

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
e-mail: mebar4@student.monash.edu

A. Anderson
e-mail: Angelika.Anderson@monash.edu

D. W. Moore
e-mail: Dennis.Moore@monash.edu

W. H. Evans
School of Education, University of West Florida, Building 85,
Pensacola, FL 32514, USA
e-mail: wevans@uwf.edu

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

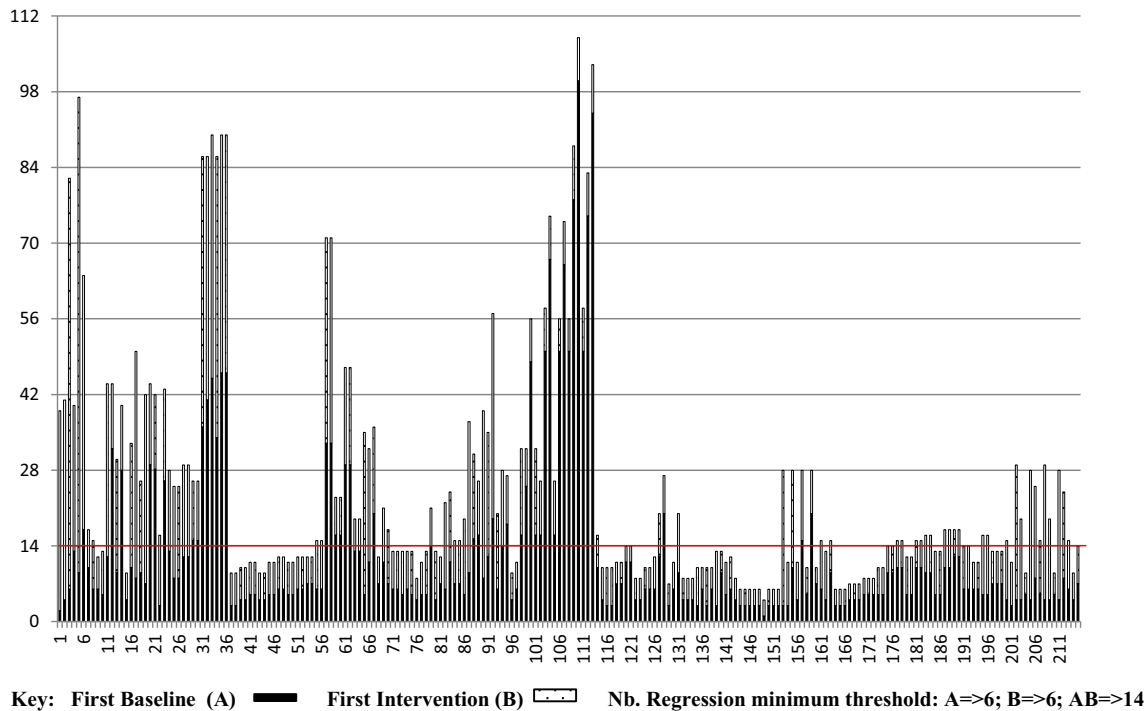
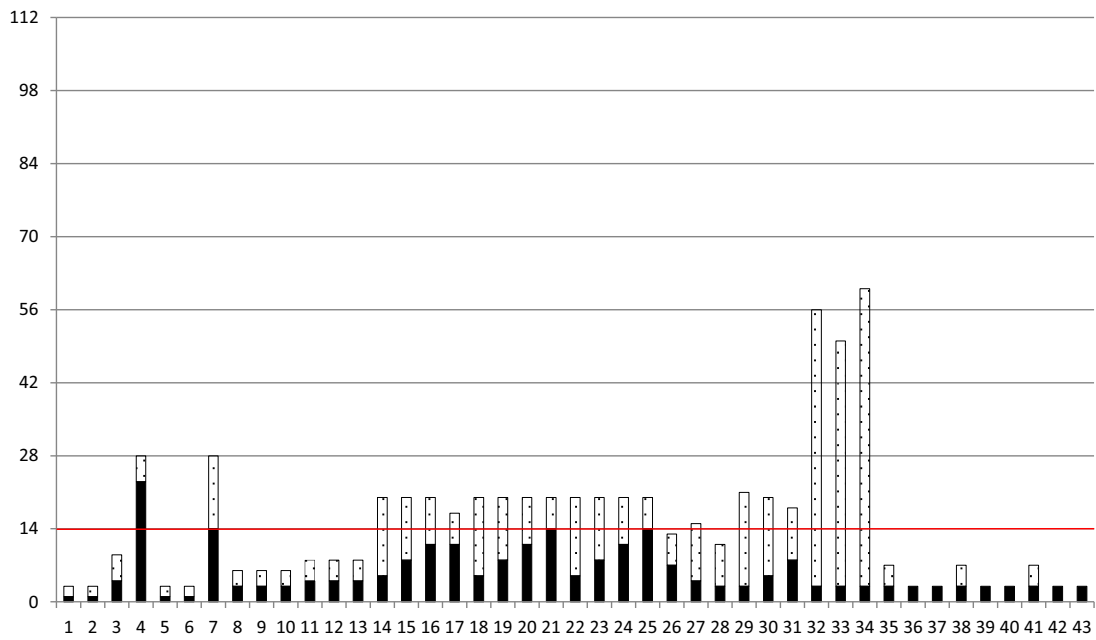


Fig. 1 Self-management data points



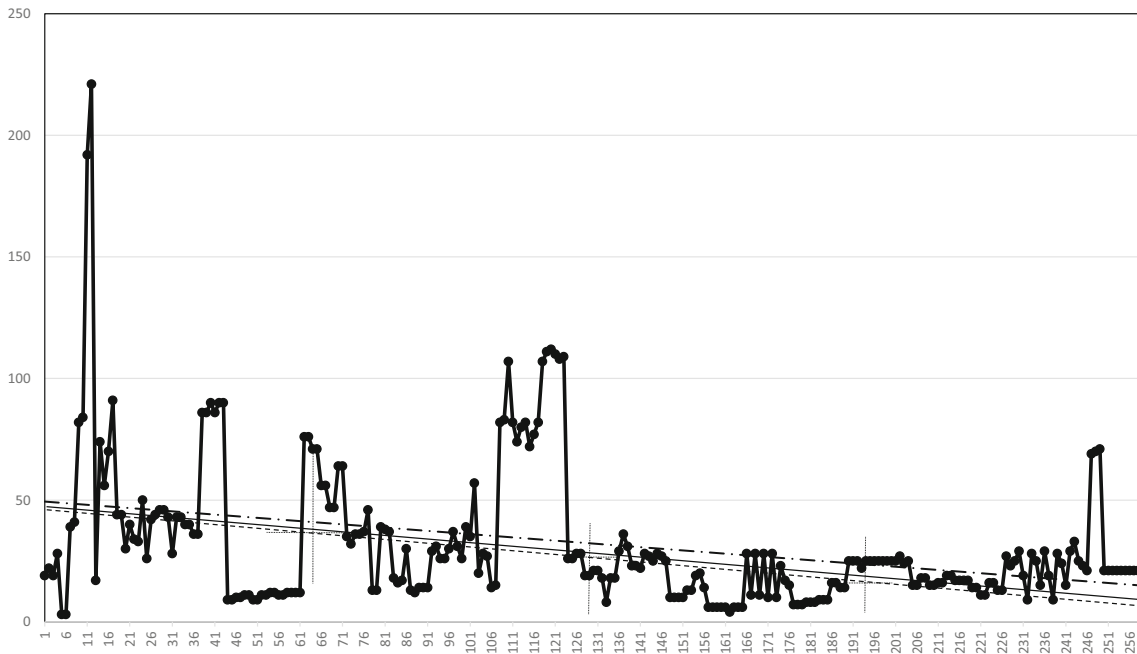
Key: First Baseline (A) First Intervention (B) Nb. Regression minimum threshold: A=>6; B=>6; AB=>14

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



Key: Quarter-Intersect Line of Progress Split-Middle Line of Progress Line of Best Fit $y = -0.1339x + 49.558$

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., 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*. 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.
- *Loftin, 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