently by hand by each coder and recorded in the coding sheet for all studies that reported sufficient data. Both the coders then met to discuss any variations in results, and a 100 % agreement for these four studies was achieved.

Subsequently, a further 14 articles (36.8 %) were selected at random, and each coder independently calculated tallies for data points for each phase of the data series and calculated PND, PAND and NAP for each AB phase comparison. Both coders then met again to compare the scores that each had calculated independently. Overall, 172 agreements were achieved from a total of 179 treatment effect calculations, and an overall inter-coder agreement was calculated at 96.7 for 47 % of the studies included in the total data set. When calculated separately for computational procedure, an inter-observer agreement (IOA) of 98.3 % was calculated for PND, 96.6 % for PAND and 93.3 % for NAP.

Subjective assessments were made on the consistency of interpretation of the treatment effect scores that were reported in Table 2. Both assessors met to discuss a method for determining consistency between interpretation scales, and consistency was operationalised using the scales provided in Table 1 as follows:

1. One rating of “ineffective” and the other rating as any other considered a disagreement
2. Ratings on the same band as each other considered an agreement
3. Ratings plus/minus one or more band considered a disagreement.

Table 1 Interpretation scale for strength of treatment effect score

PND		NAP	
0–50 %	Ineffective		
50–70 %	Questionable	0–65 %	Weak effect
70–90 %	Effective	66–92 %	Medium effect
90 %+	Very effective	93–100 %	Strong effect

PND percentage of nonoverlapping data, NAP nonoverlap of all pairs

Table 2 Participant treatment effect scores

HEADING Author (Year)	Treatment effect score			Score variances		Interpretation		NAP	Consistent
	PND	PAND	NAP	PAND cf PND	NAP cf PND	PND			
*Koegel and Koegel (1990)	Student 1	100.0 %	100.0 %	0.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Student 2	98.3 %	98.4 %	99.6 %	0.1 %	1.3 %	Very effective	Strong effect	Agree
	Student 3	40.8 %	95.1 %	84.9 %	54.3 %	44.1 %	Ineffective	Strong effect	Disagree
	Student 4	94.1 %	82.5 %	83.2 %	-11.7 %	-10.9 %	Very effective	Medium effect	Disagree
*Sainato et al. (1990)	C1	12.5 %	77.9 %	75.2 %	65.4 %	62.7 %	Ineffective	Medium effect	Disagree
	C3	0.0 %	68.8 %	61.3 %	68.8 %	61.3 %	Ineffective	Weak effect	Disagree
	C4	62.5 %	94.4 %	96.3 %	31.9 %	33.8 %	Questionable	Strong effect	Disagree
	Tony	53.9 %	83.9 %	92.8 %	30.0 %	38.9 %	Questionable	Medium effect	Disagree
*Koegel et al. (1992)	Adam	89.5 %	88.3 %	96.0 %	-1.2 %	6.5 %	Effective	Strong effect	Disagree
	Howard	50.0 %	81.4 %	77.7 %	31.4 %	27.7 %	Effective	Medium effect	Agree
	Ian	62.2 %	97.8 %	98.0 %	35.6 %	35.8 %	Very effective	Strong effect	Agree
	Bruce	100.0 %	100.0 %	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
*Stahmer and Schreibman (1992)	Justin	73.2 %	83.8 %	82.9 %	10.6 %	9.7 %	Effective	Medium effect	Agree
	Claire	100.0 %	100.0 %	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Chris	Not coded	90.8 %	92.4 %	n/a	n/a	Not coded	Medium effect	n/a
	Andre	Not coded	73.3 %	80.9 %	n/a	n/a	Not coded	Medium effect	n/a
Koegel and Frea (1993)	Jon	100.0 %	n/a	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Howard	100.0 %	n/a	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Robby	100.0 %	n/a	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Barret	33.7 %	93.8 %	97.3 %	49.4 %	44.3 %	Ineffective	Strong effect	Disagree
*Strain et al. (1994)	Aubrey	96.1 %	99.0 %	100.0 %	4.0 %	5.9 %	Very effective	Strong effect	Agree
	Sidney	100.0 %	100.0 %	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Scott	96.7 %	94.3 %	99.3 %	-2.4 %	2.7 %	Very effective	Strong effect	Agree
	Peter	90.5 %	81.3 %	97.4 %	-9.2 %	6.9 %	Very effective	Strong effect	Agree
*Newman et al. (1995)	Alex	50.0 %	77.8 %	75.0 %	27.8 %	25.0 %	Ineffective	Medium effect	Disagree
	North	80.0 %	87.5 %	94.0 %	7.5 %	14.0 %	Effective	Strong effect	Disagree
	Broderick	100.0 %	100.0 %	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Jarred	50.0 %	84.2 %	98.1 %	34.2 %	48.1 %	Ineffective	Strong effect	Disagree
Dixon et al. (1995)	Student 1 F	28.6 %	n/a	51.2 %	32.9 %	22.6 %	Ineffective	Weak effect	Disagree
	Marv	89.5 %	93.5 %	94.0 %	4.1 %	4.6 %	Effective	Strong effect	Disagree
	James	78.9 %	90.3 %	95.2 %	11.4 %	16.2 %	Effective	Strong effect	Disagree
	Andy	83.3 %	93.1 %	97.9 %	9.8 %	14.6 %	Effective	Strong effect	Disagree
*Strain and Danko (1995)	Child 1	50.0 %	n/a	71.9 %	12.5 %	21.9 %	Ineffective	Medium effect	Disagree
Shearer et al. (1996)									

Table 2 (continued)

HEADING Author (Year)	Treatment effect score PND	PAND	NAP	Score variances PAND cf PND	NAP cf PND	NAP cf PAND	Interpretation PND	NAP	Consistent
Kern et al. (1997) *Callahan and Rademacher (1999) *Reinecke et al. (1999)	Child 2	25.0 %	n/a	98.3 %	47.7 %	73.3 %	Ineffective	Strong effect	Disagree
	Child 3	25.0 %	n/a	60.0 %	44.2 %	35.0 %	Ineffective	Weak effect	Disagree
	Charlie	100.0 %	n/a	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Seth	48.8 %	85.0 %	61.2 %	36.2 %	12.3 %	Ineffective	Weak effect	Disagree
	Jack-school	58.3 %	92.3 %	98.5 %	34.0 %	40.2 %	Questionable	Strong effect	Disagree
	Paul-school	60.0 %	88.5 %	97.6 %	28.5 %	37.6 %	Questionable	Strong effect	Disagree
	Stan-home	94.7 %	96.7 %	99.5 %	1.9 %	4.8 %	Very effective	Strong effect	Agree
	Keri Study 1	94.4 %	96.8 %	98.8 %	2.4 %	4.4 %	Very effective	Strong effect	Agree
	Keri Study 2	92.1 %	99.2 %	99.4 %	7.1 %	7.3 %	Very effective	Strong effect	Agree
	Evan	92.9 %	95.0 %	97.0 %	2.1 %	4.2 %	Very effective	Strong effect	Agree
*Newman et al. (2000)	Dan	85.7 %	92.9 %	96.9 %	7.1 %	11.2 %	Effective	Strong effect	Agree
	Nancy	88.9 %	96.3 %	94.4 %	7.4 %	5.6 %	Effective	Strong effect	Agree
	Larry	100.0 %	n/a	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Karen	85.1 %	93.7 %	99.8 %	8.6 %	14.7 %	Effective	Strong effect	Disagree
	Rick	84.3 %	99.0 %	99.8 %	14.7 %	15.5 %	Effective	Strong effect	Disagree
	Jason	100.0 %	100.0 %	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Stewart	95.3 %	97.3 %	100.0 %	1.9 %	4.6 %	Very effective	Strong effect	Agree
	Alice	50.0 %	73.1 %	89.3 %	23.1 %	39.3 %	Ineffective	Medium effect	Disagree
	Bob	50.0 %	76.9 %	71.9 %	26.9 %	21.9 %	Ineffective	Medium effect	Disagree
	Keith	100.0 %	100.0 %	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
*Wehmeyer et al. (2003) Apple et al. (2005)	Roger	100.0 %	n/a	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Abby	100.0 %	n/a	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Alex	100.0 %	n/a	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Jim	100.0 %	100.0 %	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Jacob	100.0 %	100.0 %	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Booker	100.0 %	100.0 %	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Alex	100.0 %	100.0 %	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	GS	100.0 %	100.0 %	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	AH	100.0 %	100.0 %	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Tom	83.3 %	96.4 %	97.9 %	13.1 %	14.6 %	Effective	Strong effect	Disagree
*Todd and Reid (2006) Delano (2007)	Mike	82.6 %	96.3 %	97.3 %	13.7 %	14.7 %	Effective	Strong effect	Disagree
	Robert	100.0 %	100.0 %	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree
	Alan	100.0 %	n/a	100.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree

Table 2 (continued)

HEADING Author (Year)	Treatment effect score		Score variances		Interpretation		NAP	Consistent
	PND	PAND	PAND cf PND	NAP	NAP cf PND	PND		
Lofstin et al. (2008)	Peter	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
	Justin	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
	Stuart	75.0 %	n/a	97.5 %	20.0 %	Effective	Strong effect	Disagree
	Anthony	72.7 %	n/a	97.5 %	16.0 %	Effective	Strong effect	Disagree
	William	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
Palmen et al. (2008)	Dick	33.3 %	n/a	55.6 %	33.3 %	Ineffective	Weak effect	Disagree
	Tom	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
	Kees	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
	Rob	66.7 %	n/a	88.9 %	16.7 %	Questionable	Medium effect	Disagree
	Linda	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
Dorminy et al. (2009)	Piet	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
	Sarah	66.7 %	n/a	94.4 %	16.7 %	Questionable	Strong effect	Disagree
	Jan	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
	Klaas	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
	Jake	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
Soares et al. (2009)	Ben	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
	Evan	84.6 %	n/a	99.0 %	11.8 %	Effective	Strong effect	Disagree
	Patrick	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
	Barry	92.3 %	95.7 %	98.5 %	3.3 %	Very effective	Strong effect	Agree
	Charlie	88.9 %	n/a	93.1 %	3.4 %	Effective	Strong effect	Disagree
Deitchman et al. (2010)	Jason	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
	Trevor	83.3 %	n/a	99.1 %	10.0 %	Effective	Strong effect	Disagree
	Mike	88.9 %	n/a	97.5 %	5.6 %	Effective	Strong effect	Disagree
	Justin	88.9 %	n/a	98.1 %	6.3 %	Effective	Strong effect	Disagree
	George	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
Cihak et al. (2010)	Adam	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
	Jordan	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
	Richard	100.0 %	n/a	100.0 %	0.0 %	Very effective	Strong effect	Agree
	Tony	73.7 %	n/a	98.6 %	19.6 %	Effective	Strong effect	Disagree
	Graham	63.6 %	n/a	99.0 %	32.2 %	Questionable	Strong effect	Disagree
*Todd et al. (2010)	Lisa	84.6 %	90.0 %	94.5 %	5.4 %	Effective	Strong effect	Disagree
	Daniel	85.0 %	95.8 %	94.7 %	10.8 %	Effective	Strong effect	Disagree
	Mark	11.1 %	85.7 %	31.5 %	74.6 %	Ineffective	Weak effect	Disagree

Table 2 (continued)

HEADING Author (Year)	Treatment effect score			Score variances			Interpretation			Consistent
	PND	PAND	NAP	PAND of PND	NAP of PND	NAP of PAND	PND	NAP		
Parker and Kamps (2010)	95.5 %	n/a	99.8 %	2.8 %	4.4 %	n/a	Very effective	Strong effect	Agree	
	98.0 %	n/a	99.8 %	0.5 %	1.8 %	n/a	Very effective	Strong effect	Agree	
Shogren et al. (2011)	100.0 %	100.0 %	100.0 %	0.0 %	0.0 %	0.0 %	Very effective	Strong effect	Agree	
	87.5 %	94.4 %	98.8 %	6.9 %	11.3 %	4.3 %	Effective	Strong effect	Disagree	

*Studies that met minimum data requirements for all three calculation methods

PND percentage of nonoverlapping data, PAND percentage of all nonoverlapping data, NAP nonoverlap of all pairs

Both assessors separately rated each participant treatment effect score then compared their findings. A total of 58 agreements and 43 disagreements were recorded by both assessors, and a further two scores were described as not applicable for this procedure. An IOA of 100 % was obtained for this process.

Results

A data set based on 38 articles that reported treatment data gathered across a variety of behaviours and settings for 103 participants was developed. Hand counts of the number of data points reported in each baseline and treatment phase were conducted for each of the 215 data series included in 177 graphs. These tallies were entered manually into an excel spreadsheet and subsequently compared to calculation guidelines to determine the suitability of applying PND, PAND and NAP calculations.

Baseline variability that included numerous ceiling data points was identified in one study (Koegel and Frea 1993), for which a PND result could not be calculated. Another three studies included ceiling data points in baseline that resulted in a 0 % PND for the respective participants (Koegel et al. 1992; Shearer et al. 1996; Stahmer and Schreibman 1992). As a result of too few data points noted in 15 studies, a PAND treatment effect score was calculated in 23 studies. A NAP treatment effect score was calculated for all 38 studies.

Treatment effect scores were calculated for all 103 participants for whom sufficient data was provided (see Table 2). Variations between PND, PAND and NAP scores are summarised, and treatment effect scores interpreted for PND and NAP.

To examine the variation in treatment effect score methods, the data set was reduced to include only studies that met the requirements for all three treatment effect score calculations. Twenty two studies were identified, and data for 57 participants was included. These 22 studies are indicated by an asterisk in Table 2.

Mean treatment effect scores for the 57 participants were PND 78.8 %, PAND 92.7 % and NAP 93.2 %. NAP indicated that treatment effect was on average 14.4 % greater than the product of the PND calculation and 0.5 % greater than that produced by PAND. PAND indicated strength of treatment effect that was 13.9 % greater when compared to PND.

A single-factor ANOVA was used to test the null hypothesis, that the variances between the mean treatment effect scores are equal. H_0 : the mean variances of treatment effect scores are equal. A summary of the mean variances for 57 participants and the ANOVA data is presented in Table 3.

The single-factor ANOVA resulted in an F_{calc} value that was greater than the F_{crit} value, and the conclusion drawn that

Table 3 Analysis of variance: $n=22$ studies, 57 participants

Summary						
Groups	Count	Sum	Average	Variance		
PND	57	44.92883	78.8 %	0.063873		
PAND	57	52.83922	92.7 %	0.006321		
NAP	57	53.13582	93.2 %	0.015822		
ANOVA						
Source of variation	SS	df	MS	F	p value	F _{crit}
Between groups	0.760334	2	0.380167	13.25908	4.5012E-06	3.049792
Within groups	4.816929	168	0.028672			
Total	5.577263	170				

PND percentage of nonoverlapping data, PAND percentage of all nonoverlapping data, NAP nonoverlap of all pairs

the mean variances are not equal. The difference was statistically significant ($F_{2,168}=13.259$, $p<.005$). The mean PND score of 78.8 % for 22 studies indicates an effective treatment. By comparison, the mean NAP score of 93.2 % indicates a strong treatment effect for the same 22 studies.

PND and NAP treatment effect scores were interpreted using the scales provided by the original authors and subjectively assessed for consistency in interpretation for all 103 participant scores. It appeared that these scales yielded a consistent strength of treatment effect for 58 scores (56.3 %) but appeared inconsistent for 43 scores (41.7 %), and a comparison was not applicable for the remaining two scores (1.9 %).

Discussion

The purpose of this research was to explore the suitability of three nonparametric calculation methods to estimate treatment effect size with SCD studies, with a specific focus on the needs of participants with ASD. In particular, PND, PAND and NAP hand calculation techniques were selected, and published data from self-management interventions conducted with participants diagnosed on the autism spectrum was used to test the calculations.

For each calculation method, a mean of all participant scores was calculated. Results from the ANOVA test of differences between the three methods suggested that the mean treatment effect scores derived using these three treatment effect calculation methods differ significantly with PND producing the most conservative estimates of effect size.

Benefits and limitations are apparent for each scoring method. A main criticism of PND is the weighting it places on extreme, possibly outlier data points in baseline phases (Parker et al. 2007). Scruggs and colleagues have defended their procedure, stating that these potential problems have rarely been encountered in the research literature, and when they are encountered, they can be easily addressed, as was

noted in their original conventions, by acknowledging such discrepancies in the research report (Scruggs and Mastropieri 2013). Consistent with their conclusion, the current data set included relatively few instances in which outlying data points in base line skewed the resultant treatment effect score.

Both PAND and NAP address the weakness acknowledged in PND by integrating additional baseline data points in the algorithm. PAND incorporates additional baseline data although eliminates all overlapping data across baseline and treatment phases. This study identified a significant proportion of participant data that does not meet the minimal threshold of 20 data points. Of further concern, the original developers proposed a minimum range of data points of 20–25 data points. Had the upper level of their suggested threshold, 25 data points, been adopted, it is likely that an even greater number of studies would have been deemed as not suitable for PAND methodology.

The NAP calculation appeared to offer the greatest advantage in this regard as the algorithm incorporates a pairwise comparison of all data points included in the data set, thus utilising every data point recorded. NAP is not restricted by a minimum number of data points and, in that sense, is preferable to PAND as a treatment effect score can be calculated for all studies with even greater precision. However, as NAP requires a more complex calculation than PND, or PAND, it is more error prone and potentially problematic with longer data series when calculated by hand.

The treatment effect scores calculated for both PND and NAP have been interpreted using rating scales in published literature, however no such scale is available for PAND. Using PND, the mean effect size for self-management interventions was described as *effective*, the second highest category under this scale. By contrast, using NAP, the mean effect size for self-management interventions was described as *strong*, the highest category on the scale. The difference in interpretative guidelines, in addition to the observed difference in scores derived by the procedures, suggests that PND reports a more conservative strength of treatment effect than that calculated

using NAP. This discrepancy underscores the warning by Brossart and colleagues (2006) that clinicians, as well as other interested stakeholders, face difficulty when interpreting studies that use various treatment effect calculation methods. Page | 47

Conclusion

Fifteen years ago, the Wilkinson and the Taskforce on Statistical Inference (1999) emphasised the importance of understanding how a given statistical measure is calculated and how to interpret the statistic. Findings from this study suggest that both issues remain of concern in SCD research conducted with participants with ASD.

Of the three treatment effect scores that were reviewed, PAND may cautiously be considered the least applicable to stakeholders in the autism community for two reasons. First, a significant percentage of the articles included in this review did not include sufficient data points to permit a PAND-based analysis, and as observed elsewhere (Carr et al. 2014), evidence suggests that researchers are not collecting more data as would be required were PAND to be adopted in the future. Second, interpretation of the PAND score, or its associated ϕ or ϕ^2 correlation coefficient, is difficult in the absence of a conversion scale like that available for both PND and NAP. Further, as reported by Brossart and colleagues (2006), differences across research fields in the interpretation of a ϕ or ϕ^2 term may further compromise interpretation of these statistics. Further research into the development of a scale to interpret a PAND/ ϕ calculation, with a particular focus on participants with ASD, is justified.

Few studies in this sample included lengthy baselines and consequently the NAP calculation was relatively straightforward to apply to the studies included in this review. Importantly, NAP utilised all data points reported for each participant. Given that a scale for interpreting strength of treatment effect has also been proposed, these factors arguably add support to the adoption of NAP as a potential improvement over PND for incorporation in research for participants with ASD. However, given the greater treatment effect scores calculated by NAP compared to those of PND, the adoption of NAP as a new standard should be treated cautiously.

The PND metric currently dominates SCD literature, and the present data show it to yield a relatively conservative result with strength of treatment effect for self-management intervention procedures described as *effective* under PND and *strong* under NAP. Such calibration differences across methods for calculating treatment effects, were they found to be generalizable, are unlikely to contribute positively to our understanding of the relative effectiveness of our intervention procedures. Arguably, this issue is relevant to the ASD community, and research reports using newer alternate treatment effect scores should be treated with caution to avoid

presenting potentially misleading information to the ASD research stakeholders. Importantly, this study has indicated that PND is widely applicable to the data that is gathered for participants with ASD, and its continued use appears justified.

References

*Studies included in the review

- *Agran, M., Alper, S., Cavin, M., Sinclair, T., Wehmeyer, M., & Hughes, C. (2005). Using self-monitoring to increase following-direction skills of students with moderate to severe disabilities in general education. *Education and Training in Developmental Disabilities*, 40, 3–13.
- Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: the case of the single case. *Behaviour Research and Therapy*, 31, 621–631. doi:10.1016/0005-7967(93)90115-B.
- *Apple, A. L., Billingsley, F., & Schwartz, I. S. (2005). Effects of video modeling alone and with self-management on compliment-giving behaviors of children with high-functioning ASD. *Journal of Positive Behavior Interventions*, 7, 33–46. doi:10.1177/10983007050070010401.
- *Asaro-Saddler, K., & Saddler, B. (2010). Planning instruction and self-regulation training: effects on writers with autism spectrum disorders. *Exceptional Children*, 77, 107–124. doi:10.1177/1053451208330895.
- Brossart, D. F., Parker, R. I., Olson, E., & Mahadevan, L. (2006). The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behavior Modification*, 30(5), 531–563. doi:10.1177/0145445503261167.
- Burns, M. K., Zaslofsky, A. F., Kanive, R., & Parker, D. C. (2012). Meta-analysis of incremental rehearsal using phi coefficients to compare single-case and group designs. *Journal of Behavioral Education*, 21, 185–202. doi:10.1007/s10864-012-9160-2.
- *Callahan, K., & Rademacher, J. A. (1999). Using self-management strategies to increase the on-task behavior of a student with autism. *Journal of Positive Behavior Interventions*, 1, 117–122. doi:10.1177/109830079900100206.
- Campbell, J. M. (2013). Commentary on PND at 25. *Remedial and Special Education*. doi:10.1177/0741932512454725.
- Carr, M. E., Anderson, A., Moore, D. W., & Evans, W. H. (2014). How should we determine treatment effectiveness with single-case design research for participants with autism spectrum disorder? *Review Journal of Autism and Developmental Disorders*. doi:10.1007/s40489-014-0030-9.
- *Cihak, D. F., Wright, R., & Ayres, K. M. (2010). Use of self-modeling static-picture prompts via a handheld computer to facilitate self-monitoring in the general education classroom. *Education and Training in Autism and Developmental Disabilities*, 45, 136–149.
- *Deitchman, C., Reeve, S. A., Reeve, K. F., & Progar, P. R. (2010). Incorporating video feedback into self-management training to promote generalization of social initiations by children with autism. *Education and Treatment of Children*, 33(3), 475–488. doi:10.1353/etc.0.0102.
- *Delano, M. E. (2007). Improving written language performance of adolescents with Asperger syndrome. *Journal of Applied Behavior Analysis*, 40, 345–351. doi:10.1901/jaba.2007.50-06.
- *Dixon, R. S., Moore, D. W., Hartnett, N., Howard, R., & Petrie, K. (1995). Reducing inappropriate questioning behaviour in an adolescent with autism: a case study. *Behaviour Change*, 12, 163–166.

- *Dorminy, K. P., Luscre, D., & Gast, D. L. (2009). Teaching organizational skills to children with high functioning autism and Asperger's syndrome. *Education and Training in Developmental Disabilities*, 44, 538–550.
- *Ganz, J. B., & Sigafoos, J. (2005). Self-monitoring: are young adults with MR and autism able to utilize cognitive strategies independently? *Education and Training in Developmental Disabilities*, 40(1), 24–33.
- *Holifield, C., Goodman, J., Hazelkorn, M., & Heflin, L. J. (2010). Using self-monitoring to increase attending to task and academic accuracy in children with autism. *Focus on Autism and Other Developmental Disabilities*, 25, 230–238. doi:10.1177/1088357610380137.
- 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. doi:10.1007/s10864-012-9152-2.
- 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.
- *Kern, L., Marder, T. J., Boyajian, A. E., Elliot, C. M., & McElhattan, D. (1997). Augmenting the independence of self-management procedures by teaching self-initiation across settings and activities. *School Psychology Quarterly*, 12(1), 23–32. doi:10.1037/h0088944.
- *Koegel, L. K., Koegel, R. L., Hurley, C., & Frea, W. D. (1992). Improving social skills and disruptive behavior in children with autism through self-management. *Journal of Applied Behavior Analysis*, 25, 341–353. doi:10.1901/jaba.1992.25-341.
- *Koegel, R. L., & Frea, W. D. (1993). Treatment of social behavior in autism pivotal social skills. *Journal of Applied Behavior Analysis*, 26, 369–377. doi:10.1901/jaba.1993.26-369.
- *Koegel, R. L., & Koegel, L. (1990). Extended reductions in stereotypic behavior of students with autism through a self-management package. *Journal of Applied Behavior Analysis*, 23, 119–127. doi:10.1901/jaba.1990.23-119.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). WWC Single-case designs technical documentation Version 1.0.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013a). What Works Clearinghouse procedures and standards handbook (Version 2.1).
- Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013b). *What Works Clearinghouse procedures and standards handbook (Version 3.0)*.
- Leland Wilkinson and the Taskforce on Statistical Inference. (1999). The taskforce on statistical inference statistical methods in psychology journals; guidelines and explanations. *American Psychologist*, 54, 594–604. doi:10.1037/0003-066X.54.8.594.
- *Lofstin, R. L., Odom, S. L., & Lantz, J. F. (2008). Social interaction and repetitive motor behaviors. *Journal of Autism and Developmental Disorders*, 38, 1124–1135. doi:10.1007/s10803-007-0499-5.
- Maggin, D. M., O'Keeffe, B. V., & Johnson, A. H. (2011a). A quantitative synthesis of methodology in the meta-analysis of single-subject research for students with disabilities: 1985–2009. *Exceptionality*, 19(2), 109–135. doi:10.1080/09362835.2011.565725.
- Maggin, D. M., Swaminathan, H., Rogers, H. J., O'Keeffe, B. V., Sugai, G., & Horner, R. H. (2011b). A generalized least squares regression approach for computing effect sizes in single-case research: application examples. *Journal of School Psychology*, 49(3), 301–321. doi:10.1016/j.jsp.2011.03.004.
- *Mancina, C., Tankersley, M., Kamps, D., Kravits, T., & Parrett, J. (2000). Brief report: reduction of inappropriate vocalizations for a child with autism using a self-management treatment program. *Journal of Autism and Developmental Disorders*, 30(6), 599–606. doi:10.1023/a:1005695512163.
- *Mithaug, D. K., & Mithaug, D. E. (2003). Effects of teacher-directed versus student-directed instruction on self-management of young children with disabilities. *Journal of Applied Behavior Analysis*, 36(1), 133–136. doi:10.1901/jaba.2003.36-133.
- *Morrison, L., Kamps, D., Garcia, J., & Parker, D. (2001). Peer mediation and monitoring strategies to improve initiations and social skills for students with autism. *Journal of Positive Behavior Interventions*, 3, 237–250. doi:10.1177/109830070100300405.
- National Standards Report. (2009) *National Autism Center*; Randolph MA.
- *Newman, B., Buffington, D., & Hemmes, N. (1996). Self-reinforcement used to increase the appropriate conversation of autistic teenagers. *Education and Training in Mental Retardation and Developmental Disabilities*, 31, 304–309.
- *Newman, B., Buffington, D., O'Grady, M., McDonald, M., Poulson, C., & Hemmes, N. (1995). Self-management of schedule following in three teenagers with autism. *Behavioral Disorders*, 20, 190–196.
- *Newman, B., Reinecke, D. R., & Meinberg, D. L. (2000). Self-management of varied responding in three students with autism. *Behavioral Interventions*, 15, 145–151. doi:10.1002/(SICI)1099-078X(200004/06)15:2<145::AID-BIN50>3.0.CO;2-3.
- *Newman, B., & Ten Eyck, P. (2005). Self-management of initiations by students diagnosed with autism. *The Analysis of Verbal Behavior*, 21, 117–122.
- *Palmen, A., Didden, R., & Arts, M. (2008). Improving question asking in high-functioning adolescents with autism spectrum disorders: effectiveness of small-group training. *Autism: The International Journal of Research and Practice*, 12(1), 83–98. doi:10.1177/1362361307085265.
- *Parker, D., & Kamps, D. (2010). Effects of task analysis and self-monitoring for children with autism in multiple social settings. *Focus on Autism and Other Developmental Disabilities*, 26, 131–142. doi:10.1177/1088357610376945.
- Parker, R. I., Brossart, D. F., Vannest, K. J., Long, J. R., De-Alba, R. G., Baugh, F. G., & Sullivan, J. R. (2005). Effect sizes in single case research: how large is large? *School Psychology Quarterly*, 34, 116–132.
- Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND): an alternative to PND. *The Journal of Special Education*, 40, 194–204. doi:10.1177/00224669070400040101.
- Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: nonoverlap of all pairs. *Behavior Therapy*, 40, 357–367. doi:10.1016/j.beth.2008.10.006.
- Parker, R. I., Vannest, K., & Davis, J. (2011). Effect size in single-case research: a review of nine nonoverlap techniques. *Behavior Modification*, 34, 303–322. doi:10.1177/0145445511399147.
- *Pierce, K. L., & Schreibman, L. (1994). Teaching daily living skills to children with autism in unsupervised settings through pictorial self-management. *Journal of Applied Behavior Analysis*, 27, 471–481. doi:10.1901/jaba.1994.27-471.
- *Reinecke, D. R., Newman, B., & Meinberg, D. L. (1999). Self-management of sharing in three preschoolers with autism. *Education and Training in Mental Retardation and Developmental Disabilities*, 34, 312–317.
- *Sainato, D. M., Strain, P. S., Lefebvre, D., & Rapp, N. (1990). Effects of self-evaluation on the independent work skills of preschool children with disabilities. *Exceptional Children*, 56, 540–549.
- Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: issues and applications. *Behavior Modification*, 22, 221–242. doi:10.1177/01454455980223001.
- Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25: past, present, and future trends in summarizing single-subject research. *Remedial and Special Education*, 34, 9–19. doi:10.1177/0741932512440730.

- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: methodology and validation. *Remedial and Special Education, 8*(2), 24–33. doi:10.1177/074193258700800206.
- *Shabani, D. B., Wilder, D. A., & Flood, W. A. (2001). Reducing stereotypic behavior through discrimination training, differential reinforcement of other behavior, and self-monitoring. *Behavioral Interventions, 28*(6), 279–286. doi:10.1002/bin.096.
- Shadish, W. R. (2014). Analysis and meta-analysis of single-case designs: an introduction. *Journal of School Psychology, 52*(1), 1–12. doi:10.1016/j.jsp.2013.11.009.
- *Shearer, D. D., Kohler, F. W., Buchan, K. A., & McCullough, K. M. (1996). Promoting independent interactions between preschoolers with autism and their nondisabled peers: an analysis of self-monitoring. *Early Education and Development, 7*, 205–220. doi:10.1027/s15566935eed0703_1.
- *Shogren, K., Lang, R., Machalicek, W., Rispoli, M., & O'Reilly, M. (2011). Self- versus teacher management of behavior for elementary school students with Asperger syndrome: impact on classroom behavior. *Journal of Positive Behavior Interventions, 13*, 87–96. doi:10.1177/1098300710384508.
- Skiba, R. J., Casey, & Center, B. (1985). Nonaversive procedures in the treatment of classroom behavior problems. *The Journal of Special Education, 19*(4), 459–481. doi:10.1177/002246698501900409.
- *Soares, D., Vannest, K., & Harrison, J. (2009). Computer aided self-monitoring to increase academic production and reduce self-injurious behavior in a child with autism. *Behavioral Interventions, 24*, 171–183. doi:10.1002/bin.283.
- *Stahmer, A. C., & Schreibman, L. (1992). Teaching children with autism appropriate play in unsupervised environments using a self-management treatment package. *Journal of Applied Behavior Analysis, 25*, 447–459. doi:10.1901/jaba.1992.25-447.
- *Strain, P. S., & Danko, C. D. (1995). Caregivers' encouragement of positive interaction between preschoolers with autism and their siblings. *Journal of Emotional and Behavioral Disorders, 3*(1), 2–12. doi:10.1177/106342669500300101.
- *Strain, P. S., Kohler, F. W., Storey, K., & Danko, C. D. (1994). Teaching preschoolers with autism to self-monitor their social interactions: an analysis of results in home and school settings. *Journal of Emotional and Behavioral Disorders, 2*, 78–88. doi:10.1177/106342669400200202.
- *Todd, T., & Reid, G. (2006). Increasing physical activity in individuals with autism. *Focus on Autism and Other Developmental Disabilities, 21*, 167–176. doi:10.1177/10883576060210030501.
- *Todd, T., Reid, G., & Butler-Kisber, L. (2010). Cycling for students with ASD: self-regulation promotes sustained physical activity. *Adapted Physical Activity Quarterly, 27*, 226–241.
- *Wehmeyer, M. L., Yeager, D., Bolding, N., Agran, M., & Hughes, C. (2003). The effects of self-regulation strategies on goal attainment for students with developmental disabilities in general education classrooms. *Journal of Developmental and Physical Disabilities, 15*(1), 79–91. doi:10.1023/a:1021408405270.

CHAPTER 5: PUBLISHED PAPER 3

Self-management interventions on students with autism: A meta-analysis of single-subject research. *Exceptional Children* (2014) Monica E. Carr, Dennis W. Moore, Angelika Anderson. DOI:10.1177/0014402914532235

Declaration for Thesis Chapter Five

Declaration by candidate

In the case of Chapter 5, the nature and extent of my contribution to the work was the following:


Nature of contribution	Extent of contribution (%)
I have identified this research topic under the guidance of Prof Dennis Moore and Dr Angelika Anderson. The original project was conceptualized collaboratively as a systematic literature review. I have independently expanded the review concept to a meta-analysis, and independently conducted data analysis. Prof Dennis Moore and Dr Angelika Anderson have provided guidance on the application of quality assessment guidelines to identify best evidence. I have revised the manuscript under their continued guidance until submission to peer review.	90

The following co-authors contributed to the work. If co-authors are students at Monash University, the extent of their contribution in percentage terms must be stated:

Name	Nature of contribution	Extent of contribution (%) for student co-authors only
Dennis W. Moore	Guidance and collaboration for each stage of project.	5
Angelika Anderson	Guidance and collaboration for each stage of project.	5

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the candidate's and co-authors' contributions to this work*.

Candidate's
Signature


	Date 21 April, 2015
---	------------------------

Main
Supervisor's
Signature

	Date 21 April 2015
---	-----------------------

*Note: Where the responsible author is not the candidate's main supervisor, the main supervisor should consult with the responsible author to agree on the respective contributions of the authors.

Self-Management Interventions on Students With Autism: A Meta-Analysis of Single-Subject Research

Exceptional Children
2014, Vol. 81(1) 28–44
© 2014 The Author(s)
DOI: 10.1177/0014402914532235
ec.sagepub.com


Monica E. Carr¹, Dennis W. Moore¹, and Angelika Anderson¹

Abstract

Self-management interventions aimed at skill acquisition and/or improving behavior of students diagnosed with autism spectrum disorders were examined. Twenty-three single-subject research design studies met inclusion criteria. Quality assessment of these studies was conducted using the What Works Clearinghouse guidelines, and treatment effect sizes were calculated using the percentage of nonoverlapping data. Results were analyzed by age, setting, functional level, and target behaviors. Results indicate that self-management interventions are effective for increasing both social and academic skills for students of all ages and levels of ability. Results generalized to other settings and untreated behaviors and were maintained over time. Sufficient evidence supports the conclusion that self-management is an evidence-based procedure for students diagnosed with autism spectrum disorders.

Although many treatment approaches are utilized for students with autism spectrum disorder (ASD), the standard of scientific proof met by intervention procedures based on applied behavior analysis has not been achieved by any other approach (Foxx, 2008); as a result, the application of applied behavior analysis is generally thought to be the current best practice for this population. One such intervention procedure is *self-management*, defined as an individual's application of techniques that achieve a desired change in behavior (Cooper, Heron, & Heward, 2007). There is a long tradition of behavioral research on self-management in schools, with researchers in the 1970s studying self-management strategies as a way to increase or maintain desired positive behavior and decrease undesired behavior (Brodén, Hall, & Mitts, 1971; Glynn & Thomas, 1974) or to teach writing skills (Ballard & Glynn, 1975). At that time, Lovitt (1973) noted that the educational system should not only teach students performance skills but also provide them with a foundation in the principles of

self-management. Although self-management treatment packages vary, a combination of self-monitoring, self-observation, self-evaluation, self-recording, and self-reinforcement components is typically included.

Previous Reviews of Self-Management Applications in Students With ASD

A PsycINFO search identified three literature reviews published prior to November 2012 that used self-management interventions for students with ASD. Machalicek et al. (2007) identified three self-management interventions among 26 studies that concentrated on

¹Monash University

Corresponding Author:

Monica E. Carr, Faculty of Education, Monash University, Building 6, Clayton Campus, Melbourne, VIC 3800, Australia.

classroom treatment of students aged 3 to 21 years with ASD who engaged in challenging behavior. Mixed findings were reported for the use of self-management techniques.

Lee, Simpson, and Shogren (2007) conducted a meta-analysis that investigated the efficacy of self-management techniques with students with ASD, and they selected only self-management-based interventions that targeted an increase in appropriate behavior of learners with ASD. The rationale for their decision is consistent with the current objectives of positive behavior support, with the authors noting the importance of improving social and academic outcomes by increasing positive behavior. Eleven studies published between 1992 and 2001 met inclusion criteria for their review. Students included in the study ranged in age from 3 to 17 years. Lee et al. measured strength of treatment effect of the interventions using the percentage of nonoverlapping data (PND) metric. An overall mean PND of 81.9% was reported, and the researchers concluded that self-management interventions were an effective treatment for increasing target behaviors. Lee et al. attributed advantages to self-management based intervention that included "the potential to increase students' self-reliance, facilitate skill generalization, and free teachers and staff from full management responsibility" (p. 3).

Southall and Gast (2011) conducted a review of empirical research published between 1994 and 2008 that used a self-management component with participants having a pervasive developmental disorder. Their review included studies that required a student to take responsibility for behavior, and 24 studies were identified. Seven of these studies were included in the meta-analysis of Lee et al. (2007).

Southall and Gast (2011) separated students by their diagnosis into categories for autistic disorder and higher-functioning autism/Asperger's syndrome (AS). Given their descriptive assessment of the studies, rather than an effect size calculation, Southall and Gast reported that self-management interventions were effective in teaching social, vocational, and communication skills to students

with either autistic disorder or higher-functioning autism/AS, who ranged in age from 3 to 25 years. In conclusion, their review suggested that any claim to a specific self-management component as the primary cause of behavior change would be premature.

Confidence in the Findings

The American Psychological Association Presidential Task Force on Evidence-Based Practice (2006) developed a policy statement in which it defined evidence-based practice in psychology as "the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences (p. 273)." Several versions of research assessment guidelines have been developed for single-subject research designs over the last 15 years (Chambless & Hollon, 1998; Horner et al., 2005; Kratochwill et al., 2010).

Although the standards that have been developed for the What Works Clearinghouse (WWC) are the most comprehensive to date, Horner and Kratochwill (2012) have acknowledged that any standard that may ultimately be adopted by the single-subject research community will continue to evolve over time. The most recent guidelines developed for the WWC (see Kratochwill et al., 2010) suggest restricting literature to the most recent 20 years. However, as noted more than 20 years ago, single-subject research designs enable the cumulative development of a treatment knowledge base (Lovaas, Koegel, Simmons, & Stevens Long, 1973). It is conceivable that for self-management specifically, the seminal work might have been done more than 20 years ago. Accordingly, we argue that the 20-year limitation should not apply to studies for this review.

A significant limitation of all three published reviews is that, other than including only peer-reviewed studies, no evaluation was made of the quality of the studies. We have applied the WWC (Kratochwill et al., 2010) benchmark for replicable precision that has been detailed for each of the main single-case design types. The standards require a minimum of three data points per phase to "meet standard with reservations" or five data points per phase

to “meet standards.” Interobserver agreement assessments are required for each case on each outcome variable, collected on at least 20% of sessions across all intervention phases (Kratochwill et al., 2010).

When combining studies to determine a sufficient evidence base, Kratochwill et al. (2010) suggest adherence to a 5-3-20 standard:

- A minimum of five single-case design research papers meet minimal quality assessment standards and are published in peer-reviewed journals.
- The studies must be conducted by at least three different research teams at three different geographical locations.
- The five or more studies include a total of at least 20 participants.

In the current version of the WWC guidelines (Kratochwill et al., 2010), calculation of treatment effect remains an unresolved issue. The guidelines require visual analysis for each outcome variable in studies that meet assessment standards with or without reservations. Subsequent to this, the calculation of treatment effect for studies with strong or moderate evidence is recommended (Kratochwill et al., 2010). However, unlike group research designs, a generally accepted method for the calculation of treatment effect size has yet to be established for single-case designs.

In single-case research, effect size may be calculated using regression-based estimates, standardized mean difference, or nonparametric methods. Early research into various treatment effect calculation methods indicated a preference for a regression based approach, described in the literature as favorable, though not a gold standard (Allison & Gorman, 1993). The WWC (Kratochwill et al., 2010) similarly suggested a preference for regression-based estimates but indicated that a nonparametric measure may be adequate as an approximate measure of treatment effect size. The WWC panel (Kratochwill et al., 2010) has acknowledged the need for procedural refinement to account for data trends before recommendation of a preferred approach to the calculation of treatment effect.

In a review of effect size calculations, Parker, Vannest, and Brown (2009) concluded that no single effect size calculation available adequately summarizes the effects of interventions on behavior rate and variability over time. Parker and Vannest (2012) suggested that visual analysis should be used to regulate quantitative analysis, with each approach complementing the other to inform understanding of research outcomes. A review on the use of PND to synthesize single-subject research reported that this approach resulted in coherent and valid summaries and has been used in more than 40 published research summaries spanning the last 25 years. Scruggs and Mastropieri (2013) here reemphasized that reviewers should identify specific reasons in the event that a PND calculation does not adequately reflect the findings reported by the original research authors.

This update has also focused on empirical studies that aimed to improve target behaviors, echoing the sentiment of Lovitt (1973) that education should aim to teach performance skills and self-management techniques.

Given the lack of consensus on determining treatment effect for single-subject research and having considered the major arguments of both Parker and colleagues (2009) and Scruggs and Mastropieri (2013), we have adopted the PND method for treatment effect calculations.

Purpose of the Study

The purpose of this review was to update three prior reviews that together span publications from 1992 to 2008. Search terms adopted for this update were expanded to include additional terms associated with self-management. This update has also focused on empirical studies that aimed to improve target behaviors, echoing the sentiment of Lovitt (1973) that education should aim to teach performance skills and self-management techniques.

Interventions that explored improvement in problem behaviors as a collateral effect of a skill development intervention have been favored over studies that targeted a reduction in problem behaviors.

Another purpose of this study was to conduct a quality assessment of the empirical research, an identified limitation in the existing reviews. Studies that met the inclusion criteria for this review were subsequently evaluated against the three quality assessment benchmarks described earlier. To determine adequacy of the evidence base, the generally agreed-on 5-3-20 standard for synthesis of the studies was adopted (Horner & Kratochwill, 2012).

A final purpose of this review was to adopt the description of ASD per the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013), in which the subdiagnostic category of AS has been absorbed into the broader description of ASD. Consequently, the search was expanded to include participants with a diagnosis of AS to explore the effectiveness of self-management techniques for students across the autism spectrum.

To determine adequacy of the evidence base, the generally agreed-on 5-3-20 standard for synthesis of the studies was adopted....

Method

Locating Studies

Studies were located by conducting a systematic search of peer-reviewed literature published prior to November 2012. Both the PsycINFO and ERIC databases were queried using the keywords *autism* and *Asperger's syndrome*, combined with the following terms typically associated with self-management: *self-management*, *self-regulation*, *self-regulate*, *self-monitoring*, *self-recording*, *self-reinforcement*, *self-evaluation*, *self-advocacy*, *self-observation*, *self-instruction*, *empowerment*, *self-determination*, and *self-control*.

The abstract of each article was examined to determine whether an article was likely to meet inclusion criteria for further review, and a review of the article was conducted when necessary. No age limits were placed on participants. Inclusion criteria required that

1. participants had an existing diagnosis of ASD or AS (in instances in which several participants with various conditions were included in a single article, only participants with a diagnosis of ASD or AS were included for further review);
2. the study utilized a single-subject research design, such as a multiple-baseline, reversal, changing-criterion, or alternating-treatment design;
3. the study presented data from each phase in graphical format for each participant, thereby enabling the calculation of treatment effect using PND;
4. the intervention targeted an increase in behavior described as positive or appropriate;
5. components of self-management were included throughout the intervention; and
6. articles were published in English in a peer-reviewed journal.

Quality Assessment

Twenty-nine identified studies were included in a quality assessment process. Each study was assessed to determine whether it met evidence standards, met evidence standards with reservation, or did not meet evidence standards (Kratochwill et al., 2010). Studies were required to describe the systematic manipulation of the independent variable by the researcher, who actively determined when and how experimental conditions changed. Studies that did not meet this condition were deemed to not meet evidence standards. Studies that adhered to this condition were assessed further using the following assessment criteria (Kratochwill et al., 2010):

1. Each outcome variable was systematically measured over time by more than one assessor, and interobserver

agreement reported for 20% (minimum) of data in each condition. Interobserver agreement met minimal thresholds of 80% agreement or 0.6 for Cohen's kappa.

2. The study included at least three attempts to demonstrate an intervention effect at different points in time or with three different phase repetitions.
3. An intervention phase met the required data point threshold as specified by design type:

Reversal/withdrawal designs must have a minimum of four phases per case with at least five data points per phase to meet standards or at least three data points to meet standards with reservation. Multiple-baseline designs must have a minimum of six phases with at least five data points per phase to meet standards or three data points to meet standards with reservation. Changing-criterion designs must have at least three different criteria with at least five data points per phase to meet standards or three data points to meet standards with reservation. Alternating-treatment designs need five repetitions of the sequence to meet standards or four repetitions to meet standards with reservations.

Calculating Treatment Effect

The PND calculation between treatment and baseline phases involved drawing a line through the highest baseline datum point (or lowest, depending on the expected treatment effect) parallel to the sessions axis and determining the proportion of treatment data points that exceed this line (Scruggs & Mastropieri, 1998). For studies that reported duplicated phases, such as multiple-baseline or reversal designs, PND was calculated by combining the total number of nonoverlapping data points by the total number of data points in the two phases (Scruggs, Mastropieri, & Casto, 1987). PND scores < 90% were interpreted as very effective treatments, 70% to 90% as effective, 50% to 70% as questionable, and > 50% as ineffective (Scruggs & Mastropieri, 1998).

PND was calculated for all studies except those in which the baseline phase included individual ceiling or floor data points that obscured the "true" treatment effects and led to results in conflict with visual inspection. In those instances, detailed consideration of any discrepancies between the calculated treatment effect score and observed intervention effects has been provided in accordance with the intended application of PND (Scruggs & Mastropieri, 2013). An arithmetic mean was calculated through individual participant treatment effect scores, to obtain an overall PND score for each coded study.

High- or Low-Functioning Classification

Participant descriptions were evaluated as high or low functioning in accordance with the description provided by Southall and Gast (2011). An assessment was determined from descriptions provided in the original research and incorporated traditional cutoff points of the Childhood Autism Rating Scale and Wechsler Intelligence Scale for Children—Fourth Edition (Mayes et al., 2011).

High-functioning autism

- Defined as such by the original author
- IQ > 80
- Functional language
- Clinical reports for Childhood Autism Rating Scale < 30

Low-functioning or severe autism

- Defined as such by the original author
- IQ < 80
- Restricted communication/language or life skills
- Clinical reports for Childhood Autism Rating Scale ≥ 30

Reliability of Data

Intercoder agreement was calculated for each of the four methodological steps, with the first and second authors separately coding each study before comparing results. Initial trial coding was performed with three studies to ensure consistency between assessors.

A 30% random sample of abstracts from all search results was reviewed independently, to determine the reliability of the article selection process. Intercooder agreement was determined by dividing the number of agreements by the total number of agreements plus disagreements and multiplying by 100. Intercooder agreement was 97%. To ensure accuracy in the quality assessment procedure, the second author independently evaluated a randomly selected 30% sample of studies. These results were then compared to the first author's assessments. Intercooder assessment was 97%. A 30% sample of the 70 participants was randomly selected for verification of PND calculations. Twenty-one individual participant PND scores were calculated separately and independently. Intercooder agreement was 100%. In the final step, high- or low-functioning classification assessments were independently reviewed for a 50% random sample of all participants. Descriptions of age, gender, use of language, cognitive ability, descriptions of problem behavior, and clinical reports provided in the original articles were reviewed, and intercooder agreement here was 93%.

Results

The search and study inclusion procedure identified 29 studies that aimed to increase performance or develop a new or existing skill. Although Richter and Test (2011) met inclusion criteria in the abstract search, their intervention procedures did not include self-management components and were consequently judged to be outside the scope of this review.

Five studies did not meet the standards of the quality assessment guidelines. The multiple-baseline, across-participants design used by Agran et al. (2005) included only two participants with an autism diagnosis. A third participant was undergoing assessment for AS but, as such, did not meet the inclusion criteria for this review. Consequently, the study reported an insufficient demonstration of replication of treatment effects. Apple, Billingsley, and Schwartz (2005) described

two experiments, although only the second included self-management procedures. Baseline data were not reported for one of the three participants; as a result, the study presented an insufficient demonstration of intervention effects. Todd and Reid (2006); Newman, Reinecke, and Meinberg (2000); and Todd, Reid, and Butler-Kisber (2010) provided insufficient information on interrater reliability.

Table 1 provides an overview of participant characteristics, target behavior, research design, quality assessment description, key findings, generalization and maintenance data, and a mean PND for each study that met all inclusion criteria.

Descriptive Statistics

The data analysis that follows is based only on studies that met quality assessment standards. The data set contained 23 papers by 20 research teams, including 70 participants—specifically, 65 males and 5 females, aged 3 to 25 years. One study took place in Canada, one in the Netherlands, and the remainder in various locations through 10 American states.

Social skills were the dependent variable in 15 of the 23 studies, the most common focus in self-management-based interventions for students with ASD in this data set. Seven studies targeted various aspects of academic behavior. Only one of these was included in the most recent review by Southall and Gast (2011). One study targeted daily living skills. Five studies reported evidence of a collateral reduction in problem behaviors for the study participants.

The number of studies with a PND score indicative of a highly effective intervention has increased from the three reported by Lee et al. (2007) to 13 with a mean PND $\geq 90\%$. A further six studies could be categorized as effective treatments, having a PND of 70% to 90%. Two studies fell within the questionable range (PND = 50%–70%), with one study producing a score in the ineffective range (PND < 50%) and with one study not coded. The mean PND calculation was 84.3%, based on the individual PND scores of 68 participants, indicating that, overall, self-management procedures can be considered an

Table 1. Description of studies that met quality assessment standards.

Author	Age, <i>y</i>	Gender (<i>n</i>)	Targeted skills	Experimental design	Quality assessment ^a	Findings	Generalization / maintenance data	Pnd ^b
Asaro-Saddler & Saddler (2010)	6–9	M (3)	Academic story writing	MB across responses and participants	Meets with reservation	Increased number of story elements for fictional stories, quality of the story, and the average number of words. Generalized from fictional to narrative writing, maintained at 4 weeks	Observational data for maintenance	92.6
Cihak, Wright, & Ayres (2010)	11–13	M (3)	Academic task engagement	ABAB reversal	Meets with reservation	Substantially increased academic task engagement using static self-modeling photos	None	100.0
Deitchman, Reeve, & Progar (2010)	5–7	M (3)	Social initiation	MB across participants	Meets with reservation	Learned to socially initiate using daily video feedback sessions to teach discrimination. Generalized to center time, lunch, playground, and buddy classrooms in general education	Observational data for maintenance	90.7
Delano (2007)	13–17	M (3)	Academic story writing	MB across responses and participants	Meets with reservation	Used video self-modeling to teach self-regulated strategy development and self-monitoring. Increased number of words, functional essay elements, in persuasive essay writing. Two participants maintained increase in number of words at 3 months; functional essay elements were not maintained	Observational data for maintenance	100.0
Dorminy, Luscre, & Gast (2009)	9–10	M (4)	Academic file organization	MB across participants	Meets standards	Immediately improved the percentage of items filed correctly and the time in seconds that it took to retrieve an item as requested by the teacher	None	96.2
Ganz & Sigafoos (2005)	20	M (1)	Vocational task completion	Changing criterion	Meets standards	Improved rate of independent task completion. Authors reported collateral effect of increased willingness to complete work	None	100.0
Holfield, Goodman, Hazelkorn, & Heflin (2010)	9–10	M (2)	Academic accuracy / task attending	MB across participants	Meets standards	Immediately increased both academic accuracy and attending to task using self-monitoring and verbal prompting in language and mathematics	None	68.7
Koegel & Frea (1993)	13–16	M (2)	Communication	MB across behavior	Meets standards	Rapidly increased level of appropriate social behavior, which was maintained after fading. Generalized to other untreated behaviors; authors reported social communicative behavior as a pivotal response	Observational data for generalization	Not coded ^c
Koegel, Koegel, Hurley, & Frea (1992)	7–11	M (4)	Social skills	MB across settings	Meets standards	Improved severe deficits in social skills, improvement associated with a reduction in disruptive behavior	Anecdotal generalization of setting	65.5 ^d
Lofin, Odom, & Lantz (2008)	9–10	M (3)	Social initiation and interaction	MB across participants	Meets with reservation	Increased peer social initiations and following interactions during lunch and recess. Reduction in repetitive motor behavior reported, maintained after 1 month	Observational data for maintenance	71.5
Morrison, Kamps, Garcia, & Parker (2001)	10–13	M (3), F (1)	Requesting, commenting, sharing	MB across skills	Meets standards	Increased peer initiations and interactions. Peer-monitoring and self-monitoring reported as equally effective	Observational data for follow-up probes	91.2
Newman, Buffington, & Hemmes (1996)	Teens	M (3)	Social appropriate conversation	Reversal	Meets standards	Increased levels of appropriate conversation, self-reinforcement reported as effective as external reinforcement.	None	83.9
Newman et al., (1995)	14–17	M (3)	Schedule following	MB across students	Meets standards	Rapidly increased accurate identification of transition times when following schedules. All students independently and appropriately identified the time to transition. Maintained after 1-month follow-up	Observational data for follow-up	79.0
Newman & Ten Eyck (2005)	6–9	M (3)	Social initiations	MB across subjects	Meets standards	Learned and maintained social initiations in exchange for reinforcers. Maintained when switched to self-management; two of three children increased social initiations during self-management condition. Student participants were older before they learned social initiation skills	Anecdotal observation of generalization of setting	95.8

(continued)

Table 1. (continued)

Author	Age, <i>y</i>	Gender (<i>n</i>)	Targeted skills	Experimental design	Quality assessment ^a	Findings	Generalization / maintenance data	Pnd ^b
Palmen, Didden, & Arts (2008)	17–25	M (7), F (2)	Question skills during tutorials	MB across participants	Meets with reservation	Substantially increased correct question asking during tutorials for four students who learned new skill and maintained appropriate question asking for five students who had previously learned skill	Observational data for follow-up	85.2
D. Parker & Kamps (2010)	9	M (1), F (1)	Functional skills (cooking, restaurant), verbal interactions	MB across three activities and two participants	Meets with reservation	Increased task completion of games, cooking, and restaurant using self-monitoring and a task analysis. Increased activity engagement and peer-directed verbalizations using either the task analysis or language social scripts. Maintained after fading	Observational data for maintenance	96.7
Pierce & Schreibman (1994)	6–9	M (3)	Daily living skills	MB across behavior	Meets with reservation	Increased engagement in daily living and decreased inappropriate behavior. Picture prompts reported effective for self-management in complete absence of supervision	Observational data for follow-up	100.0
Reinecke, Newman, & Meinberg (1999)	4	M (3)	Sharing skills	Reversal	Meets with reservation	Increased sharing under self-management; accuracy of self-management was not high. Findings suggest that the act of sharing became a reinforcing event	Anecdotal observation response not maintained	71.0
Shearer, Kohler, Buchan, & McCullough (1996)	5	M (3)	Independent peer interactions	AT and MB	Meets with reservation	Increased interactive play; adult monitoring and self-monitoring were equally effective. Authors suggested that adult support would be required to maintain frequent and high-quality social exchanges over time	Observational data maintenance; anecdotal long- term decline	33.3 ^c
Shogren, Lang, Machalicek, Rispoli, & O'Reilly (2011)	5	M (2)	Academic behavior and engagement	ABACABAC	Meets standards	Improvement under token economy and self-management conditions for levels of appropriate classroom behavior and for engagement in classroom activities	Observational data for maintenance	93.8
Soares, Vannest, & Harrison (2009)	13	M (1)	Academic production	Reversal	Meets with reservation	Increased academic completion and decreased maladaptive behaviors. Academic engagement and task completion may be useful in decreasing problem behavior	None	92.3
Stahmer & Schreibman (1992)	7–13	M (2), F (1)	Appropriate play with toys	MB across subjects	Meets standards	Increased level of appropriate play. Generalized across settings; two of three children maintained high levels of appropriate play at 1-month follow-up. Self-stimulatory behavior decreased as appropriate play increased	Observational data maintenance and generalization	91.1
Strain, Kohler, Storey, & Danko (1994)	3–5	M (3)	Social exchanges with peers/siblings	MB across children and setting	Meets standards	Increased social interactions with peers/siblings. Adult prompts and reinforcement were successfully faded. Poster picture prompt condition had little impact on interactions, not considered to contribute to effectiveness	None	73.4

Note. AT = alternating treatment; F = female; M = male; MB = multiple baseline; PND = percentage of nonoverlapping data

^aBased on What Works Clearinghouse (Kratochwill et al., 2010).

^bMI = 84.3%.

^cCeiling data points in baseline for nonverbal mannerisms and perseveration of topic; study not coded.

^dCeiling data points in baseline for Howard community assumed normal baseline variability; PND coded as 0%.

^eCeiling data points in baseline for Child I engagement considered in line with baseline variability; PND coded as 0%.

Table 2. Participant characteristics.

	<i>n</i>	Percentage of nonoverlapping data	
		<i>M</i>	Description
Gender			
Male	65	83.8	Effective
Female	5	90.0	Highly effective
Age			
Preschool (2–5)	12	68.4	Questionable
School age (6–12)	33	88.0	Effective
Adolescence (13–17)	18	89.9	Effective
Adulthood (≥18)	7	81.4	Effective
Ability			
High	38	90.8	Highly effective
Low	32	77.1	Effective
Education			
General	35	84.8	Effective
Special	15	83.0	Effective
Setting			
Home	19	87.3	Effective
Community	8	74.8	Effective
Clinic	17	82.1	Effective
Multiple	24	82.8	Effective
Behavior			
Social skills	49	79.8	Effective
Academics	18	93.3	Highly effective
Daily living	3	100.0	Highly effective

effective treatment for students on the autism spectrum. Two participants without a PND score were omitted from the mean calculation.

Table 2 provides additional details regarding gender, age, ability, intervention setting, and target behavior for each of the 70 participants in these studies.

Examination of studies grouped by age revealed the following mean PND calculations: preschool-aged learners, 68.4%; school-aged learners, 88.0%; adolescent learners, 89.9%; and adults, 81.4%. Although PND for preschool-aged learners is marginally below the effective threshold, it is important to note that one study in this group (Shearer, Kohler, Buchan, & McCullough, 1996) had a ceiling

effect in baseline data for one child and high baseline scores for the other two. The original authors described effective outcomes, both for increasing interactive play with either an adult or child intervention agent and for maintaining social engagement in the absence of adult prompts. Self-management interventions appear effective across all age ranges, particularly with adolescents.

PND mean calculations were also conducted when participants were grouped according to target behaviors. One study was conducted with three participants that targeted daily living skills, and a mean PND of 100% was obtained. Fourteen studies that included 47 participants addressed social skills, resulting in a mean PND of 79.8%. Seven studies that involved 18 participants targeted academics and yielded a PND of 93.3%. These findings suggest that self-management procedures are highly effective for increasing a range of behaviors.

A mean PND was also calculated for the participants when grouped by the various settings included in this review. The following mean PNDs were derived: general education settings, 84.8%; special education settings, 83.0%; homes, 87.3%; community settings, 74.8%; and clinics, 82.1%. The consistency in these PND scores indicated that self-management procedures have been shown to be effective in all these settings.

Thirty-eight students (54%) in the cohort were classified as high functioning and 32 students (46%) as low functioning. PND scores were omitted for two of the high-functioning students (Koegel & Frea, 1993). Mean PND scores of 90.8% and 77.1% were obtained for high- and low-functioning participants, respectively, reflecting differential self-management treatment effects for these two groups (highly effective, effective).

Analysis of Studies

Studies were grouped by the intervention characteristics summarized in Table 3. Mean PND calculations were conducted for each category, derived from the individual participant scores. A self-recording procedure was included in 16 studies (70%), and a

Table 3. Intervention characteristics.

	AC	SS	DL	SM	SRC	SRE	SGS	ADR	NR	P	PV	VSP	VRP	CR	NSC
Koegel et al. (1992)		x		x	x	x							x	x	3
Stahmer & Schreibman (1992)		x		x	x	x								x	3
Koegel & Frea (1993)		x		x	x	x									3
Strain et al. (1994)		x		x	x	x				x		x	x		3
Newman et al. (1995)		x		x		x							x		2
Newman et al. (1996)		x		x		x							x		2
Shearer et al. (1996)		x		x	x			x					x		2
Reinecke et al. (1999)		x		x		x							x		2
Morrison et al. (2001)		x		x	x	x				x					3
Newman & Ten Eyck (2005)		x		x		x							x		2
Ganz & Sigafoos (2005)		x		x				x					x		1
Loftin et al. (2008)		x		x	x	x				x			x	x	3
Palmen et al. (2008)		x		x	x	x						x			3
Deitchman et al. (2010)		x		x				x			x		x		1
D. Parker & Kamps (2010)		x		x	x				x	x		x	x		2
Pierce & Schreibman (1994)			x	x		x						x		x	2
Delano (2007)	x			x	x		x		x		x	x	x		3
Dorminy et al. (2009)	x			x	x				x				x		2
Soares et al. (2009)	x			x	x				x			x	x	x	2
Asaro-Saddler & Saddler (2010)	x			x	x	x	x					x			4
Cihak et al. (2010)	x			x	x				x		x				2
Holifield et al. (2010)	x			x	x			x					x		2
Shogren et al. (2011)	x			x	x			x					x		2
Total studies, %	30.0	65.0	4.0	100	70.0	57.0	9.0	22.0	22.0	17.0	13.0	30.0	70.0	22.0	
Mean PND	93.3	79.8	100	84.3	83.0	83.2	96.3	72.5	97.7	82.8	96.9	89.6	80.2	81.1	

Note. AC = academics; SS = social skills; DL = daily living; SM = self-monitoring; SRC = self-recording; SRE = self-reinforcement; SGS = self-goal setting; ADR = adult-delivered reinforcement; NR = no reinforcement; P = peers; PV = photo/video prompts; VSP = visual prompts; VRP = verbal prompts; CR = collateral reduction in problem behavior; NSC = number of self-components; PND = percentage of nonoverlapping data.

mean PND of 83.0% was calculated for these interventions. Manual recording methods were noted in the most recent and highly effective studies and included recording sheets, blank bar charts, daily organizational chart with plus (+) and minus (−) signs to denote correct behavior, pencils, and marker pens (Delano, 2007; Dorminy, Luscre, & Gast, 2009; Parker & Kamps, 2011).

Self-reinforcement was the most common reinforcement method described in the data set, which occurred in 13 studies (57%). Newman and Tan Eyck (2005) described the use of

a token exchange system, in which students choose a preferred activity from a picture menu of available reinforcers, and they noted that inaccuracies tended to reflect forgetfulness rather than participants taking undeserved tokens. Newman, Buffington, and Hemmes (1996) also noted a pattern of under-reinforcement. Reinforcement was provided by adults in five studies (22%) and was not mentioned in the remaining five. Mean PND scores of 83.2% and 72.5% were achieved for the self-reinforcement and adult reinforcement studies, respectively. Studies that did not

report a reinforcement component produced the highest mean PND score, 97.7%. Four of these five studies targeted academic-related behaviors, and one targeted social skills. All participants involved in these five studies were described as high functioning.

Prompts were incorporated in the majority of studies. Only four did not describe the use of any type of prompt. Verbal prompts were most common, occurring in 16 studies (70%) and yielding a mean PND of 80.2%. Soares, Vannest, and Harrison (2009) achieved a highly effective increase in academic productivity and a simultaneous decrease in problem behavior using verbal prompts and self-recording. The participant substituted his preference for a large plush Mickey Mouse to an age-appropriate version of Mickey that was cut and pasted in a computerized recording sheet. A typical morning prompt was provided: "Don't forget we are going to cut and paste Mickey." Photos or videos were used in three studies (13%), and other visual prompts, including flowcharts and written instructions, were reported in another seven (30%). Though less common, these visually based prompts proved highly effective. Mean PND scores of 96.9% and 89.6% were calculated for photos/videos and other visual prompts, respectively.

Four studies included mention of peer involvement during intervention. The specific role of peers varied, with some taking a tutoring role (Parker & Kamps, 2010) or facilitating peer monitoring (Morrison, Kamps, Garcia, & Parker, 2001) and with others acting as confederates in social interaction processes (peer-directed verbalization, modeling, and feedback; Strain, Kohler, Storey, & Danko, 1994) or providing positive reinforcement for appropriate social initiations (Loftin, Odom, & Lantz, 2008). Overall, these studies produced a mean PND of 82.8%, reflecting effective treatment.

Goal setting was described in two studies (9%), both of which used the self-regulated strategy development procedure, which includes self-modeling, goal setting, and self-recording to improve academic story writing. Delano (2007) focused on increasing the

number of words written by three teens. Asaro-Saddler and Saddler (2010) used the same procedure to improve story writing for three school-aged children. Participants worked with their instructor to develop a goal of including more story elements in light of feedback received. Students were observed independently setting goals for the following day on occasions when they had not included all story elements in their writing task (Asaro-Saddler & Saddler, 2010). A mean PND of 96.3% was calculated for both studies. Although few in number, the positive results described by the authors of these studies suggest that goal setting is a self-management strategy worthy of further exploration.

Discussion

This study updates earlier self-management reviews (Lee et al., 2007; Machalicek et al., 2007; Southall & Gast, 2011). Search terms typically associated with self-management have been extended to include *self-recording*, *self-control*, *self-determination*, and *empowerment*, and studies conducted with individuals with AS have been included. This study

*A reduction of problem behaviors
has been described as a collateral
effect of self-management
interventions....*

also extended prior work by assessing the quality of studies that were included, and it has explored many of the themes identified previously as gaps in the knowledge base. The data set has grown significantly in the last decade, from 11 studies involving 34 participants (Lee et al., 2007) to 23 studies involving 70 participants. Of these, 12 met the WWC quality assessment guidelines, and a further 11 met standards with reservation.

WWC also recommended a 5-3-20 standard for synthesis of literature. Eight research teams conducted 12 studies that met WWC standards and included a total of 34 participants, resulting in a ratio of 12-8-34,

exceeding the minimum recommended threshold. This review thus found adequate evidence to support self-management intervention procedures. These studies reflect effective treatment for students aged from 3 to 20 years. Further, the 11 studies that met standards with reservation were conducted by 11 research teams with 36 participants. This resulted in a ratio of 11-11-36 and may be considered as additional emerging evidence of successful self-management interventions conducted with participants aged 5 to 25 years.

A reduction of problem behaviors has been described as a collateral effect of self-management interventions for five studies reflecting 14 participants. For three studies, the dependent variables included an increase in various social skills; improvements in problem behaviors were also measured (Koegel, Koegel, Hurley, & Frea, 1992; Loftin et al., 2008; Stahmer & Schreibman, 1992). A reduction in stereotypic behavior was described in an intervention that taught various daily living skills to three low-functioning children (Pierce & Schreibman, 1994). Soares et al. (2009) used a self-monitoring intervention to increase academic output for one teenaged participant with AS. Using data collected during intervention, these researchers graphically demonstrated an inverse relationship between task completion and tantrums/self-injury. Findings illustrated that as task completion increased, problem behaviors decreased.

Findings of this review have indicated that self-management was highly effective for academic behaviors.

Goal setting has emerged as a new theme, having been discussed in two recent academic studies involving six participants described as high functioning. Both these studies used self-regulated strategy development as a technique to improve students' writing, with goal setting described as an important aspect of self-regulation. The authors of these studies

noted the independent use of goal setting by the students following training and feedback.

The application of self-management procedures to academic behaviors has been identified as a gap in the published literature (Lee et al., 2007); however, since then, seven studies have been published targeting academic behaviors, such as a student's ability to stay on task or improve the quality or volume of academic work. Findings of this review have indicated that self-management was highly effective for academic behaviors.

Although early research on the application of self-management interventions with children with ASD was conducted in clinics, homes, and special education classrooms, research has recently extended into general education settings, and highly effective outcomes have been reported. The earlier meta-analysis of Lee et al. (2007) suggested that additional self-management research involving younger children was warranted. Recently, research evidence has continued to grow, with preschool-aged children now accounting for 17% of the data set. Research has also been extended to adult learners on the autism spectrum.

Although interventions for young children have typically targeted various social skills, Shogren, Lang, Machalicek, Rispoli, and O'Reilly (2011) recently reported a highly effective treatment directed at improving classroom engagement of 5-year-olds. Further research is clearly warranted exploring the effects of self-management training with young children with ASD in the classroom.

Adult learners in this review included six individuals classified as high functioning and one as low functioning. Palmen, Didden, and Arts (2008) used self-recording and self-reinforcement to substantially increase question-asking skills during tutorial sessions. Skills were found to generalize from training conducted in a therapy room to the natural tutorial settings. For three of these participants, results were highly effective, with PND scores of 100%, and were within the range of questionable to ineffective for the other three students aged 18 years or older. However, high baseline rates of correct question asking

account for lower individual treatment effects for these participants (Palmen et al., 2008). Ganz and Sigafos (2005) also found self-monitoring, when combined with time on preferred activities as reinforcement, a highly effective treatment for improving independent task completion for a 20-year-old in a vocational training setting.

Consistent with previous research findings regarding visual learning styles of students with ASDs (Quill, 1997), the use of visual prompts within self-management procedures has also been associated with positive outcomes. Although students have initially responded favorably to the use of static picture prompts using a handheld device, Cihak et al. (2010) suggested that future research explore the novelty effect and use of technology over a longer time frame. Although verbal prompts have been used more frequently, the PND metric indicates that visual prompts were associated with highly effective treatments. Continued research is recommended on the effects of visual supports on self-management procedures for children with ASD.

The data set provided limited information on a number of dimensions. Six of the seven studies that targeted academic behaviors involved 16 participants classified as high functioning. Only one study (Holifield, Goodman, Hazelkorn, & Heflin, 2010) involved two students classified as low functioning. Although attending to task and academic accuracy improved relative to baseline measures for both participants, effect sizes were not as great as those achieved with higher-functioning participants in the other academic studies. There is a need for further research targeting academic behaviors across a broader range of learners.

Peer support for self-management programs may be emerging as a research focus. Four studies, involving a total of 12 participants, included peers in various facilitating or reinforcing roles in interventions that targeted gains in sharing or verbal initiations or interactions. The results suggest that these were effective treatments. Consequently, the role of peers in self-management interventions appears worthy of further research.

Although the WWC panel (Kratochwill et al., 2010) has cautioned against PND, evidence in support of the use of PND has recently emerged, with proponents claiming that the metric remains versatile, meaningful, and reliable when used appropriately (Scruggs & Mastropieri, 2013). This meta-analysis has adopted this approach to measuring treatment effects. Employing conventional approaches to the use of PND, only one study in this set (Koegel & Frea, 1993) was judged unsuitable to score. Koegel and Frea (1993) concluded that treatment effects were rapid and the intervention highly effective, suggestive of pivotal behavior change. Overall, PND scores varied from the original outcomes reported in the original studies on only two occasions (Koegel & Frea, 1993; Shearer et al., 1996).

Both Machalicek et al. (2007) and Southall and Gast (2011) assessed studies using descriptive methods. One study (Mancina, Tankersley, Kamps, Kravits, & Parrett, 2000) was included in both reviews, and intervention effects were described as mixed and successful, respectively. By contrast, both Lee et al. (2007) and this current review adopted PND, and mean scores were noted to remain consistent between reviews. This outcome may suggest that a treatment effect score has merit in ensuring uniformity among reviewers.

Southall and Gast (2011) suggested that future research evaluate the relative contribution of components of a self-management program toward overall effectiveness. Based on PND in this current update, highly effective treatment outcomes were reported across target behaviors, and a prevailing pattern regarding component inclusion was not apparent among studies. Accordingly, the findings of this review are consistent with those of Southall and Gast and support their suggestion that further research into the effectiveness of specific components is warranted.

Although earlier reviews did not assess the quality of studies, this update has addressed this limitation. This current review has identified a pattern of collateral reduction in problem behavior associated with self-management interventions that target an increase in

positive behavior. PND scores reflecting the expanded data set have continued to describe self-management interventions within the effective range, and, for this review, PND appears to report an accurate reflection of the original findings for the majority of studies.

Conclusion

The ability to function as independently as possible within the general community is an important lifelong goal (Lovitt, 1973). This review was restricted to studies that targeted the development of a variety of performance skills while teaching individual participants the principles of self-management. Findings of this review indicate that there is sufficient research evidence to conclude that self-management should be considered an evidence-based treatment: Clinicians, teachers, and parents now have evidence of successful outcomes from the application of self-management procedures in a variety of naturalistic settings and across age groups. Specifically, this review has identified successful applications of self-management techniques in the areas of social skills development and, recently, in academic development among students of all ages with ASD. In addition, collateral benefits in the form of a reduction of challenging behaviors have been reported. Continued exploration of self-management procedures across a variety of educational, home, and community settings and activities may prove valuable for developing improved ways of assisting individuals, young and old, with an ASD to function with greater independence in an increasingly competitive world.

References

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

Agran, M., Alper, S., Cavin, M., Sinclair, T., Wehmeyer, M., & Hughes, C. (2005). Using self-monitoring to increase following-direction skills of students with moderate to severe disabilities in general education. *Education and Training in Developmental Disabilities, 40*, 3–13.

- Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behaviour Research and Therapy, 31*, 621–631. doi:10.1016/0005-7967(93)90115-B
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- American Psychological Association Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *The American Psychologist, 61*, 271–285. doi:10.1037/0003-066X.61.4.271
- Apple, A. L., Billingsley, F., & Schwartz, I. S. (2005). Effects of video modeling alone and with self-management on compliment-giving behaviors of children with high-functioning ASD. *Journal of Positive Behavior Interventions, 7*, 33–46. doi:10.1177/10983007050070010401
- *Asaro-Saddler, K., & Saddler, B. (2010). Planning instruction and self-regulation training: Effects on writers with autism spectrum disorders. *Exceptional Children, 77*, 107–124. doi:10.1177/1053451208330895
- Ballard, K. D., & Glynn, T. (1975). Behavioral self-management in story writing with elementary school children. *Journal of Applied Behavior Analysis, 8*, 387–398. doi:10.1901/jaba.1975.8.387
- Broden, M., Hall, R. V., & Mitts, B. (1971). The effect of self-recording on the classroom behavior of two eight-grade students. *Journal of Applied Behavior Analysis, 4*, 191–199. doi:10.1901/jaba.1971.4-191
- Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*, 7–18. doi:10.1037/0022-006X.66.1.7
- *Cihak, D. F., Wright, R., & Ayres, K. M. (2010). Use of self-modeling static-picture prompts via a handheld computer to facilitate self-monitoring in the general education classroom. *Education and Training in Autism and Developmental Disabilities, 45*, 136–149.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.
- *Deitchman, C., Reeve, S. A., Reeve, K. F., & Progar, P. R. (2010). Incorporating video feedback into self-management training to promote generalization of social initiations by children with autism. *Education and Treatment of Children, 33*, 475–488.
- *Delano, M. E. (2007). Improving written language performance of adolescents with Asperger

- syndrome. *Journal of Applied Behavior Analysis*, 40, 345–351. doi:10.1901/jaba.2007.50-06
- *Dorminy, K. P., Luscre, D., & Gast, D. L. (2009). Teaching organizational skills to children with high functioning autism and Asperger's syndrome. *Education and Training in Developmental Disabilities*, 44, 538–550.
- Foxx, R. M. (2008). Applied behavior analysis treatment of autism: The state of the art. *Child and Adolescent Psychiatric Clinics of North America*, 17, 821–834. doi:10.1016/j.chc.2008.06.007
- *Ganz, J. B., & Sigafoos, J. (2005). Self-monitoring: Are young adults with MR and autism able to utilize cognitive strategies independently? *Education and Training in Developmental Disabilities*, 40(1), 24–33.
- Glynn, E. L., & Thomas, J. D. (1974). Effect of cueing on self-control of classroom behavior. *Journal of Applied Behavior Analysis*, 7, 299–306. doi:10.1901/jaba.1974.7-299
- *Holifield, C., Goodman, J., Hazelkorn, M., & Heflin, L. J. (2010). Using self-monitoring to increase attending to task and academic accuracy in children with autism. *Focus on Autism and Other Developmental Disabilities*, 25, 230–238. doi:10.1177/1088357610380137
- 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. doi:10.1007/s10864-012-9152-2
- 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, 165–179
- *Koegel, L. K., Koegel, R. L., Hurlley, C., & Frea, W. D. (1992). Improving social skills and disruptive behavior in children with autism through self-management. *Journal of Applied Behavior Analysis*, 25, 341–353. doi:10.1901/jaba.1992.25-341
- *Koegel, R. L., & Frea, W. D. (1993). Treatment of social behavior in autism pivotal social skills. *Journal of Applied Behavior Analysis*, 26, 369–377. doi:10.1901/jaba.1993.26-369
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.
- Lee, S. H., Simpson, R. L., & Shogren, K. A. (2007). Effects and implications of self-management for students with autism: A meta-analysis. *Focus on Autism and Other Developmental Disabilities*, 22, 2–13. doi:10.1177/10883576070220010101
- *Loflin, R. L., Odom, S. L., & Lantz, J. F. (2008). Social interaction and repetitive motor behaviors. *Journal of Autism and Developmental Disorders*, 38, 1124–1135. doi:10.1007/s10803-007-0499-5
- Lovaas, O., Koegel, R., Simmons, J., & Stevens Long, J. (1973). Some generalization and follow-up measures on autistic children in behavior therapy. *Journal of Applied Behavior Analysis*, 6, 131–166. doi:10.1901/jaba.1973.6-131
- Lovitt, T. C. (1973). Self-management projects with children with behavioral disabilities. *Journal of Learning Disabilities*, 6, 138–150. doi:10.1177/002221947300600303
- Machalicek, W., O'Reilly, M. F., Beretvas, N., Sigafoos, J., & Lancioni, G. E. (2007). A review of interventions to reduce challenging behavior in school settings for students with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 1, 229–246. doi:10.1016/j.rasd.2006.10.005
- Mancina, C., Tankersley, M., Kamps, D., Kravits, T., & Parrett, J. (2000). Brief report: Reduction of inappropriate vocalizations for a child with autism using a self-management treatment program. *Journal of Autism & Developmental Disorders*, 30, 599. doi:10.1023/A:1005695512163
- Mayes, S. D., Calhoun, S. L., Murray, M. J., Morrow, J. D., Yurich, K. K. L., Cothren, S., Purichia, H., . . . Petersen, C. (2011). Use of the childhood autism rating scale (CARS) for children with high functioning autism or Asperger syndrome. *Focus on Autism and Other Developmental Disabilities*, 27, 31–38. doi:10.1177/1088357611406902
- *Morrison, L., Kamps, D., Garcia, J., & Parker, D. (2001). Peer mediation and monitoring strategies to improve initiations and social skills for students with autism. *Journal of Positive Behavior Interventions*, 3, 237–250. doi:10.1177/109830070100300405
- *Newman, B., Buffington, D., & Hemmes, N. (1996). Self-reinforcement used to increase the appropriate conversation of autistic teenagers. *Education and Training in Mental Retardation and Developmental Disabilities*, 31, 304–309.
- *Newman, B., Buffington, D., O'Grady, M., McDonald, M., Poulson, C., & Hemmes, N. (1995). Self-management of schedule

- following in three teenagers with autism. *Behavioral Disorders*, 20, 190–196.
- Newman, B., Reinecke, D. R., & Meinberg, D. L. (2000). Self-management of varied responding in three students with autism. *Behavioral Interventions*, 15, 145–151. doi:10.1002/(SICI)1099-078X(200004/06)15:2<145::AID-BIN50>3.0.CO;2-3
- *Newman, B., & Ten Eyck, P. (2005). Self-management of initiations by students diagnosed with autism. *The Analysis of Verbal Behavior*, 21, 117–122.
- *Palmen, A., Didden, R., & Arts, M. (2008). Improving question asking in high-functioning adolescents with autism spectrum disorders: Effectiveness of small-group training. *Autism: The International Journal of Research and Practice*, 12(1), 83–98. doi:10.1177/1362361307085265
- *Parker, D., & Kamps, D. (2010). Effects of task analysis and self-monitoring for children with autism in multiple social settings. *Focus on Autism and Other Developmental Disabilities*, 26, 131–142. doi:10.1177/1088357610376945
- Parker, R. I., & Vannest, K. J. (2012). Bottom-up analysis of single-case research designs. *Journal of Behavioral Education*, 21, 254–265. doi:10.1007/s10864-012-9153-1
- Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children*, 75, 135–150.
- *Pierce, K. L., & Schreibman, L. (1994). Teaching daily living skills to children with autism in unsupervised settings through pictorial self-management. *Journal of Applied Behavior Analysis*, 27, 471–481. doi:10.1901/jaba.1994.27-471
- Quill, K. A. (1997). Instructional considerations for young children with autism: The rationale for visually cued instruction. *Journal of Autism and Developmental Disorders*, 27, 697–714. doi:10.1023/A:1025806900162
- *Reinecke, D. R., Newman, B., & Meinberg, D. L. (1999). Self-management of sharing in three preschoolers with autism. *Education and Training in Mental Retardation and Developmental Disabilities*, 34, 312–317.
- Richter, S., & Test, D. (2011). Effects of multimedia social stories on knowledge of adult outcomes and opportunities among transition-aged youth with significant cognitive disabilities. *Education and Training in Autism and Developmental Disabilities*, 46, 410–424.
- Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: Issues and applications. *Behavior Modification*, 22, 221–242. doi:10.1177/01454455980223001
- Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25: Past, present, and future trends in summarizing single-subject research. *Remedial and Special Education*, 34, 9–19. doi:10.1177/0741932512440730
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education*, 8, 24–33. doi:10.1177/074193258700800206
- *Shearer, D. D., Kohler, F. W., Buchan, K. A., & McCullough, K. M. (1996). Promoting independent interactions between preschoolers with autism and their nondisabled peers: An analysis of self-monitoring. *Early Education & Development*, 7, 205–220. doi:10.1027/s15566935eed0703_1
- *Shogren, K., Lang, R., Machalicek, W., Rispoli, M., & O'Reilly, M. (2011). Self- versus teacher management of behavior for elementary school students with Asperger syndrome: Impact on classroom behavior. *Journal of Positive Behavior Interventions*, 13, 87–96. doi:10.1177/1098300710384508
- *Soares, D., Vannest, K., & Harrison, J. (2009). Computer aided self-monitoring to increase academic production and reduce self-injurious behavior in a child with autism. *Behavioral Interventions*, 24, 171–183. doi:10.1002/bin.283
- Southall, C. M., & Gast, D. L. (2011). Self-management procedures: A comparison across the autism spectrum. *Education and Training in Autism and Developmental Disabilities*, 46, 155–171.
- *Stahmer, A. C., & Schreibman, L. (1992). Teaching children with autism appropriate play in unsupervised environments using a self-management treatment package. *Journal of Applied Behavior Analysis*, 25, 447–459. doi:10.1901/jaba.1992.25-447
- *Strain, P. S., Kohler, F. W., Storey, K., & Danko, C. D. (1994). Teaching preschoolers with autism to self-monitor their social interactions: An analysis of results in home and school settings. *Journal of Emotional and Behavioral Disorders*, 2, 78–88. doi:10.1177/106342669400200202
- Todd, T., & Reid, G. (2006). Increasing physical activity in individuals with autism. *Focus on Autism and Other Developmental Disabilities*, 21, 167–176. doi:10.1177/10883576060210030501

Todd, T., Reid, G., & Butler-Kisber, L. (2010). Cycling for students with ASD: Self-regulation promotes sustained physical activity. *Adapted Physical Activity Quarterly*, 27, 226–241.

About the Authors

Monica E. Carr, Krongold Centre, Faculty of Education; Professor **Dennis W. Moore**, Krongold

Centre, Faculty of Education; Dr. **Angelika Anderson**, Krongold Centre, Faculty of Education, Monash University, Melbourne, Victoria, Australia.

This study was completed in partial fulfillment of the requirements for a PhD for the first author.

Manuscript received June 2013; accepted October 2013.

CHAPTER 6: PUBLISHED PAPER 4

Goal setting interventions: Implications for participants on the autism spectrum. *Review*

Journal of Autism and Developmental Disorders (2014) Monica E. Carr, Angelika Anderson,

Dennis W. Moore. DOI:10.1007/s40489-014-0022-9

Declaration for Thesis Chapter Six

Declaration by candidate

In the case of Chapter 6, the nature and extent of my contribution to the work was the following:


Nature of contribution	Extent of contribution (%)
I have conceptualized this research topic independently. Prof Dennis Moore has provided guidance and collaboration regarding inclusion criteria during data set development. Dr Angelika Anderson has provided guidance and collaboration of thematic analysis of emergent themes. I have conducted the data analysis and drafted the original manuscript independently, and Dr Anderson and Prof Moore have provided guidance to refine the manuscript through to submission to peer review.	90

The following co-authors contributed to the work. If co-authors are students at Monash University, the extent of their contribution in percentage terms must be stated:

Name	Nature of contribution	Extent of contribution (%) for student co-authors only
Angelika Anderson	Guidance and collaboration of thematic literature review, and refinement of manuscript.	5
Dennis W. Moore	Guidance and collaboration of data set development, and refinement of manuscript.	5

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the candidate's and co-authors' contributions to this work*.

Candidate's
Signature

	Date 21 April, 2015
---	------------------------

Main
Supervisor's
Signature

	Date 21 April 2015
---	-----------------------

*Note: Where the responsible author is not the candidate's main supervisor, the main supervisor should consult with the responsible author to agree on the respective contributions of the authors.

Goal Setting Interventions: Implications for Participants on the Autism Spectrum

Monica E. Carr, Dennis W. Moore & Angelika Anderson

**Review Journal of Autism and
Developmental Disorders**

ISSN 2195-7177

Rev J Autism Dev Disord
DOI 10.1007/s40489-014-0022-9



Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media New York. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Goal Setting Interventions: Implications for Participants on the Autism Spectrum

Monica E. Carr · Dennis W. Moore · Angelika Anderson

Received: 13 January 2014 / Accepted: 23 May 2014
© Springer Science+Business Media New York 2014

Abstract Single-case research design studies that included a goal setting component in interventions for a broad array of participants engaged in a learner role were reviewed. A systematic search of the empirical literature identified 38 studies that met the inclusion criteria. These studies were evaluated in terms of participant characteristics, who set the goals (self or other), how goals were incorporated into interventions, type of behavior change goal setting was applied to and whether maintenance and generalization of intervention effects was assessed. Results highlight the potential benefits of including a self-set goal component in interventions aimed at assisting participants across the autism spectrum to achieve independent functioning. An agenda is proposed for future research exploring goal setting for this population.

Keywords Autism spectrum disorder · Goal setting · Self-management · Systematic review

This study was completed in partial fulfilment of the requirements for a PhD for the first author.

M. E. Carr (✉)
Faculty of Education, Monash University, Building 6, Clayton
Campus, Clayton, Victoria 3800, Australia

D. W. Moore · A. Anderson
Krongold Centre, Faculty of Education, Monash University,
Building 6, Clayton Campus, Clayton, Victoria 3800, Australia

D. W. Moore

A. Anderson

Introduction

Goals have been defined as the object or aim of an action that an individual is trying to accomplish, and it is generally understood that goal attainment involves meeting a specific standard of task proficiency within an acceptable time frame (Locke et al. 1981). Several published reviews have reported that goal setting and goal attainment are important components of self-determination for students with various developmental or learning disabilities (Algozzine et al. 2001; Fowler et al. 2007; Konrad et al. 2007; Palmer and Wehmeyer 2003). Self-determination has been identified by the US Department of Education as an important educational outcome for students with disabilities (Algozzine et al. 2001). The Division of Career Development and Transition (DCDT) has suggested that by the age of 14 years, students should be encouraged to the full extent of their capabilities to assume a maximum amount of responsibility in planning their futures (Halpern 1994).

The skills required to develop self-determination are derived from the classic career development literature (Super 1983) and include developing an understanding of the relationship of time to goal attainment (Field et al. 1998). Research by Palmer and Wehmeyer (2003) conducted amongst children with learning disabilities, speech impairments, giftedness and mental retardation has shown that problem solving and goal setting skills develop over time. Their findings demonstrated that children with disabilities can, with the support of teachers, set goals and work through a self-determined skill development model at as young as 5 years of age.

Copeland and Hughes (2002) conducted a review on the effects of goal setting upon task performance for individuals with a diagnosis of mental retardation (MR). The authors reported an improvement in awareness and task performance upon the introduction of goal setting, with stronger effects

noted with more training. Self-management strategies were taught in addition to goal setting in 7 of the 17 studies reviewed and achieved positive outcomes. Similarly, Konrad et al. (2007) published a review of the effects of self-determination interventions on the academic skills of students with learning disabilities (LDs) and/or attention deficit/hyperactivity disorder (ADHD). Again, stronger increases in students' productivity were achieved when goal setting was added to self-management.

The literature has indicated that goal setting skills can direct attention and effort towards relevant activities and positively affect task performance (Copeland and Hughes 2002; Konrad et al. 2007; Locke and Latham 2002). While it is evident that these skills can be learned over time, it is possible that individuals with autism spectrum disorder (ASD) are not always given adequate opportunity to acquire the necessary skills to set and attain goals or opportunity to practice them. The ability to independently set challenging, attainable and appropriate goals is an important skill for students with ASD to develop and may contribute to improved awareness, task performance, fulfilment and independent functioning.

Recently, attention has been drawn to the significant number of individuals with high-functioning autism or Asperger's syndrome who were diagnosed during the early 1990s and, as a result of early intervention and supports throughout school, are both capable and motivated to enter higher education (Pinder-Amaker 2014; VanBergeijk et al. 2008). Elsewhere, it has been reported that children with ASD may face an elevated risk for developing obesity (Curtin et al. 2014). Curtin and colleagues (2014) postulated that contributing factors may include disordered sleep, atypical eating patterns and challenges associated with engaging in physical activity. It is plausible that goal setting skills may prove an important component in treatment programs aimed to overcome these barriers and develop a healthy and independent lifestyle.

However, a recent review of self-management interventions for students with ASD (Carr et al. 2014) identified a paucity of goal setting research for students on the autism spectrum. Yet, evidence of the success of goal setting techniques has been reported for individuals with mental retardation or cognitive disabilities. Therefore, to address this knowledge gap, we conducted a broader systematic literature search of goal setting research which included an array of participants not restricted by diagnostic category. The aim of this review was to map the current knowledge base regarding goal setting as an intervention, or part thereof, for students with varied additional learning needs and challenges. Given our focus on the use of goal setting as part of individualised self-management training for individuals with ASD, the search was restricted to research employing single-case research designs (SCDs). The ideographic nature of SCD research makes this methodology particularly appropriate when examining the effects of individualised interventions (Southall and

Gast 2011), such as often seen in special education research (Horner et al. 2005). These findings are then discussed with reference to their implications for students with ASD with a view to proposing an agenda for future research in this area. The following research questions were investigated:

1. What are the profiles of participants in goal setting research?
2. How have goal setting interventions been structured?
3. What target behavior was goal setting applied to?
4. Who is involved in setting participant goals (self or other)?
5. Are goal setting skills maintained over time and was generalization assessed?
6. What does the current literature suggest for applying goal setting to participants with ASD?

Method

Studies were located for this review by conducting a systematic search of peer-reviewed literature published prior to November, 2013. Keywords typically associated with goal setting were identified from existing literature reviews, and a PsycINFO database search was subsequently conducted. The following search terms were queried: "goal setting", "goal attainment" and "goal orientation". A hand search of the reference lists of existing goal setting reviews, published in peer-reviewed journals, was also conducted to identify any additional relevant articles that may have been omitted from the database search results.

The following inclusion criteria were adopted for this review:

1. The study utilised a SCD such as a multiple baseline, reversal or alternating treatment design.
2. The study presented data from each phase in graphical format for each participant.
3. The intervention included a goal setting component.
4. Participants were students engaged in a learning process.
5. Articles were published in English in peer-reviewed journals.

An independent review of both the title and the abstracts was conducted to ensure the reliability of the article selection process. A graduate student familiar with the broader self-management literature was provided a randomly selected sample of 30 % of all search results and examined both the titles and the abstracts to determine whether these articles met the inclusion criteria for further review. Inter-assessor agreement (IOA) for the article selection procedure was determined by dividing the number of agreements by the total number of

agreements plus disagreements and multiplying by 100. An IOA of 97 % was achieved for this process.

Results and Discussion

This search and selection process yielded 38 studies describing research conducted with 186 participants, who ranged in age from 6 to 54 years. Participants were attending school, higher education, elite athlete or sheltered workshop job skills training programs. The 38 studies were conducted by 29 separate research teams in various locations. A total of 30 studies were from within the USA, conducted in the following states: Alabama (2), California (1), Florida (3), Georgia (1), Hawaii (1), Illinois (2), Kansas (2), Louisiana (2), Missouri (1), Montana (1), Nebraska (2), New York (3), Ohio (3), Oklahoma (2), Oregon (1), Tennessee (2) and Washington (1). The remaining eight studies were conducted in the following locations: Canada (3), New Zealand (2) and the UK (3). Further, for one study by researchers from Oregon, data collection was conducted in South Korea.

Table 1 provides a descriptive overview of each study included in the review.

The descriptive statistics obtained from these data are further summarised in Table 2.

What Are the Profiles of Participants in Goal Setting Research?

Of the 186 participants, 128 were males (69 %) and 58 were females (31 %). For studies that reported an age range only, we have classified the participants using the uppermost age. Where studies reported an average age, we have classified the participants using the average age provided. The remaining studies provided an age for each participant. Accordingly, there were 82 elementary students aged 6 through 12 years (44 %), 73 high school students aged 13 through 17 years (39 %) and 31 adults aged 18 years or older (17 %).

Participant descriptions were classified using the primary diagnosis provided in the original study. Accordingly, there were 5 participants with autism (3 %), 4 with Asperger's syndrome (AS) (2 %), 14 with ADHD (8 %), 14 with emotional behavioral disorder or difficulties (8 %), 45 with learning disability or difficulties (24 %), 21 with mental retardation (11 %), 12 with non-compliance/conduct disorder (6 %), 1 participant who was manic/depressive (1 %), 2 gifted students (1 %), 15 participants described as typically developing (8 %), 49 athletes (26 %) and 4 participants with physical disabilities (2 %).

Mixed results were reported with young students. Figarola et al. (2008) used goal setting and self-graphing to improve the math fact fluency of three young participants that included a 7 year old with a LD and 7 and 8 year olds with ADHD. A

goal was established for the end of the year, and participants graphed their daily performance scores. Participants were given verbal praise for correctly entering their data, as well as for meeting or exceeding their aim line. Results showed that performance met or exceeded goals 83 % of the time for the 8-year-old participant and 90 % of the time for the 7 year old with LD. Modifications in the form of shorter problem sets, verbal prompts and reinforcement were required for the other 7 year-old participant before achieving consistent improvement.

By contrast, Grossi and Heward (1998) reported more consistent results in a study with older participants. Four adults aged 20–37 years old diagnosed with mild mental retardation were included in an intervention that was conducted in a community based restaurant training program. During training, each participant established a goal to increase production, goal setting being guided by the experimenter if the goal appeared unattainable. Participants were taught to self-monitor and record their performance. The authors reported that all participants increased their work productivity without compromising the quality of their work and without achieving complete accuracy in self-recording.

Similarly, Balcazar et al. (1991) reported consistent results and generalization effects among older physically disabled students. Three university students aged in their 20s and one aged 40 years were included in an intervention that used role-play to teach the skills necessary for individuals to recruit the help they may need to attain their own self-set personal goals. A goal attainment scale was used to identify either no change, realistic attainment or best anticipated attainment outcomes, and the authors reported that for the 20 goals that were set by the four participants, realistic attainment was reported eight times and best anticipated attainment was achieved for seven goals. No change was reported in only five instances. A generalization probe also reported that intervention effects were observed in natural situations in which participants asked for help.

To date, intervention research has been conducted across a broad age range of participants. While less consistency has been demonstrated with younger children, the literature we reviewed has indicated preliminary support for goal setting employed across all ages. Importantly, these findings indicate that adults have been able to learn and apply goal setting skills. While our search revealed little research conducted among participants diagnosed with ASD, goal setting has been employed in successful interventions with participants who experience learning difficulties or cognitive challenges. These preliminary findings may present significant implications for future research with this population as the behavioral issues often exhibited by individuals with ASD that may serve to mask a self-determination skill deficit amenable to goal setting intervention.

Table 1 Description of studies

Author (year)	Diagnosis	Age	M/F	Setting	Target behavior	Research design	Intervention elements	Findings	generalization/ maintenance data
Goals set by others									
Barry and Messer (2003)	ADHD	12	5 M	Classroom of elementary school	On-task behavior, disruptive behavior, academic performance	Multiple baseline across five students with embedded withdrawal	Goals based on classroom behavior of other students, self-recording, teacher verbal/written prompting, reinforcement, fading	Self-management intervention was effective, gradual fading required; also influenced academic performance as well as classroom behavior	Yes
Brobst and Ward (2002)	Soccer players	15–17	3 F	General education high school	(i) Movement with the ball. (ii) Movement during restarts. (iii) Movements after ball passed	Multiple baseline across behaviors for each participant	Public posting of daily practice results placed near playing field; Coach established goal setting; Oral feedback; praise for meeting goals or encouragement if not met	Intervention was effective in improving practice performance for all three behaviors; generalization to games evident for movement with the ball for all three participants but unclear for other behaviors	Yes
Coddling, Lewandowski and Eckert (2005) *	ADHD	9–12	1 M, 1 F	Elementary school	Increase mathematics fluency	Alternating treatments	Comparison: performance feedback and experimenter goal setting and Performance feedback and student selected goal setting or instruction; performance feedback; experimenter set goals (percentile shaping) student selected prize for exceeding goal	Both participants performed better during the self-set goals condition; both participants indicated that PFSG was their favourite intervention	No
Duhon et al. (2004)	Failing grades in written language, spelling math	8–10	4 M	General education	Math: work completed correctly per minute; reading/writing: words correct per minute	Alternating treatments	Practice, guided organization, or instruction; performance feedback; experimenter set goals (percentile shaping) student selected prize for exceeding goal	Two student rewards were effective in the brief assessment suggesting a performance deficit hypothesis; for two a skill deficit hypothesis; extended analysis conducted to compare skills-based treatment with performance-based treatment	No
Farrell and McDougall (2008)	SED/ADHD/BI/Polar (1) ADHD (2) SLD—Math (4) ADHD/Tourettes (1)	14–15	4 M, 2 F	Special education math at high school	Correct digits, incorrect digits, accuracy	Combined multiple baseline across participant, changing criterion	Self-monitoring pace using: MotivAider tactile cuing, visual cuing on worksheets, daily self-graphing, teacher-set goals	One participant omitted given unstable baseline; five participants met or exceeded lower performance limit in nearly every session; increase in pace not associated with a decrease in accuracy, most participants also increased overall percentage of digits completed accurately	Yes
Figarola et al. (2008)	Learning disability (1) ADHD (2)	7–8	3 F	Special education resource room in public elementary school	Performance of 40 one-digit by one-digit sums from 0 to 18	ABAB withdrawal (2) AB with reduction in sets (1)	Goal setting based upon baseline, self-graphing of daily results, verbal praise	Goal setting met or exceeded aim line on 83 and 90 % of occasions; multiple modifications to reduce sets, with verbal prompting for third participant, described as less successful	No
Gross and Duhon (2013)	Math skill deficits	8–9	3 F	Elementary school	Math problem accuracy	Non-concurrent multiple baseline across subjects	Computer-aided instruction with prizes; accuracy goal set at 90 %; teacher training in instruction, modelling and rehearsal with performance; feedback	Findings demonstrated an increase in math fact accuracy. All three students met accuracy goal at some point using CAI	No
Hanel and Martin (1980)	Mental retardation	19–54	8 M	Residential sheltered workshop	Coffee pack assembly rate per hour; percentage correct	Combined multiple baseline, reversal; delayed treatment control group	Self-monitoring, self-administration of token reinforcement, goal setting determined by other	Increased productivity of all clients, some considerably, some minimally	No
Lambert et al. (1999)	Gymnasts	12–13	4 F	Gym	Within subjects alternating treatment	Self-set goals, coach-set goals		Gymnasts with an internal locus of control spent more time on-task	No

Table 1 (continued)

Author (year)	Diagnosis	Age	M/F	Setting	Target behavior	Research design	Intervention elements	Findings	generalization/ maintenance data
Ward et al. (1998)	PE: volleyball	Grade 4	14 M 10 F	General education—PE volleyball skills	On-task behavior when working on beam during competition training Opportunity to respond: number of trials performed, number correct per minute	Multi-treatment A-B-A-C reversal	Group instruction, peer dyads, peer mediated accountability, teacher determined goal setting, peer recording and public posting, reinforcement	during self-set goal condition; gymnasts with an external locus of control spent more time on-task during coach-set goal condition Little difference between group instruction and peer dyad condition; highest number of or's and correct trials during peer-mediated accountability phase; using peers as data collectors created an instructional environment where teacher could hold students accountable for performance, recording created a product of performance All students increased journal writing quantity; goals set to be attainable but challenging; all teachers reported intervention easy to implement and required few resources, student like the intervention, one student appeared to improve performance in other academic areas	No
Winn et al. (2004)	N/A; referred for behavioral consultation	7–9	2 M, 1 F	General education in three inner-city schools	Number of words written in daily journal	Non-concurrent multiple baseline using three AB designs	Goal setting; self-recording: counting words and recording on goal charts, reinforcers	Increased number of story elements used when writing fictional stories, quality of the story and average number of words. Generalised across task, from fictional to narrative writing, and maintained at four week follow-up	Anecdotal, no data
Guided goal setting Asaro-Saddler and Saddler (2010)	Asperger's syndrome (1) Autism (2)	6–9	3 M	Public elementary, various empty classrooms	Improve story writing length, number of story elements, quality	Multiple probes across multiple baseline design	Goal setting (participant and instructor), self-monitoring, self-recording, self-reinforcement	Increased number of story elements used when writing fictional stories, quality of the story and average number of words. Generalised across task, from fictional to narrative writing, and maintained at four week follow-up	Yes
Grossi and Heward (1998)	Mild mental retardation	20–37	4 M	Community-based restaurant training program	(a) Work productivity, (b) Work quality/accuracy, (c) Accuracy of self-monitoring work performance	Multiple baseline across tasks	Self-evaluation training, self-set goals with guidance, self-monitoring, verbal direction, modelling, rehearsal, corrective feedback, social praise, self-adjustment of goal	Increase in work productivity for all four trainees, trainees increased their rates of work without sacrificing work quality, results suggest complete accuracy is not necessary for improved work performance	No
Lee and Tindal (1994)	Typically developing students	10–12	3 M, 2 F	Regular education classroom in Korea	(i) Percentage of time on-task, (ii) Mathematics productivity rate completed accurately per minute	Alternating treatments	Self-recording; self-set goals with guidance; combined goal setting and reward for achieving daily goal (teacher provided snack); feedback	All students increased on-task levels upon implementation of SR and GS; all students levels in SR and GS higher than baseline for math productivity; because of ceiling effects and results suggest effects of SR and GS not consistently different	Yes
McCarthy et al. (2010)	Junior multi event athletes	12 (2), 13 (1)	3 F	Athletes club track	(i) Positive affect, (ii) Negative affect	Multiple baseline across participants	Coach assisted goal setting: diary to detail process goals and goal setting; feedback and reinforcement	Demonstrated statistically significant increase in positive affect for all participants; no statistically significant decrease in negative	No

Table 1 (continued)

Author (year)	Diagnosis	Age	M/F	Setting	Target behavior	Research design	Intervention elements	Findings	generalization/ maintenance data
Merriman and Coddling (2008)	ADHD	HS	2 M, 1 F	General high school, special ed teacher support	(i) Homework completion. (ii) Homework accuracy	Multiple baseline across participants	Coaching with guided goal setting, self-monitoring of homework completion, accuracy and progress, graphing of completion and accuracy, fading	affect; participants set appropriate process and performance goals with guidance and mastered goals All students increased percentages for homework completion and accuracy, maintained during fading and follow-up for two participants (not implemented for third participant as did not meet accuracy, fading long-term goal, however increase considered adequate)	Yes
Miller and Kelley (1994)	Conduct and learning problems, hyperactivity, inattention/passivity	9–11	2 M, 2 F	General education	(i) Accuracy of completed homework. (ii) On-task behavior	Reversal ABAB multiple baseline	Child, parent goal suggestions then determined a compromised goal, parent monitoring, contingency contracting, daily and weekly rewards or sanctions	Increased work accuracy for all participants and increased stability of responding across phases for three of four participants, two participants demonstrated increases in percentage of on-task behavior	No
Moore et al. (2001)	Typically developing students	8	3 M	General education	On-task behavior	Multiple baseline across subjects	Self-monitoring with self-recording using intervals of on-task at the time, tallied at end and graphed; goal setting with assistance and on-going performance feedback; fading	Substantial increase in on-task behavior for all three students during self-monitoring/self-recording; performance levels maintained during goal setting; maintained during fading for two participants	Yes
Srikameswaran and Martin (1984)*	Mentally handicapped	25–36	2 M, 2 F	Sheltered workshop training centre	(i) Production of bags of ceiling tile clips. (ii) Percentage of correct production	Modified multiple baseline with reversal-replication	Self-monitoring plus goal setting (guided during training); token exchange quarters; edibles/activities; general/ specific prompts verbal interactions; feedback; pay system	SRP(SMGS+reinforcer) was effective at increasing production rate above baseline, two participants did not respond to quarters but demonstrated small increases for edible/activity reinforcers; SRP more effective than SM or SMGS for three of four participants	No
Tollefson et al. (1986)	Learning disabled	Grade 7 (6) Grade 8 (2)	7 M, 1 F	Middle school resource room	Classroom assignment completion, homework assignment completion	Multiple baseline across eight students classroom assignment comparison group for homework assignment	Contract: goal statement, plan to meet goal, and evaluation of appropriateness of goal and plan. Student selected achievement goal; student developed a specific goal statement/plan, student chart, feedback	Students can learn to set realistic achievement goal, develop plans to reach these goals and monitor progress towards goals. Students increased rates of assignment completion and were more confident of their ability to plan	No
Goals set by participants * also included Coddling et al. (2005); Srikameswaran and Martin (1984); Lambert et al. (1999)									
Albertson and Billingsley (2001)	Gifted students	13	1 M 1 F	Enrolled in school honours program,	Creative writing planning, text production, rates of writing, reviewing and writing quality	Multiple baseline across participant	Student-set goal setting: planning and time recording; story writing and time recording; reviewing,	Strategy instruction plus self-regulation (goal setting and graphing) enhanced story length, fluency (words/minute) and story	Yes

Table 1 (continued)

Author (year)	Diagnosis	Age	M/F	Setting	Target behavior	Research design	Intervention elements	Findings	generalization/ maintenance data
Balcazar et al. (1991)	Physical disabilities	20–40	2 M, 2 F	conducted in home office of researcher Large mid-western university	Independently recruit help to attain personal goals	Interrupted time series multiple replication across participants	changes and time recording; self-recording and graphing; feedback with verbal praise Role-play assessment; written training manual of help-recruiting skills with list of 25 verbal responses; feedback and praise	element scores for both participants; participants met or exceed goals 88 % of time; written quality of stories written quickly during intervention was judged as equal to or better than prior sessions Comparison of scores before and after training indicated significant improvement in participants help-recruiting skills; participants set 20 goals with total of five reported as no change, eight met realistic attainment and seven best anticipated attainment outcomes Yes	
Balcazar et al. (1995)	behavioral disorder (4), mental retardation (1), manic-depressive (1)	17–19	6 M	Residential treatment school	(i) Transition goals, (ii) Action towards reaching goals. (iii) Help-recruiting skills	Multiple baseline design across behaviors (as a group)	Self-goal setting, self-recording in student diary, role-play of help-recruiting skills	Consistent improvements in help-recruiting skills of group; all students demonstrated substantial efforts and reported in goal attainment (11/17 goals for group attained, 2/2 goals dropped and 4/17 goals in progress) Yes	
Benitez et al. (2005)	Emotional behavioral disorder	15–17	5 M	Special education	Conflict resolution (1), assertiveness (2), career exploration (2)	Five AB designs, identical goals and outcome measures	Goal setting with goal training, goal attainment measures, self-monitoring	Increased correct responses by condition for all participants, all participants achieved their target goals Yes	
Caldwell and Joseph (2012)	Conduct disorder and substance abuse (2); emotional disturbance (1)	14, 17, 18	3 F	Maximum-security juvenile facility	Time on-task, academic productivity, academic accuracy during independent math practice	Single-case reversal design across participants	Set and record productivity and accuracy goals, self-monitoring, fading	Students were taught to graph performance data and set performance and accuracy goals answering question to maintain or increase previous result. Corrective feedback was provided. Accuracy did not increase during intervention, direction instruction on math problems was not provided. Sample size small, incentive rewards not immediate. Baselines long, possible that intervention should have occurred sooner No	
Delano (2007)	Asperger's syndrome	13–17	3 M	General education classes, instruction in conference room	Words written, functional essay elements	Multiple baseline design across responses and participants	Self-monitoring, self-recording, self-goal setting, verbal/written story starter/video prompts	Used video self-modelling to teach SRSD and self-monitoring. Increased number of words, functional essay elements in persuasive essay writing. Two participants maintained increase in number of words at 3-month follow-up, functional essay elements were not maintained Yes	
Graham et al. (1992)	Learning disability	11 (2), 12 (1), 13 (1)	3 M, 1 F	Resource room in elementary school	Improve writing composition	Multiple probe design across subjects	Students directed to list ways to meet goals; self-prompt evaluation, reflection;	Improved performance in essay writing in each goal area: components, length and Yes	

Table 1 (continued)

Author (year)	Diagnosis	Age	M/F	Setting	Target behavior	Research design	Intervention elements	Findings	generalization/ maintenance data
Holt et al. (2012)	Professional soccer club academy players	10–12	5 M	Indoor coaching	Technical drill “passing square”: awareness, passing and first touch	Single-subject, multiple baseline with reversal	Peer-assessed feedback with carting, goal setting and group contingency; players self-set personal goals and following agreement with coach charted goal line	convincing. Post-treatment two times as many elements and four times more basic parts. Essays were two to three times longer, more convincing. Maintained gains over time, student and teacher reported strategy improved writing performance	Yes
Kelley and Stokes (1984)	High school dropouts with academic and social deficits	16–21	7 M 1 F	Vocational training program	Number and percentage of academic items answered correct	Contracting; reversal; multiple baseline	Baseline payment based on attendance; contingency contracting payment for number of correct workbook items plus \$13.50 for completing weekly goal (or % thereof); student-set goal setting daily/weekly payment	Student-teacher contracting with pay was effective in increasing student productivity; student goal setting effective for maintaining student productivity substantially higher than baseline, although generally lower than during contracting conditions	No
Lenz et al. (1991)	Learning disability	13	5 M 1 F	6-week summer program	Improve completion of project-type assignments	Multiple baseline across subjects	Goal setting: task evaluation, options, specification; Actualisation: plan identification, plan expansion, self-monitoring, feedback	Effective in increasing the number of projects successfully completed and the quality of goal setting and goal actualisation responses; all students demonstrated an increase of at least 20 % over their baseline averages in both goal setting and goal actualisation	No
Lyman (1984)	Conduct disorder	12	6 M	Residential treatment facility	On-task behavior	Multiple baseline across six subjects	Private student-set goal; public student-set goal, goal card on bulletin board at front of class	Percentage of time on-task slightly higher in private goal setting than baseline; percentage of time on-task substantially higher in public goal setting than either baseline or private goal setting; public goal setting may be more effective than private goal setting	No
Moore et al. (1989)	Severe mental retardation	19–21	2 M, 2 F	Sheltered workshop	Rate for sorting washers; self-instructions	Multiple baseline across subjects	Self-talk training, self-goal setting, self-reinforcement, verbal praise, corrective feedback, modelling	All workers increased production to criterion levels, increases maintained up to 3 months, productivity consistent throughout work periods, work production increase under self-reinforcement was modest and perhaps money was not reinforcing	No
O'Brien et al. (2009)	Elite (3) and non-elite (3) boxers	15–17	6 M	Amateur boxing association		Multiple baseline across individuals	Participant goal determination, participant goal setting, goal	All elite participants immediately improved performance, sustained	No

Table 1 (continued)

Author (year)	Diagnosis	Age	M/F	Setting	Target behavior	Research design	Intervention elements	Findings	generalization/ maintenance data
Todd et al. (2010)	Autism	15–17	2 M 1 F	Cycling course at special education school	(i) Performance (number of punches), (ii) Competition outcomes	Multiple baseline changing criterion	reviewing, reinforcement for goal attainment, feedback	during retention phase; all non-elite participants immediately improved performance, not sustained during retention phase. Non-elite mean percentage of wins increased during intervention and retention for two participants and showed no increase during intervention and a decrease at retention for one participant	
Trammel et al. (1994)	Learning disability	13–16	6 M 2 F	General education resource room support	Homework assignment completion	Multiple baseline across subjects	Self-monitoring, self-evaluation, goal setting, self-graphing	Immediate increase of homework assignment completion under both self-monitoring and self-graphing conditions; results maintained at high levels during maintenance phase and at 20- and 40-day follow-up probes	Yes
Wanlin et al. (1997)	Speed skating performance	12–17	4 F	Provincial team preparing for National Winter Games	(i) Lap frequency, (ii) Number of drills completed, (iii) 500-m race times, (iv) Off-task behavior	Multiple baseline across individuals (1 control)	Mission development, long-term self-setting, sub-goal and practice goal setting, self-talk, goal visualisation, self-monitoring and recording; prompt, self-evaluation and set new practice goals; public goal sharing	Immediate increase in laps for all three participants, control initially increased then dropped below mean levels of both baseline and intervention; all participants improved number of drills, control participant maintained stable level of completion; all participants improved mean racing times, control participant chose not to race; off-task behaviors immediately dropped for two participants, one participant reported an almost complete absence of off-task behavior, control participant had high frequency of off-task behavior throughout study; goal achievement at 72, 76 and 71 %	No
Goals introduced in latter phase Maag et al. (1992)	Learning disability	7–11	4 M 2 F	Regular classroom	(i) On-task behavior, (ii) Academic productivity	Multiple treatment design across participant	(i) Self-observation, (ii) Self-observation and self-recording, (iii) Self-observation, self-recording	Substantial changes in on-task behavior and academic productivity for each student in each experimental condition	Yes

Table 1 (continued)

Author (year)	Diagnosis	Age	M/F	Setting	Target behavior	Research design	Intervention elements	Findings	generalization/ maintenance data
Skinner et al. (1993)	behavioral disorders	9	1 M	Private special education	(i) Percentage of division problems correct (ii) Number of digits and/or problems correct per minute	Multiple baseline across-problems within subjects	and contingent reinforcement phase 1. (iv) Self-observation, self-recording, contingent reinforcement phase 2 Guided goal setting Cognitive cover, copy, and compare (CCC); one participant required assessment feedback and teacher-set goal setting to reach mastery level	except self-observation alone in which effects were negligible. Increased performance of on-task behavior and academic productivity most substantial during contingent-reinforcement phases CCC effective at increasing students' rate of accurate responding to division problems; assessment feedback and goal setting suggest that these procedures alone may result in increases in rates of accurate responding	Yes

How Have Goal-Setting Interventions Been Structured?

Interventions that included goal setting as part of treatment packages and alongside other intervention components were frequently described. These included contracting (3), self-monitoring (25), reinforcement (15) and feedback (17). The use of technology was also described and included computer instruction (1) and video modelling (1). An approach to teach goal setting was described in six studies (16 %), and the use of self-regulated strategy development (SRSD) techniques was reported in two studies (5 %). Public goal sharing was included as an intervention element in four studies (11 %).

Lyman (1984) explored the effects of private student-set goals and public sharing of student-set goals with six students schooled in a residential treatment facility as a result of conduct disorder. During the public condition, a goal card was placed on the bulletin board at the front of the class. Participants reportedly spent a slightly higher percentage of time attending to task during the private goal setting condition and substantially more time on task upon introduction of public goal sharing.

Goal setting was introduced during the latter stage of intervention in two studies (5 %). One study examined the effects of self-monitoring with contingent reinforcement under two conditions: comparison to prior performance and contingent to meeting a goal (Maag et al. 1992). Another study introduced goals in the latter phase in order for one participant to achieve mastery of math division (Skinner et al. 1993).

Maag et al. (1992) introduced goal setting in the final phase of an intervention that included the following phases: a self-observation condition; self-observation and self-recording; self-observation (reversal), self-recording and contingent verbal reinforcement for performance improvement; and self-observation, self-recording and contingent reinforcement using teacher- and student-set goals for successive increases in mathematics problem task performance. The largest gain in academic productivity occurred during the phase of contingent reinforcement plus goal setting, during which the teacher not only provided verbal reinforcement for an increase over prior performance but also provided verbal encouragement in the event a goal was not met. The authors concluded that treatment effects can be increased by the simultaneous use of goal setting. However, the effect of goal setting alone was not explored.

While the original authors of these 38 studies have consistently reported positive treatment effects, the majority of the studies have confounded goal setting with other dependant variables. Three studies were identified which enable the effectiveness of goal setting alone to be examined. Lambert et al. (1999) compared self-set goals with coach-set goals, and

Table 2 Summary of descriptive statistics

	Descriptor	Participant numbers
Sex	Male	128
	Female	58
Age	School aged (6–12 years)	82
	Adolescence (13–17 years)	73
	Adulthood (18 years and older)	31
Diagnosis	Autism	5
	Asperger's syndrome	4
	ADHD	14
	Emotional behavioral disorder	14
	Learning disability	45
	Mental retardation	21
	Non-compliance/conduct	12
	Manic depressive	1
	Giftedness	2
	Typically developing	15
	Athletics	49
	Physical disability	4
Method	Taught prerequisite skills	1
	Taught goal setting	6
	Taught self-regulated strategy development (SRSD)	2
	Goals set by other	11
	Guided goal setting	9
	Guided goal setting, set by participant	20
	Public goal sharing	4
	Goals introduced in latter phase	2
Other components	Contracting	3
	Self-monitoring	25
	Reinforcement	15
	Feedback	17
	Used computer instruction	1
Target behavior	Used video modelling	1
	Transition planning	2
	Problem behavior	8
	Mathematics	9
	Writing	6
	Reading	1
	Academics—all subjects	3
	Homework	4
	Projects	1
	Social skills	1
	Athletics	7
	Development of physical activity skills	1
	Vocational tasks	4

both Maag et al. (1992) and Skinner et al. (1993) introduced a goal condition additively to achieve an effective treatment thereby enabling intervention effects to be measured before and after the introduction of goal setting. In each case, goal setting was associated with clear behavioral gains.

What Types of Target Behavior Was Goal Setting Applied To?

These interventions addressed a variety of target behaviors including transition planning (2), decreasing problem

behavior (8), mathematics (9), writing (6), reading (1), all academic subjects (3), homework (4), projects (1), social skills (1), athletic performance (7), development of physical activity skills (1) and vocational tasks (4). Target behavior was plotted by number of participants, as illustrated in (Fig. 1).

Barry and Messer (2003) incorporated goal setting with self-management to increase on-task behavior and academic performance and decrease disruptive behavior in an intervention conducted with five boys with ADHD. The teacher set participant goals using the approximate average for the class in an intervention that included self-recording, teacher prompting and reinforcement. Goals for academic achievement and on-task behaviors were gradually increased during fading. While results again showed this to be an effective intervention procedure, researchers noted that gradual fading of supports was necessary to ensure maintenance.

Tollefson et al. (1986) targeted an increase in completion of classroom assignments and an increase in homework assignment completion for eight middle school students described as learning disabled. During training with the research assistant, each student selected an achievement goal: daily or weekly in-class assignments, or homework assignments for the week, and developed a goal statement of what was required and by when. A study plan was specified, and each student self-monitored work completion. Subsequently, students met with the research assistant to evaluate the goal and the plan, addressing reasons for success or failure in goal attainment. A homework contract phase was introduced for four students, intended to promote generalization of goal implementation skills. Three students demonstrated a marked increase in homework assignment completion. The authors reported that

time spent teaching goal setting and goal implementation strategies not only increased the rate of assignment completion but also made students more confident in their ability to plan.

A goal setting model that included goal determination, goal setting and goal reviewing was used to examine the cumulative effects on skill performance among six high school boxers (O'Brien et al. 2009). During goal determination, participants met with the researcher to determine a self-generated performance behavior. All participants rated the number of punches landed as most important, thus determining this as the target behavior. During goal setting, a numeric system to measure changes in the participants' target expectations over time as well as a means to measure progress during intervention was developed. Three elite level participants immediately improved performance, and improvement was sustained during a follow-up phase. However, while the three non-elite participants also immediately improved performance, their increase was not sustained during follow-up.

Individuals with ASD are often described as having an uneven profile of executive functioning skills, an area of importance for achieving academic or vocational success (Geller and Greenberg 2010). While high-functioning students on the spectrum may be academically capable, problems with organisation and planning may negatively impact success in higher education or vocational settings and present a challenge for independent functioning. The application of goal setting to behavior management, attention to task and time management described in the existing literature raises the possibility that effective goal setting instruction may be beneficial to students with ASD in learning skills necessary to

Target behavior by participants

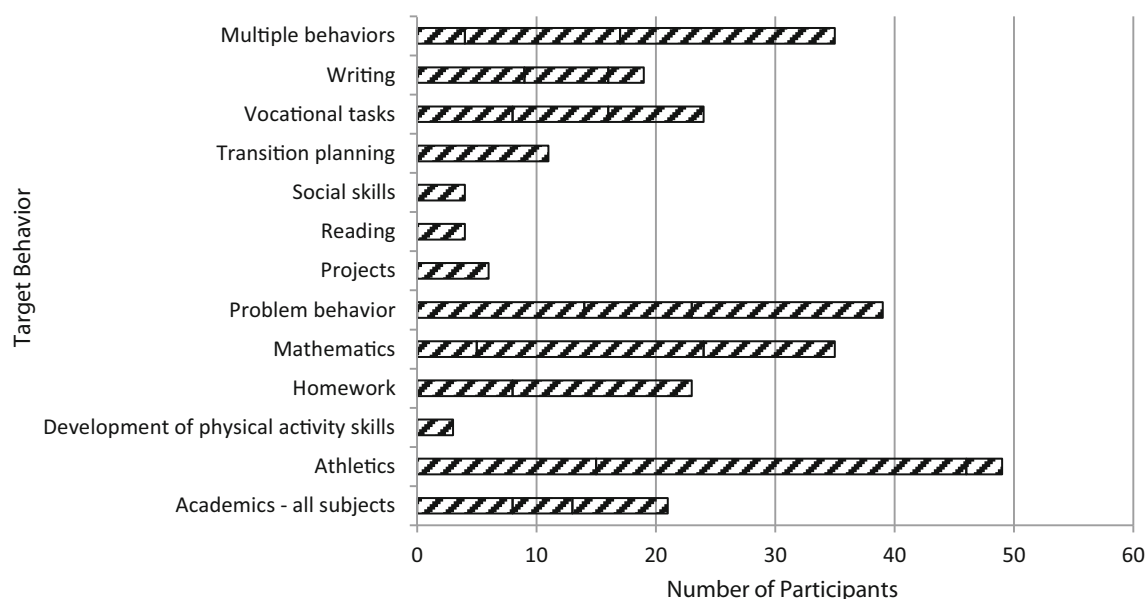


Fig. 1 Target behavior by participants

overcome these challenges and to perform tasks independently.

Who Is Involved in Setting Participant Goals?

The studies in this data set described a variety of stakeholders involved in the goal setting process. Twenty studies (53 %) described goal setting by the participants, 9 studies (24 %) described the participant setting goals under guidance, and in 11 studies (29 %), goals were set by someone other than the participant.

Two studies employed an alternating treatment design to compare the effects of goal setting by either the experimenter/coach with goals that were set by the participant (Coddington et al. 2005; Lambert et al. 1999). While Coddington and colleagues (2005) reported that participants not only performed better during the student selected goals condition but also favoured this condition over the experimenter-selected goal condition, Lambert and colleagues (1999) reported that self-set and coach-set goals functioned differently depending on the participants' locus of control. In particular, an internal locus of control was associated with greater gains under a self-set goal condition, and an external locus of control was associated with greater gains under a coach-set goal condition.

Participant descriptions were plotted against the differences in the identity and role of the goal setter (self vs other vs guidance provided) for the studies reviewed, as illustrated in Fig. 2. The graph illustrates a potpourri of self, other and guided approaches to goal setting. The exclusive use of independently set goals was noted amongst participants described

either as gifted or physically disabled, while goals were either set independently by the participants or under guidance for participants with ASD, AS, conduct disorder and students with manic depression. Typically developing students in this data set did not set goals independently and received guidance with setting goals only marginally more frequently than having goals set by an adult. A few athletes received guidance with goal setting, and athletes set goals independently in just over one third of the data. The majority of participants with learning disabilities self-set goals, although both guidance and goals set by another adult were observed. Few students with ADHD set goals independently, and occasional provision of guidance was observed in these studies though the majority of these participants had goals set by an adult. This diversity in approaches may reflect different purposes on the part of the researchers when adopting goal setting in interventions. For studies targeting increased performance, such as in the case of athletes, it may arguably be less relevant for the participants to independently set their own performance goals. By contrast, where the aim of the intervention is to enhance independent functioning, developing social skills, transition planning or reducing problem behavior for example, teaching participants to set their own goals either independently or with guidance may be of greater relevance.

By way of example, three non-verbal high school students diagnosed with ASD were included in an intervention that aimed to increase participation in sustained physical activity (Todd et al. 2010). The students possessed the physical skills required to cycle and were able to discriminate correctly between more or less necessary to set distance goals. Prior to

Goal setting method used by participants

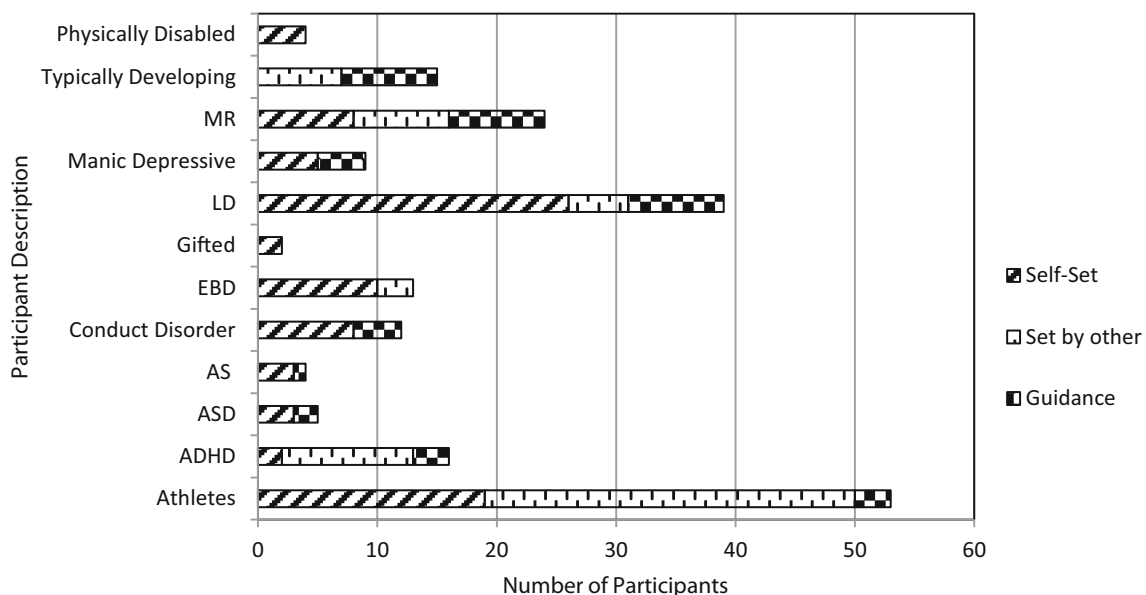


Fig. 2 Goal setting method used by participants

independently setting cycling distance goals, participants were given guidance to understand their cycling capabilities and were taught to self-monitor. All three students developed accuracy in setting both distance and intensity goals, and two of the students increased the distance they cycled.

However, for interventions in which successful completion of task is less related to attaining independence, the identity of the goal setter may be less significant. For example, guided goal setting, combined with self-monitoring and performance feedback, formed the basis of a coaching intervention in an intervention conducted with three high school students with ADHD (Merriman and Coddling 2008). An initial goal setting meeting was conducted between each student and the coach. Students used their baseline data to set goals for mathematics homework completion and accuracy, and the coach gave guidance in the event of unrealistic goal suggestions. Coaching was conducted daily with feedback and goal reformulation, and praise for progress was provided until the students achieved their long-term goals. During systematic fading, coaching sessions went to alternate days then weekly until eventual termination. Results showed that coaching improved the completion and accuracy of homework for all three participants, with improvements maintained during fading and follow-up for the two participants that met their long-term goals.

In contrast, interventions that do not impact the participants' ability to achieve independent functioning may not require the participants themselves to set goals in order to achieve positive outcomes. For example, Brobst and Ward (2002) reported a study with three teenaged female soccer players to increase performance of existing ball skills in both practice sessions and games. The head coach and the researcher established a performance level for the targeted skills, and an intervention package that included public posting, goal setting and oral feedback was implemented. Both the goal and the rationale behind it were explained to the participants, and practice results were reported on a performance chart displayed near the playing field. Players were praised for meeting goals and given encouragement if they did not. Results demonstrated improvement during practice. Generalization results were less consistent, and improvement during games was apparent for only one of the three ball skills.

Are Goal Setting Effects Maintained Over Time and Was Generalization Assessed?

While frequent informal reports of generalization and maintenance over time were included by the original authors of studies in this set, data confirming such effects was reported in 18 of the 38 studies (47 %). Three of the 38 studies included participants with ASD (Asaro-Saddler and Saddler 2010; Delano 2007; Todd et al. 2010), and notably, all three of these

articles included data for either generalization or maintenance of treatment effects over time.

Inspection of generalization and maintenance data for the participants with ASD revealed variable findings. Asaro-Saddler and Saddler (2010) included a generalization task in an intervention conducted with two elementary students with AS, and one with autism, in which the writing requirement altered from fiction to personal recount. While two participants appeared to successfully generalize writing skills to the new task, one participant included a fictional name rather than "I". The authors reported that participants may not have clearly distinguished the required difference. Similarly, data gathered to explore maintenance of treatment effects over time also resulted in variable findings. Todd et al. (2010) examined cycling distance goals in the maintenance phase of an intervention with three non-verbal teens. Gains in excess of baseline were maintained for two of the three participants during the phase. Delano (2007) collected follow-up data at both 1-week and 3-month time frames for three teens with AS. Although a gain in the number of words written was maintained by all three participants at 3 months, the number of functional essay elements was not maintained. Delano (2007) suggested that a longer intervention period may be required. Asaro-Saddler and Saddler (2010) also collected maintenance probes after 4 weeks and acknowledged the limitation of their data as continued gains beyond this time frame remain unknown.

Stokes and Baer (1977) described generalization across subjects, settings, people, behaviors and/or time as the occurrence of relevant behavior under different untrained conditions in the absence of conditions that had been scheduled during training. At that time, attention was drawn to the importance of actively programming to achieve generalization, rather than to passively expect it as an outcome of training procedures. In addition, the use of stimuli found in the generalization settings, which included the role of peers as tutors, was highlighted as significant to future research when structuring training sessions.

Authors of the original studies from our broader data set generally concur that graduated fading is necessary in order to maintain intervention effects over time or to observe generalization of effects to new settings or behaviors. Several studies reported successful maintenance of treatment effects over time among all participants in research that included systematic stimulus fading (Moore et al. 2001; Merriman and Coddling 2008).

Conclusions

Our aim in this study was to map the existing single-subject research evidence for interventions that have included a goal

setting component with a view to informing a research agenda for participants with ASD. This data set as a whole provides preliminary support for the effectiveness of goal setting techniques in a wide variety of interventions. However, to date, little research on goal setting has been performed with individuals who have a diagnosis of ASD.

Second, in the research identified, goal setting effects have almost always been confounded with interventions typically involving a combination of components additional to goal setting. The most common of these, self-monitoring, was included in two thirds of the interventions. Several of these studies noted that participants value both establishing their goal independently and self-monitoring their behavior. Feedback was also included in almost half of the interventions, and again, original author reports suggest that information derived from feedback played a significant role in positive intervention outcomes.

Southall and Gast (2011), in their qualitative review of self-management interventions for students with ASD, emphasised the need for researchers to examine the effects of specific components of current self-management intervention packages to identify the relative contribution of each. Our findings support this conclusion, highlighting in particular the need to explore the contribution goal setting has, either in isolation or in combination with other intervention components, on intervention outcomes and on the development of self-determination more generally.

Another important observation arising from this data set concerns the generalization and maintenance of observed changes. Goal setting skill training may be an important component of treatment packages that aim to develop independent functioning for participants with ASD, and effective goal setting may be a vital skill for high-functioning students who pursue higher education. Given that goal setting skills are developed over time, it appears particularly important to gather generalization and maintenance data. While limited in volume, the preliminary maintenance data for participants with ASD is variable, including accounts of maintenance failure post-intervention. Accordingly, we reiterate the importance that future research includes longer intervention phases, in addition to programming for generalization and maintenance over longer time frames and/or including peers.

With respect to how participants have been involved in the goal setting process within an intervention, no clear pattern is evident in our data set. The basis on which researchers have decided either to set goals themselves or how and to what extent to involve the participants in this process is unclear, and importantly, the implications of these decisions are largely unexplored. This review has identified goal setting research that has been conducted successfully with a wide variety of individuals engaged in the process of learning. Although the authors of this work have consistently reported successful outcomes in interventions that have included goal setting

and these accounts suggest that goal setting skills are teachable, further research that explores the relative merit of engaging participants in the goal setting process and that identifies effective strategies for teaching goal setting appears warranted.

Finally, although the literature has included successful accounts of individuals improving sports performance through the use of goal setting skills, there is a paucity of research with participants with ASD regarding the development and/or improvement of skills required to promote a physically active lifestyle. This finding is of significance to this population given the recent reports of elevated chance of developing obesity, in addition to the loss of opportunity to engage in social interactions with typically developing peers. Research that explores treatments aimed to overcome challenges that may result from a sedentary lifestyle, food selectivity or disrupted sleep appears justified.

Proposed Research Agenda for Participants with ASD

While limited in quantity, the SCD research with participants on the autism spectrum included in this review suggests that goal setting instruction may be successful with this population. We propose the following foci as elements of future research to aid in developing treatment packages that foster independent living:

1. Research that explores the role of goal setting in isolation or in combination with other intervention components
2. Research that explores the generalization and maintenance effects of goal setting interventions, particularly including peers as tutors to assist in programming generalization
3. Research that explores how to best teach goal setting skills to participants with a view to attaining the ability to perform this skill independently
4. Research that explores the relative effects of goal setting by the participants themselves compared to goals set by other stakeholders

References

*Studies included in the review

- *Albertson, L. R., & Billingsley, F. (2001). Using strategy instruction and self-regulation to improve gifted students' creative writing. *The Journal of Secondary Gifted Education*. doi:10.4219/jsge-2000-648.

- Algozzine, B., Browder, D., Karvonen, M., Test, D. W., & Wood, W. M. (2001). Effects of interventions to promote self-determination for individuals with disabilities. *Review of Educational Research*. doi:10.3102/00346543071002219.
- *Asaro-Saddler, K., & Saddler, B. (2010). Planning instruction and self-regulation training: effects on writers with autism spectrum disorders. *Exceptional Children*. doi:10.1177/1053451208330895.
- *Balcasar, F. E., Fawcett, S. B., & Seekins, T. (1991). Teaching people with disabilities to recruit help to attain personal goals. *Rehabilitation Psychology*. doi:10.1037/h0079072.
- *Balcasar, F. E., Keys, C. B., & Garate-Serafini, J. (1995). Learning to recruit assistance to attain transition goals: a program for adjudicated youth with disabilities. *Remedial and Special Education*. doi:10.1177/074193259501600407.
- *Barry, L. M., & Messer, J. J. (2003). A practical application of self-management for students diagnosed with attention-deficit/hyperactivity disorder. *Journal of Positive Behavior Interventions*. doi:10.1177/10983007030050040701.
- *Benitez, D., Lattimore, J., & Wehmeyer, M. (2005). Promoting the involvement of students with emotional and behavioral disorders in career and vocational planning and decision-making: the self-determined career development model. *Behavioral Disorders*, 30, 431–447.
- *Brobst, B., & Ward, P. (2002). Effects of public posting, goal setting, and oral feedback on the skills of female soccer players. *Journal of Applied Behavior Analysis*. doi:10.1901/jaba.2002.35-247.
- *Caldwell, S., & Joseph, L. M. (2012). Helping female juveniles improve their on-task behavior and academic performance using a self-management procedure in a correctional facility. *Contemporary School Psychology*, 16, 61–74.
- Carr, M., Moore, D. W., & Anderson, A. (2014). Self-management interventions on students with autism: a meta-analysis of single-subject research. *Exceptional Children*. doi:10.1177/0014402914532235. in press.
- *Coddling, R. S., Lewandowski, L., & Eckert, T. (2005). Examining the efficacy of performance feedback and goal-setting interventions in children with ADHD: a comparison of two methods. *Journal of Evidence-Based Practices for Schools*, 6, 42–58.
- Copeland, S. R., & Hughes, C. (2002). Effects of goal setting on task performance of persons with mental retardation. *Education and Training in Mental Retardation*, 37, 40–54.
- Curtin, C., Jovic, M., & Bandini, L. G. (2014). Obesity in children with autism spectrum disorder. *Harvard Review of Psychiatry*. doi:10.1097/HRP.0000000000000031.
- *Delano, M. E. (2007). Improving written language performance of adolescents with Asperger syndrome. *Journal of Applied Behavior Analysis*. doi:10.1901/jaba.2007.50-06.
- *Duhon, G. J., Noell, G. H., Witt, J. C., Freeland, J. T., Dufrene, B. A., & Gilbertson, D. N. (2004). Identifying academic skill and performance deficits: the experimental analysis of brief assessments of academic skills. *School Psychology Review*, 33, 429–443.
- *Farrell, A., & McDougall, D. (2008). Self-monitoring of pace to improve math fluency of high school students with disabilities. *Behavior Analysis in Practice*, 1(2), 26–35.
- Field, S., Martin, J., Miller, R., Ward, M., & Wehmeyer, M. (1998). Self-determination for persons with disabilities: a position statement of the division on career development and transition. *Career Development for Exceptional Individuals*. doi:10.1177/088572889802100202.
- *Figarola, P. M., Gunter, P. L., Reffel, J. M., Worth, S. R., Hummel, J., & Gerber, B. L. (2008). Effects of self-graphing and goal setting on the math fact fluency of students with disabilities. *Behavior Analysis in Practice*, 1(2), 36–41.
- Fowler, C. H., Konrad, M., Walker, A. R., Test, D. W., & Wood, W. M. (2007). Self-Determination interventions' effects on the academic performance of students with developmental disabilities. *Education and Training*, 42, 270–285.
- Geller, L. L., & Greenberg, M. (2010). Managing the transition process from high school to college and beyond: challenges for individuals, families, and society. *Social Work in Mental Health*. doi:10.1080/15332980902932466.
- *Graham, S., MacArthur, C., Page-Voth, V., & Schwartz, S. (1992). Improving the compositions of students with learning disabilities using a strategy involving product and process goal setting. *Exceptional Children*, 58(4), 322.
- *Gross, T. J., & Duhon, G. (2013). Evaluation of computer-assisted instruction for math accuracy intervention. *Journal of Applied School Psychology*. doi:10.1080/15377903.2013.810127.
- *Grossi, T. A., & Heward, W. L. (1988). Using self-evaluation to improve the work productivity of trainees in a community-based restaurant training program. *Education and Training in Mental Retardation and Developmental Disabilities*. doi:10.1080/153779032013.810127.
- Halpern, A. S. (1994). The transition of youth with disabilities to adult life: a position statement of the Division on Career Development and Transition, The Council for Exceptional Children. *Career Development for Exceptional Individuals*. doi:10.1177/088572889401700201.
- *Hanel, F., & Martin, G. (1980). Self-monitoring, self-administration of token reinforcement, and goal-setting to improve work rates with retarded clients. *International Journal of Rehabilitation Research*. doi:10.1017/00004356-198012000-00004.
- *Holt, J. E., Kinchin, G., & Clarke, G. (2012). Effects of peer-assessed feedback, goal setting and a group contingency on performance and learning by 10–12-year-old academy soccer players. *Physical Education & Sport Pedagogy*. doi:10.1080/17408989.2012.690568.
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S. L., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71, 165–179.
- *Kelley, M. L., & Stokes, T. F. (1984). Student-teacher contracting with goal setting for maintenance. *Behavior Modification*. doi:10.1177/01454455840082005.
- Konrad, M., Fowler, C. H., Walker, A. R., Test, D. W., & Wood, W. M. (2007). Effects of self-determination interventions on the academic skills of students with learning disabilities. *Learning Disability Quarterly*. doi:10.2307/30035545.
- *Lambert, S. M., Moore, D. W., & Dixon, R. S. (1999). Gymnasts in training: the differential effects of self-and coach-set goals as a function of locus of control. *Journal of Applied Sport Psychology*. doi:10.1080/10413209908402951.
- *Lee, C., & Tindal, G. A. (1994). Self-recording and goal-setting: effects on on-task and math productivity of low-achieving Korean elementary school students. *Journal of Behavioral Education*. doi:10.1007/bf01539545.
- *Lenz, B. K., Ehren, B. J., & Smiley, L. R. (1991). A goal attainment approach to improve completion of project-type assignments by adolescents with learning disabilities. *Learning Disabilities Research & Practice*, 6, 166–176.
- Locke, E., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: a 35-year odyssey. *American Psychologist*. doi:10.1037/0003-066X.57.9.705.
- Locke, E., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969–1980. *Psychological Bulletin*. doi:10.1037/0033-2909.90.1.125.
- *Lyman, R. D. (1984). The effect of private and public goal setting on classroom on-task behavior of emotionally disturbed children. *Behavior Therapy*. doi:10.1016/s0005-7894(84)80008-8.
- *Maag, J. W., Rutherford, R. B., & DiGangi, S. A. (1992). Effects of self-monitoring and contingent reinforcement on on-task behavior and

- academic productivity of learning-disabled students: a social validation study. *Psychology in the Schools*. doi:10.1002/1520-6807(199204)2-1:2<157.
- *McCarthy, P. J., Jones, M. V., Harwood, C. G., & Davenport, L. (2010). Using goal setting to enhance positive affect among junior multievent athletes. *Journal of Clinical Sport Psychology*, 4, 53–68.
- *Merriman, D. E., & Coddling, R. S. (2008). The effects of coaching on mathematics homework completion and accuracy of high school students with attention-deficit/hyperactivity disorder. *Journal of Behavioral Education*. doi:10.1007/s10864-008-9072-3.
- *Miller, D. L., & Kelley, M. Lou. (1994). The use of goal setting and contingency contracting for improving children's homework performance. *Journal of Applied Behavior Analysis*. doi:10.1901/jaba.1994.27-73.
- *Moore, S. C., Agran, M., & Fodor-Davis, J. (1989). Using self-management strategies to increase the production rates of workers with severe handicaps. *Education and Training in Mental Retardation*, 24, 324–332.
- *Moore, D. W., Prebble, S., Robertson, J., Waetford, R., & Anderson, A. (2001). Self-recording with goal setting: a self-management programme for the classroom. *Educational Psychology: An International Journal of Experimental Educational Psychology*. doi:10.1080/01443410123466.
- *O'Brien, M., Mellalieu, S., & Hanton, S. (2009). Goal-setting effects in elite and nonelite boxers. *Journal of Applied Sport Psychology*. doi:10.1080/10413200903030894.
- Palmer, S. B., & Wehmeyer, M. L. (2003). Promoting self-determination in early elementary school: teaching self-regulated problem-solving and goal setting skills. *Remedial and Special Education*. doi:10.1177/07419325030240020601.
- Pinder-Amaker, S. (2014). Identifying the unmet needs of college students on the autism spectrum. *Harvard Review of Psychiatry*. doi:10.1097/HRP.0000000000000032.
- *Skinner, C. H., Bamberg, H. W., Smith, E. S., & Powell, S. S. (1993). Cognitive cover, copy, and compare: subvocal responding to increase rates of accurate division responding. *Remedial and Special Education*. doi:10.1177/074193259301400107.
- Southall, C. M., & Gast, D. L. (2011). Self-management procedures: a comparison across the autism spectrum. *Education and Training in Autism and Developmental Disabilities*, 46, 155–171.
- *Srikameswaran, S., & Martin, G. L. (1984). A component analysis of a self-management program for improving work rates of mentally handicapped persons. *The Mental Retardation and Learning Disability Bulletin*, 12, 39–52.
- Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*. doi:10.1901/jaba.1977.10-349.
- Super, D. E. (1983). Assessment in career guidance: toward truly developmental counseling. *The Personnel and Guidance Journal*. doi:10.1111/j-2164-4918.1983.tb0009.x.
- *Todd, T., Reid, G., & Butler-Kisber, L. (2010). Cycling for students with ASD: self-regulation promotes sustained physical activity. *Adapted Physical Activity Quarterly*, 27, 226–41.
- *Tollefson, N., Tracy, D. B., Johnsen, E. P., & Chatman, J. (1986). Teaching learning disabled students goal-implementation skills. *Psychology in the Schools*. doi:10.1002/1520-6807(198604)23:2<194::AID-PITS2310230213>3.0.CO;2-C
- *Trammel, D. L., Schloss, P. J., & Alper, S. (1994). Using self-recording, evaluation, and graphing to increase completion of homework assignments. *Journal of Learning Disabilities*. doi:10.1177/002221949402700202
- VanBergeijk, E., Klin, A., & Volkmar, F. (2008). Supporting more able students on the autism spectrum: college and beyond. *Journal of Autism and Developmental Disabilities*. doi:10.1007/s10803-007-0524-8.
- *Wanlin, C. M., Hrycaiko, D. W., Martin, G. L., & Mahon, M. (1997). The effects of a goal-setting package on the performance of speed skaters. *Journal of Applied Sport Psychology*. doi:10.1080/10413209708406483.
- *Ward, P., Crouch, D. W., & Patrick, C. A. (1998). Effects of peer-mediated accountability on opportunities to respond and correct skill performance by elementary school children in physical education. *Journal of Behavioral Education*. doi:10.1023/a:1022868808642
- *Winn, B. D., Skinner, C. H., Allin, J. D., & Hawkins, J. A. (2004). Practicing school consultants can empirically validate interventions. *Journal of Applied School Psychology*. doi:10.1300/J370v20n02_07

CHAPTER 7: DISCUSSION

7.1 Methodological Implications of Treatment Effect Calculations

The papers presented above have attempted to identify issues and challenges that are of particular concern when evaluating SCD literature for participants with ASD. While additional research with other behavioural interventions employed with participants with ASD is required before these preliminary findings can be confidently generalised across this population, several important points for consideration have been identified.

Most importantly, the first of the studies reported that relatively short data series are being collected, with a declining trend in the volume of data points observed over time. Self-management studies were used to represent an established treatment and physical activity studies to represent an emerging treatment. Drawing from both data sets, more recent studies were found to have included fewer data points than older studies. However, behavioural challenges were described for many participants. Collecting a greater number of data points, particularly in base line conditions poses a significant ethical dilemma to researchers and may often not be in the best interests of the participant, other students, or teachers. Given this current trend it appears that in future, SCD data for participants with ASD may not be suitable for more complex treatment effect calculations should minimum data requirements not meet the required thresholds of these algorithms.

In particular, regression based calculations do not appear suitable for the type of data that is collected for participants on the autism spectrum, making a nonparametric calculation the better of the currently available options. Excluding studies which provide an insufficient volume of data points from meta-analyses may result in distorted findings. Furthermore, omission of studies on this basis arguably does not assist researchers collating information to determine best evidence. Therefore, the second study explored in further detail which of the

available nonparametric calculations may be most appropriate currently for all stakeholders to evaluate best treatment practices for individuals with ASD.

The second study was guided by the statements made by the Taskforce on Statistical Inference (1999), that emphasized the importance of understanding how a given statistical measure is calculated, and further, how to interpret the statistic. The three indices included in Study 2 were selected on the grounds that they can be calculated by hand in a relatively straightforward manner, without requiring extensive training or additional software applications. Arguably, this may mean that teachers, clinicians, or other stakeholders in underfunded communities, or remote location, could access and interpret treatment reports.

Data were extracted from a systematic search of published self-management intervention studies. PND, PAND, and NAP were calculated for all participants for whom sufficient data had been provided, with PND used as the basis for comparison. A significant percentage of these articles contained insufficient data to employ the PAND metric.

Given that the current data collection trend observed in the reviewed studies appears unlikely to change in the near future, short data sets appear typical in intervention research with participants with ASD. Although PAND has received favourable feedback in the literature when compared to PND, and has appeared recently in several published systematic reviews conducted with students with disabilities, PAND appears inappropriate given the short data sets that were typical for participants on the autism spectrum.

While the sensitivity analysis of the three treatment effect scores intended to identify the most appropriate method to calculate treatment effect in the immediate term was the main focus of this study, a secondary issue of great significance was identified as a study outcome. Interpretation of derived scores is currently not a straightforward procedure. At present, an interpretation scale is in wide use for PND, a separate scale has been proposed for NAP, and

an interpretative scale is absent for PAND. While it is mathematically relatively straightforward to calculate a *Phi* or *Phi*² correlation coefficient for PAND scores, the literature review has revealed that interpretation standards currently vary across research fields. Reporting treatment effect in this manner may address the aim to integrate both group and SCD research in meta-analyses, yet additional research is currently required to facilitate meaningful interpretation of either *Phi* or *Phi*² correlation coefficients. Furthermore, additional research is necessary in order to determine the applicability of the PAND metric to data sets for interventions with participants with ASD, as the research with self-management data has shown that studies frequently include fewer than the required 20 data points.

Also of concern, the tentative interpretative scale that has been developed for NAP presents bandings of scores that are inconsistent to those of the widely employed PND metric. Until these issues are further researched, interpretation of newer calculation methods should be treated cautiously. In particular, it appears that newer methods may report inflated treatment effect scores relative to PND. This may be potentially misleading, as readers may perceive studies as more effective should a newer treatment effect score be adopted. This in turn may contribute to false expectations on behalf of treatment providers should they implement a new treatment reported to have a greater treatment effect when compared to older established treatments that have been reported using a more conservative PND score. Consequently, it may appear in the short term that PND is the most appropriate calculation for evaluating research specific to individuals with ASD.

7.2 Developing Independent Functioning

“As Lovitt (1973) observed more than 30 years ago, the fact that systematic instruction of self-management skills is not a part of most schools’ curricula is a paradox because ‘one of

the expressed objectives of the educational system is to create individuals who are self-reliant and independent’ (p. 139)” (Cooper et al., 2007, 583).

Arguably, support services for individuals around the world face immediate challenges in keeping pace with the rapid increase in prevalence of ASD as described earlier. Findings from research reviews of treatments conducted with participants with ASD have been interpreted in different ways by major United States health insurance coverage providers. In a recent report it was noted that coverage for ABA services had been extended to individuals under one insurance carrier, and not another, resulting in funding anomalies (Campbell, 2013). Accordingly, some students may receive support services while others do not, as a result of the insurance plan guidelines.

Providing appropriate supports for individuals with ASD is a big challenge that impacts a tremendous number of children and adults alike. It appears that ASD affects families regardless of ethnicity or social or economic factors, and the condition remains to varying extents throughout the lifetime of affected individuals. As larger numbers of adolescents on the autism spectrum enter higher education the ability to function independently with as little support as possible is of utmost concern.

In its most simple form, self-management may be understood as a behaviour performed by a person that influences another of their behaviours. Cooper and colleagues (2007) have defined self-management within the ABA framework as “the personal application of behavior change tactics that produces a desired change in behavior” (p. 578). When a behaviour modification intervention is implemented by a teacher, clinician or parent, this party may manipulate variables in the environment, arrange stimuli, provide visual or verbal prompts, deliver consequences, and observe and record occurrences or non-occurrences of targeted behaviour over a set time period or in terms of frequency. Self-

management of the behaviour change intervention involves the participant performing some, or all, elements of the program.

Self-management has been applied to a variety of situations to help individuals to become more effective and efficient in daily living, break bad habits and establish good ones, achieve difficult tasks and attain personal goals. Further, self-management has been identified in the processes of generalising and maintaining behaviour change gains to new tasks or settings, and over extended periods of time (Baer et al., 1968).

Managing one's self independently in social, academic, and work environments is frequently reported to be challenging for children and adults with ASD. Given that self-management techniques offer the potential to assist individuals to attain greater independence, the literature was systematically examined to identify what is known about self-management specifically with individuals diagnosed on the autism spectrum. Goal setting was associated with highly effective treatment outcomes in the general self-management literature review, but found to be included relatively rarely in the ASD literature. As goal setting may be considered an important skill for older students and young adults who aspire to independence an additional goal setting literature search was conducted subsequent to the self-management review.

7.3 Findings from Self-Management and Goal-Setting Systematic Reviews

The meta-analysis of self-management interventions adopted the WWC SCD pilot guidelines as a general framework for assessing the quality of the evidence in the research base. However, in a departure from the WWC guidelines, studies that exceeded 20 years since publication date were retained. Further, participants were included regardless of age, a variation to the current NSR (2009) report that capped age of participants at 22 years. The 5-3-20 rule was adopted to synthesise the collection of studies.

The review found adequate evidence to support self-management procedures as a best practice for participants with ASD. More specifically, self-management was described as an effective treatment for students aged three to 25 years old. One important finding of this review was that for three studies, one targeting improvements in social skills, one the development of daily living skills, and one an increase in academic skills, collateral benefits in the form of reductions in problem behaviours, stereotypy and self-injury were described. While earlier studies reported success in clinics, home and/or special education settings, more recent studies were conducted in general education classrooms and playgrounds where self-management interventions were also described as highly effective. In addition, self-management was applied to various aspects of academic skills that included attention to task, and improvements in either the quality or volume of academic work. Self-management interventions were found to be effective with pre-school aged children, and also with adults that were described as either high or low functioning.

Several areas of the self-management topic appear under researched to date. In particular, the majority of studies that targeted academics were conducted with high functioning students. Evidence regarding the inclusion of peers in interventions was relatively sparse, with peers being described in either a facilitator or reinforcing role. Importantly, goal setting appears to be associated with successful treatment outcomes but has not been frequently included in self-management interventions. Goal setting was applied in interventions that targeted improvements in academic writing and was described as an important aspect of self-regulation. Following training and feedback, it was observed that independent goal setting was often observed.

The ability to independently set realistically attainable and appropriate goals is of utmost importance to individuals with autism aiming for independent living, yet appeared under researched in the self-management literature. Consequently, a broader systematic

search of the literature was conducted to better understand what is known on teaching and applying goal setting skills.

This second literature review identified preliminary support for the effectiveness of goal setting techniques in a variety of settings and for a range of target behaviours, although few interventions included participants with ASD. It was noted that goal setting techniques were almost always confounded with other intervention components in a treatment package: two thirds of the studies with self-monitoring; and one half of the studies with feedback.

While no clear pattern on how goals were set was evident from the studies included in the review, the results did suggest that goal setting skills are teachable. Several studies used goal setting to successfully improve sports performance. This may prove a valuable line of research for individuals with ASD by promoting the development of various skills required to engage in a physically active lifestyle and providing opportunities to socialize with typically developing peers. A more physically active lifestyle may also assist in alleviating the challenges associated with a sedentary life style, food selectivity, and disrupted sleep that are frequently described in the ASD literature.

Goal setting skills appear to be developed over time. However, maintenance probes suggested that treatment gain was not always maintained over time. As such, it may be crucial to program interventions to promote generalization of goal setting skills across behaviours and settings, and to improve the likelihood of skill maintenance over time.

7.4 Summary

Individuals on the autism spectrum may have unique profiles that often include special talents or unique strengths in addition to social, sensory or behavioural challenges. Many highly desirable personal attributes are often associated with ASD, such as an ability to concentrate deeply on specific tasks, a strong sense of fairness, and honesty. Geller and

Greenberg (2010) noted that often high functioning individuals on the autism spectrum may be intellectually advanced, passionate about studying, focused on academic or work goals, loyal and hardworking, and may have highly marketable skills. Ideally, behavioural interventions that can assist individuals to overcome barriers to social inclusion and participation, yet allow retention of their unique strengths, are highly desirable.

Intervention techniques that can be taught to very young children may ultimately help alleviate the costs associated with providing support services later in life. However, many higher functioning individuals on the spectrum are not diagnosed until older and miss the opportunity for early intervention. Similarly, children in underfunded communities, or who may be geographically isolated from support services, may be overlooked at a young age. Therefore, intervention techniques that are effective with older learners are also very important. Self-management was identified as an evidence-based intervention technique that may be used across the life-span with individuals regardless of level of cognitive functioning.

Further, the ability to manage independently, and self-advocate in the absence of family or teacher support is a priority to older children or those who fall outside of a support system. In this context teaching effective goal setting looks to be a promising strategy that appears to be teachable to young and older learners alike. The review of goal setting research presented above has indicated that further research that promotes the acquisition of this important skill is warranted.

CHAPTER 8: CONCLUSION

8.1 Recount of Study

Lovass (1987), drawing on the findings of the original early intervention program, reported that the assignment of one fulltime special education teacher for two years would cost an estimated \$40,000, in contrast to the nearly \$2 million it would cost for life-long institutionalisation of one individual. When adjusted for inflation into the current dollar value, the cost of a special education teacher would now equal a little over \$80,000, with costs for life-long institutionalisation exceeding \$4,038,000 (Friedman, 2015). Long-term institutionalisation is neither in the best interests of individuals or societies. Soaring support care costs place a burden on families and support service providers, and are arguably unsustainable. A solution that assists in the promotion of independent functioning is urgently required.

Transition beyond secondary school into adulthood has been described in the literature as particularly challenging for individuals diagnosed on the autism spectrum. Common difficulties for this age group may include: developing or maintaining conversation or relationships; patterns of thinking that may be perseverative; difficulty in understanding non-verbal cues or perspectives of others; visual, auditory or tactile hyper or hypo sensitivity; physical clumsiness or unusual movements; difficulty understanding oral directions; and emotional dysregulation (Geller & Greenberg, 2010). An uneven profile of strengths and deficits may result in underperformance or social withdrawal. Post-secondary faculty, employers, peers and/or colleagues may struggle to understand an apparent mismatch between a student or employee's potential, and their performance, as a result of impairments in organisational or social skills. The literature has highlighted that appropriate support

systems are sparse, and much work is required in order to improve outcomes for this population.

This current study has explored two pertinent issues central to the understanding of how to improve outcomes for youth and adults who are diagnosed with ASD. The research question that guided this research was:

How can we effectively develop independent functioning of children and adults with Autism Spectrum Disorder (ASD)?

The first core area of research involved developing an understanding of how to assess the literature in order to identify an evidence base of best practice that may be interpreted by various stakeholders including clinicians, teachers, parents and researchers. This entailed identifying the most current taskforce protocols in the literature regarding the conduct of quality assessment of interventions, and performing meta-analyses in SCD research. A review of this literature highlighted the ongoing debate that surrounds the most appropriate method to determine the strength of treatment effects. A series of two studies was undertaken to better understand the nature of SCD data that are collected in interventions conducted with participants of all ages, who are diagnosed with ASD, and compare outcomes from different methods that were considered most suitable to calculate treatment effect scores. In summary, the findings suggested that PND remains a widely applicable calculation method that produces a comparatively conservative treatment effect score.

The second core area of research involved exploring intervention components that may teach the requisite skills to function independently in settings that offer little or no support to high functioning transition age youth or young adults with ASD. In particular, self-management techniques that were used in interventions that sought to develop skills or increase performance of a targeted behaviour formed the focus of the review. Goal setting

skills were identified as largely absent from this self-management research base, and a further systematic review that extended into the broader population was conducted. PND was applied in the self-management review to studies that met the WWC SCD quality assessment guidelines. Accordingly, self-management was found to be an effective evidence-based treatment. Further, goal setting was identified as a teachable skill that may prove beneficial to those with ASD.

8.2 Significance of Findings

The first study identified that shorter data sets are being included in recent studies when compared to older studies. More complex regression based treatment effect algorithms have been argued by some as a favourable calculation approach. However, the volume of data points that has been presented in this data set was found to be insufficient for use with such methods.

The second study compared PND with two newer nonparametric formulae, PAND and NAP, both described in the literature as promising improvements. The findings of this study were significant in three important ways. PAND was found to be unsuitable for a significant number of studies that did not include a high enough volume of data points. Firstly, when identifying evidence-based practice, omitting studies is problematic. This may result in an insufficient volume of studies to satisfy the 5-3-20 guideline, or may misrepresent the results of meta-analysis. Secondly, PND, while often rejected in critiques favouring newer nonparametric calculations, was found to be applicable for the majority of studies, and to report a conservative strength of treatment effect compared to both PAND and NAP. Thirdly, while a primary focus of ongoing debate in the literature is centred on which treatment calculation method to adopt in SCD research, interpretation of derived scores has emerged as an equally important issue.

The third study, an examination of the self-management literature, revealed that this technique can be successfully taught to older students and young adults and used in settings in which the participants are required to function independently. Support systems for many older learners are scarce, and limited funding may present challenges to stakeholders when designing and implementing treatment packages. This meta-analysis has identified that a reduction in problem behaviour may not need to be the target behaviour of an intervention in order to achieve improved outcomes. Rather, a reduction in such behaviours was identified as a collateral effect of interventions that focused on skill development or improvement in performance of an already acquired skill. This review also identified that the inclusion of peers in either a facilitator or feedback/support role was associated with effective treatments, however this appears to be an under researched phenomenon.

The ability to independently set appropriate goals is an essential skill for successful transition beyond secondary education for youth and young adults with ASD. Given the paucity of ASD specific research, the systematic review of goal setting intervention literature was extended to the general population. Findings of the review have been focused on their significance to individuals on the autism spectrum. Firstly, this review has indicated that goal setting skills are teachable. Secondly, it also identified that goal setting is often combined with self-monitoring. However, the latter was not apparent in the earlier review of self-management, and it may be argued that this important component of treatment packages has been overlooked by researchers working with individuals with ASD.

8.3 Implications for Practice

Several important implications have emerged as an outcome of this study. The first core component of the study examined various methods of calculating treatment effect scores. A significant implication to the variety of stakeholders who may access this

information is that extreme care should be exercised when interpreting derived scores. While older studies have typically measured strength of treatment effect using PND, newer algorithms may yield a greater strength of treatment effect when applied to the existing literature, or to newly published studies. This information is potentially misleading as stakeholders may unintentionally anticipate a greater result from treatment packages.

It has been noted in a recent study that friendships play an important role in mental health, and that accordingly it may be reasonable to assume that assisting individuals on the autism spectrum to better develop and maintain friendships may lead to improvements in the core deficits, and comorbid symptoms, of ASD (Miller et al., 2014). The review of self-management literature has identified that inclusion of peers as facilitators or reinforcers has proven effective, yet has been employed relatively infrequently. Research with teachers or parents that includes peers in these capacities is encouraged, to support positive change in classrooms and other settings.

An explicit recommendation of the EBP taskforce was that goals should be developed collaboratively between the patient and the clinician (Goodheart et al., 2006). Arguably, in order to meet this objective, an individual on the autism spectrum requires an understanding of goal development and goal attainment. Furthermore, many individuals face situations in which they do not have access to a clinician for support, or may reach an age at which such support services are discontinued. For these individuals, it is imperative to develop the ability to independently set realistically attainable goals. The review of goal setting literature has identified that these skills are teachable. However, both the self-management meta-analysis and the goal setting systematic review have highlighted the issue that this essential skill appears to have been overlooked to date. Ongoing research appears justified to support changes in practice.

8.4 Limitations of the Study

While this research has made a contribution to the literature through a series of publications, choices were necessary throughout the research journey resulting in limitations to the study.

The ability to generalise the findings from the research conducted for both Papers 1 and 2 may be considered a limitation of this study. While many intervention types are used for individuals with ASD, the overall research focus for this thesis was on increasing independent functioning. It was hypothesised that self-management may promote independent functioning for this population, and was selected as the intervention type to explore in the research of issues that surround the calculation of a treatment effect score.

Initially the psycINFO data base was queried for self-management interventions with a resulting data set that covered the time period from 1990 through 2011. Data were drawn from 38 published articles and provided 215 data series that reported on results for 102 participants. The data were plotted in a line graph over time in excel, and the line of best fit function was used to identify trend in the data. Arguably drawing data from only one intervention type may impact the ability to generalize findings to other intervention types. In turn, this may negatively impact the ability to effectively answer the research questions of whether studies report a sufficient volume of data points to conduct a treatment effect calculation and whether any trends are apparent in the data.

Upon reflection, the data set that was used to identify any apparent trends in data collection was expanded to also include exercise interventions. It was theorised that many benefits associated with leading a more physically active life style may also contribute positively to improving independence amongst individuals with ASD. PsychINFO was queried again for exercise interventions, and additional data drawn from eight studies that

provided 43 data series for 20 participants were included in the revised analysis. In addition, a split middle line of progress was plotted in addition to the excel line of best fit. The trend lines closely mirrored each other, although the y-intersect value varied slightly.

The sensitivity analysis conducted for Paper 2 was based solely on the self-management intervention data given that this overall research project was subject to the time constraints of PhD candidature. Parker and colleagues (2007) described the PAND data point minimum threshold in the original literature as between 20-25 data points. A conservative minimum threshold of 20 data points was selected for this research in an attempt to include as many studies as possible. Accordingly a PAND score was calculated for 22 studies and reflected 57 participants. It may be argued that this decision has improved confidence in the findings of the comparison between the three treatment effect scores, as it has been based on the largest sample possible for this data set. Should a minimum of 25 data points have been adopted it is likely that a PAND score would be calculated for less than these 22 studies. Future research may overcome the limitation of Paper 2 by adopting the greater minimum threshold of 25 data points, and replicating this sensitivity analysis with other intervention types that are frequently used for participants with ASD.

The third paper in this research program used PND to calculate treatment effect scores in a meta-analysis of self-management interventions. PND was selected on merit as a result of the feasibility study, and subsequent sensitivity analysis of treatment effect scores conducted in Papers 1 and 2 respectively. A particular weakness of this method is the inability to calculate a confidence interval for these scores, as there is reportedly no known distribution for this non-parametric approach. The absence of confidence intervals on mean treatment effect scores reported in the meta-analysis may be considered a limitation of the study, and this may impact the reliability of these scores. Further, self-management interventions were only included in this study if the target behaviour had been either an

increase in performance or the acquisition of a new skill. Studies that targeted a reduction in challenging behaviours using self-management techniques were omitted from the meta-analysis. The rationale behind this choice was that this overall research project was focused on improving independent functioning. Future research may address the limitation of absence of confidence intervals by drawing from the larger sample of self-management interventions that may also include studies that have targeted a reduction in problem behaviours.

Finally, the fourth paper examined goal setting interventions and drew from research conducted with the broader population in addition to that with ASD. As only two studies included participants with ASD, this phase of the research is arguably of limited value when generalizing to the specific ASD population. Accordingly, a frame work of proposed significant issues was developed for future researchers working with individuals with ASD.

8.5 Recommendations for Future Research

The NSR (2009) has identified 11 established treatments, one of which was self-management, and 21 emerging treatments, one of which was exercise. Additional research that examines the nature of data gathered for participants on the autism spectrum that examines additional treatments is highly warranted to be able to generalise the findings of this aspect beyond what was identified with self-management and exercise interventions. Should additional research on the volume of data points gathered in SCD studies for participants with ASD support the initial findings of this current research, treatment effect calculation considerations for this sub-community may be accounted for in the recommendations developed at for the broader field of educational psychology.

Arguably, the ability to independently set appropriate goals is an essential skill for children and adults on the autism spectrum. This line of research has been identified as a

teachable skill yet appears largely under-researched with this population. Ongoing research that explores ways in which goal setting skills can be effectively taught appears highly warranted.

The findings from the meta-analysis have identified self-management based interventions as an evidence-based practice for this population. While this technique has been applied to social skill development, it has been noted that physical activity has been largely overlooked in the studies included in the review. Ongoing research incorporating self-management techniques may help children and adults on the autism spectrum to develop the skills required to engage in a physically active lifestyle. In turn, this may assist in expanding opportunities to engage socially with neurologically typically developing peers. It is plausible that overcoming these barriers to participation may contribute positively to regulation of sleeping and eating patterns, and potentially assist in alleviating co-morbid depression. Pursuing this line of research appears justified.

8.6 Contribution to Knowledge

This study has made a contribution to the literature via a series of four published papers. Each paper has presented the research methods and subsequent analysis in a format that will allow other researchers in the field to replicate the processes that were adopted. The papers have each been subjected to the rigour of peer review prior to publication in notable US academic journals of specific interest to stakeholders in the autism field. A contribution to knowledge has been established by building upon current research issues, and through the analysis and interpretation of findings from the data collated for this thesis.

The first paper has identified that short data sets appear to be the norm for SCD research working with participants with ASD. In particular, graphs of this data revealed a declining trend in length of data sets that are being collected, with older studies presenting a

greater volume of data points than more recent studies. However, examination of the participant behaviour descriptions revealed that for many participants involved in interventions for which the target behaviour was development or improvement of a skill, challenging behaviour was also an issue. Such behavioural issues may present challenges to researchers, peers, or the participant and the collection of additional data presents a significant ethical dilemma. The likelihood of collecting greater volumes of data points in future interventions appears improbable in many instances. This finding is of particular importance to the current debate in the literature in which the merit of regression based treatment effect calculations has not been dismissed. The research conducted from this current paper has suggested that a non-parametric based treatment effect calculation is preferable for the ASD sub community of SCD researchers and regression calculations appear not suitable.

The second paper compared the widely adopted PND treatment effect score to two newer non-parametric calculations that can be performed by hand, PAND and NAP, both identified in recent literature as improving upon PND. However, the research conducted for this sensitivity analysis surprisingly found that while PAND has gained recent popularity in published educational psychology literature, this calculation appears the least preferable for the ASD sub-community of researchers working with SCD data. A significant number of studies reported an insufficient volume of data for the PAND calculation to be applied. Further, while the broader literature noted that a *phi* or *phi*² correlation coefficient can be calculated, this study identified that meaningful interpretation guidelines are currently absent from the literature for the researchers working in the autism field.

Findings from the current examination of the NAP calculation with self-management intervention data that is specific to participants with ASD have also made a timely contribution to the ongoing debate in the field. Specifically, unlike either PND or PAND, the

NAP hand calculation was applicable for all data series. The apparent short nature of the self-management data for participants with ASD resulted in a relatively straight forward hand calculation. Another important contribution to the ongoing debate made by the findings of this current sensitivity analysis was that the tentative interpretation scale currently proposed by the original authors of NAP uses a four banding scale that varies from that of the widely adopted PND. In the concluding comments of the published sensitivity analysis, stakeholders reviewing research that has employed NAP have been cautioned to be aware of potentially misinterpreting treatment effect scores and descriptions of strength of treatment effect when comparing to earlier research that has employed PND scores. In particular, care is urged such that newer research reported using NAP is not unintentionally perceived as resulting in a greater strength of treatment effect. Surprisingly, in contrast to widely accepted critique, PND was found to be applicable to the majority of data that was included in the sensitivity analysis. In light of this finding, and in consideration of the potential misinformation in the absence of an agreed upon interpretation scale, current continued use of the PND method was argued.

The meta-analysis on self-management has also made a contribution to the field, as this current study was the first to apply quality assessment guidelines to the body of self-management interventions conducted with participants on the autism spectrum. The findings of the meta-analysis that focused on skill development or acquisition of a new skill as the target behaviour, suggested that there is sufficient evidence to conclude that self-management is an evidence-based treatment for this population. Unlike earlier systematic reviews in this specific area, this current meta-analysis reported that intervention outcomes are successful not only with very young children but also with older students and young adults. This meta-analysis has also identified recent extension of self-management techniques to the development of various academic tasks. The meta-analysis excluded from the data set any

studies in which a reduction in problem or challenging behaviours served as the target behaviour. Surprisingly, the study identified that a reduction in these undesired behaviours was described as a collateral benefit of almost one third of the studies that focused on skill acquisition or improvement.

Finally, while goal setting has been widely studied in the broader literature, the self-management study and subsequent systematic review of goal setting interventions conducted in Paper 4 has ironically revealed that these techniques are largely overlooked in interventions conducted with participants with ASD. This systematic review has identified that goal setting components are almost always confounded with other intervention components. The most common was self-monitoring, which was included in two thirds of the studies, and feed-back which was included in half of the studies. This finding is consistent with an earlier claim made by Southall and Gast (2011) in their self-management literature review, in which those authors argued the merit of further research into the specific components of self-management interventions to identify the relative contribution of each. The goal setting review conducted for this thesis has contributed to this argument by adding further evidence that components appear to be confounded.

The goal setting review, while noting that goal setting skills appear teachable, observed no clear differential pattern is evident between studies in which a researcher or participants set goals. A proposed research agenda specific to the ASD population was suggested in Paper 4. The research agenda has made a timely contribution to the field by identifying a paucity of research that has used goal setting techniques to assist individuals with ASD to develop a physically active lifestyle. In looking forward, this review has also suggested that future researchers program interventions to optimize generalization and maintenance of treatment effects, as the information to date gathered in the review reported variable findings in these areas.

REFERENCE LIST

- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*, 91–7.
doi:10.1901/jaba.1968.1-91
- Baker, J. P. (2013). Autism at 70 - Redrawing the boundaries. *The New England Journal of Medicine, 369*, 1089–1091. doi:10.1056/NEJMp1306457
- Birnbrauer, J. S., & Leach, D. J. (1993). The Murdoch early intervention program after 2 years. *Behaviour Change, 10*, 63–74.
- Blampied, N. M. (1999). A legacy neglected : Restating the case for single-case research in cognitive-behaviour therapy. *Behaviour Change, 16*, 89–104.
doi:10.1375/bech.16.2.89
- Boyd, B. A., Odom, S. L., Humphreys, B. P., & Sam, A. M. (2010). Infants and toddlers with autism spectrum disorder: Early identification and early intervention. *Journal of Early Intervention, 32*, 75-98. doi:10.1177/1053815110362690
- Brossart, D. F., Parker, R. I., Olson, E. A., & Mahadevan, L. (2006). The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behavior Modification, 30*, 531–63. doi:10.1177/0145445503261167
- Bryson, S. E., Zwaigenbaum, L., McDermott, C., Rombough, V., & Brian, J. (2008). The autism observation scale for infants: Scale development and reliability data. *Journal of Autism and Developmental Disorders, 38* 731-738. doi:10.1007/s10803-007-0440-y

Campbell, J. M. (2013). Commentary on PND at 25. *Remedial and Special Education*, 34, 20–25. doi:10.1177/0741932512454725

Carr, M. E., Anderson, A., Moore, D. W., & Evans, W. H. (2014). How should we determine treatment effectiveness with single-case design research for participants with autism spectrum disorder? *Review Journal of Autism and Developmental Disorders*. doi:10.1001/s40489-014-0030-9

Carr, M. E. (2014). A sensitivity analysis of three nonparametric treatment effect scores for single-case research for participants with autism. *Review Journal of Autism and Developmental Disorders*. doi:10.1007/s40489-014-0037-2

Carr, M. E., Moore, D. W., & Anderson, A. (2014). Self-management interventions on students with autism: A meta-analysis of single-subject research. *Exceptional Children*. doi:10.1177/0014402914532235

Carr, M. E., Moore, D. W., & Anderson, A. (2014). Goal setting interventions: Implications for participants on the autism spectrum. *Review Journal of Autism and Developmental Disorders*. doi:10.1007/s40489-014-0022-9

CDC MMWR. (2012). Prevalence of autism spectrum disorders - autism and developmental disabilities monitoring network, 14 sites, United States, 2008. *Centers for Disease Control and Prevention*, 61(3), 1–19. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/22456193> July 7, 2014

CDC MMWR. (2014). Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *Morbidity and mortality weekly report. Surveillance summaries (Washington, D.C. :*

2002), 63(2), 1–21. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/24670961>

July 7, 2014

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66, 7–18. doi:10.1037/0022-006X.66.1.7

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied Behavior Analysis* (2nd ed.). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

Crosland, K. A., Clarke, S., & Dunlap, G. (2012). A trend analysis of participant and setting characteristics in autism intervention research. *Focus on Autism and Other Developmental Disabilities*, 28, 159–165. doi:10.1177/1088357612468029

FDA. (2014). Beware of false or misleading claims for treating autism. *Consumer Health Information*, 1-2. Retrieved from <http://www.fda.gov/downloads/forconsumers/consumerupdates/ucm394800.pdf>
December 1, 2014

Fombonne, E. (2005). The Changing Epidemiology of Autism. *Journal of Applied Research in Intellectual Disabilities*, 18, 281–294. doi:10.1111/j.1468-3148.2005.00266.x

Foxx, R. M. (2008). Applied behavior analysis treatment of autism: The state of the art. *Child And Adolescent Psychiatric Clinics Of North America*, 17, 821–834.
doi:10.1016/j.chc.2008.06.007

Friedman, M. (2015). *The inflation calculator*. Retrieved from <http://www.westegg.com/inflation.cgi> January 26, 2015

- Geller, L., & Greenberg, M. (2010). Managing the transition process from high school to college and beyond: Challenges for individuals, families, and society. *Social Work in Mental Health*, 8(1), 92–116. doi:10.1080/15332980902932466
- Gillberg, C. (1990). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 31, 99–119. doi:10.1111/j.1469-7610.1990.tb02275.x
- Goodheart, C. D., Levant, R. F., Barlow, D. H., Carter, J., Davidson, K. W., Hagglund, K. J., ... Zane, N. W. (2006). Evidence-based practice in psychology. *The American Psychologist*, 61, 271–85. doi:10.1037/0003-066X.61.4.271
- Green, V. A., Pituch, K., Itchon, J., Choi, A., O'Reilly, M., & Sigafoos, J. (2006). Internet survey of treatments used by parents of children with autism. *Research in Developmental Disabilities*, 27, 70–84. doi:10.1016/j.ridd.2004.12.002
- Honda, H., Sgunuzu, Y., Misumi, K., Niimi, M., & Ohashi, Y. (1996). Cumulative incidence and prevalence of childhood autism in children in Japan. *British Journal of Psychiatry*, 169, 228–235.
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S. L., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71, 165–179. doi:10.1177/001440290507100203
- 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. doi:10.1007/s10864-012-9152-2

- 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, 269–290. doi:10.1353/etc.2012.0011
- Howlin, P. (2008). Redressing the balance in autism research. *Nature Clinical Practice Neurology*, 4, 407. doi:10.1038/ncpneuro0860
- Hubbard, R., Parsa, R. A., & Luthy, M. R. (1997). The spread of statistical significance testing in psychology. *Theory and Psychology*, 7, 545–554.
- Kanner, L. (1943). Autistic disturbances of affective contact. *Pathology*.
- Kielinen, M., & Moilanen, I. (2000). Autism in Northern Finland. *European Journal of Child and Adolescent Psychiatry*, 167, 162–167.
- Kim, Y. S., Leventhal, B. L., Koh, Y., Fombonne, E., Laska, E., Lim, E., ... Grinker, R. R. (2011). Prevalence of autism spectrum disorders in a total population sample. *American Journal of Psychiatry*, 168, 904–913.
- Koyama, T., Tachimori, H., Osada, H., Takeda, T., & Kurita, H. (2007). Cognitive and symptom profiles in Asperger's syndrome and high-functioning autism. *Psychiatry and Clinical Neurosciences*, 61, 99–104. doi:10.1111/j.1440-1819.2007.01617.x
- Kratochwill, T. R. (2007). Preparing psychologists for evidence-based school practice: Lessons learned and challenges ahead. *The American Psychologist*, 62, 826–43. doi:10.1037/0003-066X.62.8.829

- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). WWC Single-case designs technical documentation Version 1.0. Retrieved July 12, 2011 http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2012). What works clearinghouse procedures and standards handbook (Version 2.1). Retrieved August 14, 2012 http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf
- Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). *What Works Clearinghouse Procedures and Standards Handbook (Version 3.0)*. Retrieved June, 13, 2013 <http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=19>
- Lambert, M. C., Cartledge, G., Heward, W. L., & Lo, Y. (2006). Effects of response cards on disruptive behavior and academic responding during math lessons by fourth-grade urban students. *Journal of Positive Behavior Interventions*, 8(2), 88–99.
- Leland Wilkinson and the Taskforce on Statistical Inference. (1999). Statistical methods in psychology journals; Guidelines and explanations. *American Psychologist*, 54, 594–604. doi:10.1037/0003-066X.54.8.594
- Loftus, G. R. (1991). On the tyranny of hypothesis testing in the social sciences. *Contemporary Psychology*, 36, 102–105.

- Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*, 3–9.
- Lovaas, O., Koegel, R., Simmons, J., & Stevens Long, J. (1973). Some generalization and follow-up measures on autistic children in behavior therapy. *Journal of Applied Behavior Analysis, 6*, 131–166. doi:10.1901/jaba.1973.6-131
- Luyster, R., Guthrie, W., Gotham, K., Risi, S., DiLavore, P., Lord, C. (2008). *The autism diagnostic observation schedule - Toddler module: Preliminary findings using a modified version of the ADOS*. Retrieved from <http://infar.confex.com/infar/2008/techprogram/P1295.HTML>. February 12, 2015
- Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., ... Lord, C. (2009). The autism diagnostic observation schedule - toddler module: A new module of a standardized diagnostic measure for autism spectrum disorder. *Journal of Autism and Developmental Disorders, 39*, 1305-1320. doi:10.1007/s10803-009-0746-z
- Maggin, D. M., O’Keeffe, B. V., & Johnson, A. H. (2011). A quantitative synthesis of methodology in the meta-analysis of single-subject research for students with disabilities: 1985–2009. *Exceptionality, 19*(2), 109–135. doi:10.1080/09362835.2011.565725
- Mayes, R., & Horwitz, A. V. (2005). DSM-III and the revolution in the classification of mental illness. *Journal of the History of the Behavioral Sciences, 41*, 249–67. doi:10.1002/jhbs.20103

- Miller, A., Vernon, T., Wu, V., & Russo, K. (2014). Social skill group interventions for adolescents with autism spectrum disorders: A Systematic Review. *Review Journal of Autism and Developmental Disorders*. doi:10.1007/s40489-014-0017-6
- Morris, E. K. (1988). Contextualism: The world view of behavior analysis. *Journal of Experimental Child Psychology*, 46, 289–323.
- Murphy, R. J., & Bryan, A. J. (1980). Multiple-baseline and multiple-probe designs: Practical alternatives for special education assessment and evaluation. *The Journal of Special Education*, 14, 325–335. doi:10.1177/002246698001400306
- National Research Council. (2001). *Educating Children with Autism*. Committee on Educational Interventions for Children with Autism. Lord, C. & McGee, J. P., eds. Division of Behavioral and Social Sciences and Education. Washington DC: National Academy Press.
- National Standards Report. (2009). *National Autism Center*. National Autism Center, Randolph MA.
- Parker, R. I., Brossart, D. F., Vannest, K. J., Long, J. R., De-Alba, R. G., Baugh, F. G., & Sullivan, J. R. (2005). Effect sizes in single case research: How large is large? *School Psychology Quarterly*, 34, 116–132.
- Parker, R. I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND): An alternative to PND. *The Journal of Special Education*, 40, 194-204. doi:10.1177/00224669070400040101
- Parker, Richard I, Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children*, 75, 135–150.

- Pinder-Amaker, S. (2014). Identifying the unmet needs of college students on the autism spectrum. *Harvard Review of Psychiatry*, 22, 125–37.
doi:10.1097/HRP.0000000000000032
- Roth, M. E., Gillis, J. M., & DiGennaro Reed, F. D. (2014). A meta-analysis of behavioral interventions for adolescents and adults with autism spectrum disorders. *Journal of Behavioral Education*, 23, 258–286. doi:10.1007/s10864-013-9189-x
- Satterfield, J. M., Spring, B., Brownson, R. C., Mullen, E. J., Newhouse, R. P., Walker, B. B., & Whitlock, E. (2009). Toward a transdisciplinary model of evidence-based practice. *The Milbank Quarterly*, 87, 368–390.
- Schneider, S. M., & Morris, E. K. (1987). A history of the term radical behaviorism: From Watson to Skinner. *The Behavior Analyst*, 10, 27–39.
- Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education*, 8(2), 24–33. doi:10.1177/074193258700800206
- Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: Issues and applications. *Behavior Modification*, 22, 221–242. doi:10.1177/01454455980223001
- Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25: Past, present, and future trends in summarizing single-subject research. *Remedial and Special Education*, 34, 9–19.
doi:10.1177/0741932512440730
- Shadish, W. R. (2014). Analysis and meta-analysis of single-case designs: An introduction. *Journal of School Psychology*, 52, 109–22. doi:10.1016/j.jsp.2013.11.009

Sigafoos, J., & Schlosser, R. (2008). Applied behavior analysis is NOT an autism therapy.

Evidence-Based Communication Assessment and Intervention, 2, 197–198.

doi:10.1080/17489530802640391

Skinner, B. (1974). Behavior Modification. *Science, New Series*, 185, 813.

doi:10.1126/science.185.4154.813

Southall, C. M., & Gast, D. L. (2011). Self-management procedures: A comparison across the autism spectrum. *Education and Training in Autism and Developmental Disabilities*, 46, 155–171.

VanBergeijk, E., Klin, A., & Volkmar, F. (2008). Supporting more able students on the autism spectrum: College and beyond. *Journal of Autism and Developmental Disorders*, 38, 1359–70. doi:10.1007/s10803-007-0524-8