CrossMark

REVIEW PAPER

Self-Monitoring Interventions for Students with ASD: a Meta-analysis of School-Based Research

John L. Davis[1] · Benjamin A. Mason[2] · Heather S. Davis[3] · Rose A. Mason[2] ·
Stephen A. Crutchfield[4]

Abstract The current study examines qualities of self-monitoring (SM) interventions for students with autism spectrum disorders (ASD) within school settings. Despite substantial research supporting the use of SM in schools, there is currently a lack of consensus regarding the student-, intervention-, and setting-level variables that support the use of this intervention. The current study specifically examines characteristics of students with ASD that react favorably to the intervention, what procedures are necessary for implementation, and how settings within schools affect outcomes. Sixteen single subject studies were identified with 28 unique participants. Results indicate strong effects for students with ASD (overall Tau-U $= .83$ CI$_{95}$ [.78, .78]) across a variety of behavioral targets and school settings. In terms of implementation, the current analysis found that much of this interventions effects are based four key components (e.g., selecting a target behavior, defining the target behavior, self-assessment, and self-recording). In addition, preliminary evidence shows that higher levels of student involvement lead to stronger intervention effects.

✉ John L. Davis
lockettdavis@gmail.com

1 Department of Educational Psychology, University of Utah, 1721 Campus Center Drive, SAEC 3229, Salt Lake City, UT 84112, USA

2 Juniper Gardens Children's Project, University of Kansas, 444 Minnesota Avenue, Kansas City, KS 66101, USA

3 HOME Program, University Neuropsychiatric Institute, 650 Komas Drive, Salt Lake City, UT 84108, USA

4 School of Education, California Polytechnic State University San Luis Obispo, Cotchett Education Building 2, Room 120A, San Luis Obispo, CA 93407, USA

Autism spectrum disorder (ASD) is a developmental disorder clinically defined by the presence of persistent deficits in social communication and social interaction across multiple contexts and restricted, repetitive patterns of behavior, interests, or activities (DSM-V, 2013). The diagnosed incidence of autism spectrum disorders has increased dramatically within the past decade, occurs in all ethnic and social groups, and is four times more likely to be diagnosed in boys than girls (Rice et al. 2007). Drawing from 2008 data provided by 14 states and published by the Center for Disease Control (CDC), prevalence of autism spectrum disorder is 1 in 88 children (Baio 2012). This represents a striking increase in diagnosed prevalence from the prior CDC estimate of 1 in 156 collected in 2002. The dramatic growth in autism diagnosis underscores the importance of early, applied intervention research with this population.

While non-school settings are often the site of ASD diagnosis, public school settings are typically the initial referral source for surveillance of symptoms and need (Wiggins et al. 2006). Moreover, schools represent the bulk of intervention opportunity, as children and adolescents with ASD spend a majority of their day in educational settings. In the past, exclusionary practices resulted in students with ASD receiving the majority of instruction in self-contained or alternative school placements (Simpson et al. 2003). Consequently, federal mandates under the Individuals with Disabilities Education Improvement Act (IDEIA 2004) and the No Child Left Behind Act (NCLB 2001) require that educators both consider the least restrictive environment as well as provide students access to grade-appropriate curricula to the maximum extent possible. This legislative drive has resulted in an

increased likelihood that students with disabilities will be educated in general education settings alongside typically developing peers. Using federal data drawn from state reports, McLeskey and colleagues (2012) found that over 65 % of students with disabilities were educated in general education settings in 2007, an increase from under 34 % in 1990. Thus, as autism prevalence has expanded, service provision trends have increasingly focused on general education as the preferred placement. Furthermore, the core deficits common to autism often contribute to difficulties in school settings. These difficulties include the following: managing the executive tasks required for planning and carrying out behaviors necessary for goal attainment (Kleinhans et al. 2005), mastering the social behaviors required to build and maintain friendships, and maintaining the on-task and appropriate behaviors expected in typical classrooms (Lord et al. 2000).

The specific behavioral needs of students with ASD in public school settings often require targeted interventions, particularly when those interventions are implemented in very different educational arrangements. Within the last decade, several initiatives (National Autism Center; National Professional Development Center on Autism Spectrum Disorders, 2010; National Standards Project, 2009; What Works Clearing House, 2009) have moved the ASD field forward by identifying and categorizing practices that are considered to have a strong foundation in scientific investigation. These initiatives have mirrored the nationwide drive towards identification and implementation of evidence-based interventions (EBIs) used within a framework of evidence-based practice (EBP). Identified by a federally funded task force and codified within the *Procedural and Coding Manual for the Review of Evidence-Based Intervention* (Kratochwill and Stoiber 2002), interventions that meet design standards and display sufficient internal and external validity should provide sufficient information about participants and settings to allow a school professional to determine if the intervention should benefit *particular student* within a *particular setting*. Thus, EBP and EBI evaluation paradigms place high importance on matching the intervention to the appropriate setting and targeted behavior (Chambless and Hollon 1998; Chambless and Ollendick 2001). Furthermore, the intervention itself must be specifically defined and described so as to allow for identification of *core components* that drive the intervention results. Critical to bridging the research to practice gap is first, identifying the core components that drive intervention results, and second, identifying the methods necessary to implement those components with fidelity in school settings (Fixsen et al. 2005).

Given this need, contemporary EBP evaluation guidelines for single case research call for several specific areas to be addressed (Horner and Kratochwill 2012). These areas include the following: (1) Operational definition of component procedure(s), (2) Designation of any competency criteria that must be met by individuals implementing the procedure(s), (3)

Designation of the context(s) in which the procedure(s) are appropriate, (4) Designation of the population(s) of individuals who are intended to benefit from the procedure(s), and (5) Designation of the valued outcomes that the procedure(s) are expected to affect. Each of these areas will be examined in the current meta-analysis. For these reasons, interventions that may be carried out in inclusive settings with limited local resources that improve behavioral challenges for students with ASD while providing sufficient information for implementers to match behavioral need to specific intervention are in high demand.

Self-Management Interventions

Self-management represents a broad array of skills and strategies individuals use to assess and regulate their behavior (Cooper et al. 2007; Snider 1987). Mooney et al. (2005) categorized the key self-management strategies as falling within one of five domains—self-monitoring, self-evaluation, self-instruction, goal-setting, and strategy instruction, with self-monitoring serving as the most common component arrangement (see Table 1). In a similar investigation of school-based self-management with more finely grained differentiation of component strategies that included exploration of implementation responsibility (e.g., teacher vs student), Fantuzzo and Polite included 11 component parts (see Table 2). The observation and recording components that comprise self-monitoring were again the most frequently reported intervention components.

Fantuzzo and Polite (1990) further investigated self-management component presence in order to determine if the number of components present within the self-management intervention resulted in stronger results. The authors found that the average component presence was 9.6 of 11 possible components and that 60 % of included studies included all 11 self-management components, but that many were managed by the teacher, researcher, or someone other than the student. Self-management interventions offer a relatively unique advantage in that some components of the intervention are student-driven. Thus exploration of component presence should also include who manages that component. Interventions that allow for greater student independence are more portable and by definition, more efficient uses of school resources.

Extending this work, Briesch and Chafouleas (2009) reviewed the effects of self-management interventions on classroom behavior using the Fantuzzo and Polite framework. The authors again found that the "observe and record" components that define self-monitoring interventions were fundamental to school-based self-management but that total numbers of intervention components had declined overall from the 9.6 total found by Fantuzzo and Polite (1990) to 7.6 of 11 in the 20 years subsequent to their review (1988 to 2008, inclusive). Notably, the reduction in overall components did not

Table 1 Self-management interventions

Intervention	Definition
Self-monitoring	Self-monitoring is defined as a multiple step process where the student observes the occurrence or non-occurrence of the behavior and records features of this behavior (Mace et al. 2001).
Self-evaluation	Self-evaluation is an intervention where a student is involved in the determination and/or evaluation of a performance goal (Mace et al., 2001).
Self-instruction	Self-instruction is an intervention based on student generated instructional statements. In this intervention, the student applies predetermined self-directed statements to guide behavior (Bornstein and Quevillon 1976; Meichenbaum and Goodman 1971).
Self-reinforcement	Self-reinforcement is an intervention where the student chooses and administers external reinforcement when a pre-determined criterion is met (Barling and Patz 1980; Morris and Messer 1978).

Definitions adapted from Mooney Ryan, Uhing, Reid, and Epstein (2005)

appear to result in a decrease in obtained effects, suggesting that more streamlined self-management interventions are possible without compromising interventional efficacy.

Meta-analytic Reviews of SM for Students with ASD

A past review of school-based interventions for behavioral challenges associated with autism spectrum disorders found positive effects for self-management intervention packages but did not use similar effect-size measures and thus could not compare interventions directly (Machalicek et al. 2007). Lee et al. (2007) conducted a meta-analysis of self-management interventions for children identified with ASD. The analysis resulted in generally positive outcomes for self-management interventions with students with autism spectrum disorders. The authors found that self-monitoring, self-reinforcement, and self-management packages all led to increases in appropriate behavior, but no significant differences were found between intervention modalities. The findings in Lee et al. demonstrate the need to evaluate specific components of self-management strategies for relative impact on

desired student outcomes but did not use EBI frameworks to assess article quality and, like Briesch and Chafouleas (2009), were likely limited by the use of percent of non-overlapping data (PND; Scruggs et al. 1987) as the nonoverlap metric. PND relies on one baseline datapoint (the highest) for comparison against all intervention datapoints (Parker et al. 2011a) and thus is overly sensitive to outliers within baseline data. A recent meta-analytic review of SM interventions for children and adults with ASD by Carr et al. (2014) did evaluate articles for inclusion according to EBI guidelines. The more stringent guidelines applied by the authors resulted in 23 articles that supported the use of SM interventions overall but was limited by the use of PND as well as the broad differences in clinic, home, and community settings that lack the specific setting-level data needed by school-based practitioners.

Purpose, Rationale, and Research Questions

The present study employs a single case meta-analysis to evaluate the efficacy of school-based self-monitoring interventions in changing behavior for students with ASD and builds

Table 2 Self-management component definitions

Component	Definition
Selection of target behavior	Selecting and prioritizing Behavioral outcome variable(s)
Definition of target behavior	Creating an operational definition of target behavior
Determination of performance goal	Determining the performance criteria for target behavior
Evaluation of performance goal	Comparing actual performance of the target with stated performance goal
Instructional prompts for target behavior	Delivering prompt(s) to engage in target behavior
Observation of target behavior	Making a judgment as to the presence or absence of the target behavior
Recording	Documenting the occurrence of target behavior
Graphing or charting behavior	Summative Documentation of the student recording, either graphically or with written notes
Selection of primary reinforcer	Choosing the primary reinforce
Administration of primary reinforcer	Dispensing or initiating the dispensation of primary external reinforcers
Administration of secondary reinforcer	Dispensing of tokens or points to be exchanged for primary reinforcement

Definitions adapted from Briesch and Chafouleas (2009) and Fantuzzo and Polite (1990)

upon past reviews in four ways. First, by using only those studies that meet WWC standards for single-case research, more confidence may be held in obtained results at the level of individual study and aggregated results. Second, by including more finely grained setting information, implementers can make clearer conclusions about generalizability to their own setting. Third, the use of a robust, distribution-free measure of single-case effect size (Tau-U) allows for more defensible comparison of effects across single-case studies. Unlike the more commonly employed PND, Tau-U compares all individual baseline data points to all individual intervention data points in the adjacent phase. The result is a score that is more representative of overall intervention results (Parker et al. 2011b) and allows for the creation of confidence intervals for obtained results. Fourth, the coding and subsequent comparison of studies employing self-monitoring with (or without) other self-management components allows for comparison within and between specific intervention packages. As the "observe and record" components found within self-monitoring are present in virtually all school-based self-management interventions for students with ASD, a meta-analysis that compares self-monitoring interventions with and without the inclusion of other self-management components is supported.

Research Questions

1. Do mean Tau-U effect sizes obtained from single-case studies that meet WWC standards support the use of self-monitoring with and without other self-management components as a behavioral intervention for students with ASD in school settings?
2. Do mean Tau-U effect sizes differ by student characteristics (grade, gender), setting characteristics (e.g., self-contained vs inclusive classrooms), or intervention characteristics (number of components, adult vs student responsibility)?

Method

A comprehensive literature review of school-based SM interventions for students with ASD was conducted using standard methods identified by Lipsey and Wilson (2001), including keyword searches from bibliographic databases, the review of identified journal articles on the topic of SM interventions, and review of references within those identified articles. The following procedures were used to locate articles published between January 1, 1960 and December 31, 2014. First, the Education Resource Information Center (ERIC), Academic Search Complete (EBSCO), PsycINFO (Proquest), and Cambridge Scientific Abstracts Databases were searched for English-language, peer-reviewed journal articles using the following

keywords: *self-monitoring*, *self-instruction*, *self-recording*, *self-evaluation*, *self-management*, *self-reinforcement*, *self-observation*, and *self-graphing*. Due to a large overlap of SM studies in disciplines outside of education, search strings were generated by combining keywords with Boolean operators and at least one of the following: *special education*, *education*, *classroom intervention*, *school*, and *teacher*.

Following this initial literature search process, a pool of 6592 possible studies were located. After reviewing the title and abstract of each of the identified articles, the number of relevant studies was reduced to 226. Studies were then carefully assessed and included in the current analysis if the following criteria were met:

1. Utilized SCR methodology with a clearly readable graph of data. Group studies were omitted to allow for continuity in comparison of effect sizes (Lipsey and Wilson 2001).
2. Employed self-monitoring as the intervention to modify a student behavior. Studies that examined specific academic skill attainment or work completion were omitted. Studies that examined both academic and behavioral dependent variables separately were included; however, only the behavioral outcomes were considered in analysis.
3. Occurred within a public school setting. Studies that occurred in residential treatment facilities, hospitals, clinics, homes, private schools, Head Start, or Easter Seals preschool programs were excluded. In studies that examined outcomes across school and other settings, dependent measures from non-school settings were excluded from analysis.
4. Included behavior of children or adolescents between the ages of 5 and 21 with the diagnosis of autism spectrum disorder (e.g., Asperger syndrome, pervasive developmental disorder—not otherwise specified, or autism) as the intervention target.
5. Met minimum SCR design requirements (Horner et al. 2005; Kratochwill et al. 2010) to demonstrate experimental control on the dependent variable (see discussion under Assessment of Methodological Quality below).
6. Examined SM intervention data in a phase immediately preceded by a baseline or nonexperimental condition phase. Studies that examined multiple intervention protocols (e.g., token economy and SM) were included if the SM intervention was evaluated in a phase adjacent to a baseline phase. The single exception to this criterion applied to studies that collected data in a student training phase between the baseline and intervention phases.

In addition, studies that met the above conditions were excluded if the behavioral outcome was solely medical (e.g., diabetes management) or athletic skill-based (e.g., swimming stroke improvement, golf, or dancing). Additionally, if SM

data collected were from individuals not targeted by the intervention, the article was excluded. Some studies included data for students that were not involved in the intervention. For example, Sainato et al. (1992) examined the use of facilitative communication strategies for student peers working with students with ASD. The SM intervention was only implemented with the student peers; however, social behaviors were also measured as a secondary outcome with the students with ASD. Since the SM intervention was only directed toward the general education peer working with the student with ASD, only the data from students directly using the SM intervention were included in the current analysis.

Assessment of Methodological Quality When conducting meta-analyses, it is important to use only studies that demonstrate experimental control of the dependent variable (Lipsey and Wilson 2001). To verify the presence of internal validity for purposes of inclusion, two graduate students with experience and training in research methodology reviewed the methods and data section of each of the included articles. Each of the students coded the results separately, and then compared assessment results. When the student disagreed on a particular study, both would review the article a second time and discuss to consensus.

Within the pool of studies targeted, three designs paradigms were used most often: multiple baseline design (MBD) between subjects or behaviors, single baseline designs (SBD), such as reversal designs and (c) changing criteria, and changing criterion designs. Evaluation procedures for each of these designs are as follows. The "points" of change were evaluated in MBDs as a phase change within a single participant. Therefore, a MBD across three participants with a single phase change (A-B) would be counted as having one point of change for each participant, giving the design a total of three. Within this criterion, the number of participants was an important consideration for determining the level of experimental control. Thus, MBDs with three points of change were included in the analysis because the design was sufficient to demonstrate experimental control according to criteria set by Horner et al. (2005). For studies with a SBD or changing criteria, the number of phase changes was also used to determine the level of experimental control. Therefore, reversal and changing-criteria designs were evaluated to determine if three experimental "points" of control were present. Only studies with three points of control were included in further analysis.

In addition to the assessment of internal validity, the presence of sufficient data and reliability were evaluated. The researcher counted the number of data points in each phase analyzed. Designs that included phases with less than three data points were excluded in the analysis. Reliability was coded, and only studies with acceptable levels of reliability were included in the analysis. Acceptable reliability was set at a minimum of .80 for percent agreement and .60 for Cohen's kappa (Kratochwill et al. 2010).

A full review of the studies resulted in exclusion of additional articles for the following reasons: 65 studies did not target a behavioral intervention, 64 studies examined participants outside of the school setting or school age, 25 did not employ SCR methods, 23 were eliminated for not meeting minimum design quality standards for SCR, 28 studies did not examine an SM intervention in a phase adjacent to a baseline, and 6 studies included illegible graphs. Finally, 87 studies did not target an individual with an ASD. Application of the additional exclusion criteria resulted in a total of 16 studies considered for further analysis.

Data Extraction Graphic data from published studies was digitized using the GetData digital ruler (GetData, 2012). Digitizing data results in exact reconstruction of the original graphic data to numeric data. Each graph was extracted and labeled separately for each of the included studies. The graphic data were then uploaded into the GetData program where the scale of the x and y axes are set in accordance with information from the graph. Values from the GetData output were rounded to whole numbers whenever necessary to ensure an appropriate match with original study data. Following this digitizing procedure, each data set was entered into an Excel spreadsheet and attached to the variables of interest (moderators, outcomes, etc.) from each study.

Coding

Articles were coded to include information for student and study (setting, intervention component, and outcome) characteristics.

Student Characteristics The Student Characteristic Variables analyzed consisted of student age and gender. The age variable had three levels: students in primary, elementary, and secondary settings. The information provided for this variable was not consistent among the studies. Some studies reported only grade, while other reported ages. Therefore, students in the primary category were defined as prekindergarten to 2nd grade or 3–7 years old. The elementary category was defined as 3rd to 6th grade or 8 to 12 years old. Finally, the secondary category was defined as students in 7th–12th grade or 13–21 years old. Gender was defined as male or female.

Setting As all articles that met inclusion criteria were conducted in school settings, the setting variable was coded for properties specific to study implementation. Due to ambiguity related to multiply defined terms, the descriptors for "inclusion" or "inclusive" settings were replaced with the codes "General Education" or "General Education with Supports" and coding

was guided by the following decision rules: If all classroom and intervention characteristics could naturally occur within a standard classroom then it was coded as "General Education." If supports provided were specific to the student(s), then the article was coded as "General Education with Supports." As an example, student teachers and teacher aides are commonly present in classroom practice and would result in a "General Education" code unless either were specifically assigned to intervention student(s) during the conduct of the study. Following this logic, if intervention supports were observed as being in place for intervention students, then the intervention was codes as "General Education with Supports." References to accommodations and/or modifications within the intervention students' Individual Education Plan (IEP) was insufficient to meet these criteria unless explicitly observed and reported within the study. Articles conducted in classrooms solely serving students meeting eligibility criteria for special education services were coded as "Self-Contained." In studies that provided intervention services individually or in a small group that was distinct from the classroom setting with or without neurotypical peers, the term "Intervention Pullout" was coded.

Intervention Component Classification Using the framework advanced by Fantuzzo and colleagues (1987), 11 components are typically contained within SM interventions. Each of the intervention components was coded to designate the presence of that component. In addition, student participation within each component was coded to determine the extent to which the student was involved in each component of the intervention. The researcher coded the presence or absence of each intervention component along with information regarding student involvement or implementation responsibility. If the student was responsible for the component implementation, it was coded with an "S." If a teacher, researcher, or other person was responsible for implementation, it was coded with an "R." This coding strategy was used to assist in determining how important certain components were to overall effects and what impact student involvement had on outcomes for studies that used these components.

Levels of Component Analysis Two levels of analysis were necessary to answer research questions; specifically, (a) to determine if effects differ based on intervention components, and (b) to determine if there are differences among sets of intervention components based on levels of student involvement.

First Level Effects for SM were calculated based on the presence of only the specific intervention components assigned to each of the intervention methods. This analysis was used to determine how many studies with similar methods aligned. The presence of the component was the only factor used to

include studies in each of the intervention analysis. Studies were aggregated based on the presence of like components. Effect sizes were calculated for each of these groups based on the presence of only the specific intervention components. This analysis was used to determine how many studies with similar methods aligned within the broader SM construct. The presence and absence of individual components were the only factors used to include studies in each analysis. This analysis step allowed for the examination of intervention components separate from overarching intervention category.

Second Level At the second level of analysis, implementation responsibility was considered. Within each of the intervention categories determined in the first level of analysis, studies were aggregated based on the use of similar methods for student implementation. Clustering studies in this manner allowed for partitioning effects based on the degree of student responsibility for SM implementation.

Outcomes In the investigation of the effectiveness of SM for behavior, several dependent variables were identified in published articles. All were collapsed into five categories that captured similar behaviors under a common label: Disruptive, Communication, On Task, Stereotypy, and Following Rules. The Disruptive category included behaviors that are distracting to others in the classroom—both verbal and nonverbal. The Disruptive category included behaviors such as, talking out, yelling, screaming, out of seat, and aggression. The Communication category included social interaction outcomes that are not strictly communicative in nature, such as sharing, positive interactions, social skill improvement, and use of social facilitation strategies. The Communication category also examined both verbal and nonverbal communication outcomes. This included verbal behavior outcomes such as requesting, initiating verbalizations, and appropriate commenting. This category also included nonverbal behaviors such as appropriate eye contact and raising head to the appropriate position to communicate. The On Task category examined student attention to presented tasks and student engagement. The On Task variable did not include task completion outcomes. Data from task completion outcomes were excluded from the current study. Finally, the Following Rules category included desirable classroom behaviors that were not strictly disruptive, social, or communicative in nature. This included several discreet outcomes such as classroom work preparation, following teacher directions, and transitioning appropriately. This category also included student outcomes in studies that aggregated classroom rule sets. The Following Rules category was not fully independent from the other categories, as many of the rule sets included on task, nondisruptive, and social behaviors among other targeted outcomes. Due to the aggregation of varying outcomes in some studies, this category functioned more as a general measure of

SM effectiveness in classrooms rather than an indicator of specific behavioral outcomes. Despite the lack of specificity in the Following Rules category, the creation of this category allowed for the expression of all study outcomes and preserved the integrity of each of the four categories mentioned above. Separate from student characteristics and dependent variables, setting variables within the schools were examined to determine differences in study effects.

Interrater Reliability

To assess the reliability of data coding, a doctoral student in special education who had not participated in the original coding and was blind to previous coding results recoded each variable for 100 % of the studies analyzed. These results were compared to the original data coding. Reliability was calculated using a simple percent agreement or (total agreement/agreement + disagreement). Initial agreement was 92 %. Cohen's kappa (Kappa) was also calculated. Kappa is a more conservative measure of reliability that adjusts for expected chance agreement (Ary and Suen 1989). Initial Kappa was an acceptable 84 %; Kappa values above 60 % are considered good agreement (Altman 1991). Following this initial assessment of reliability, the coders resolved disagreements by discussion until agreement was 100 %.

Data Analyses

Phase Contrast Selection Selecting which phase contrasts to evaluate is an important consideration to protecting the integrity of results. Only phase contrasts that represented independent manipulation of the independent variable were evaluated with an ES. This resulted in the forward evaluation of any adjacent baseline-to-intervention phases. Data from subsequent intervention phases were not aggregated with a prior intervention phase. Each phase and phase combination in the design was only evaluated once to preserve the independence of all contrasts. For example, for designs that employed reversal logic, separate effect sizes were calculated for each baseline/intervention combination. Each of the separate effect sizes in this case was aggregated to reflect the overall outcome on the dependent variable. Therefore, an ABABAB design produced three separate effect sizes, which were then aggregated into one omnibus effect size for the design.

MBDs were treated with similar logic. In the current application, given an appropriate baseline and intervention phase, each tier of the MBD was evaluated separately for effect, and then these ESs were aggregated using the methods described below.

Effect Size For the current study, the Tau-U ES was used to determine intervention effects. Tau-U is a method for measuring data nonoverlap between two phases (A and B) that compares each datapoint in the A phase to each datapoint in the B phase and may be interpreted as *the percent of data that improve over time*. When data do not conform to parametric data assumptions, which is common in SCR, the power of a nonparametric statistic can exceed the parametric statistical analogue (Cliff 1993; Delaney and Vargha 2002; Wilcox 2010). Tau-U follows the "S" sampling distribution (Parker et al. 2011a), making it possible to calculate exact *p* values and confidence intervals. Tau-U analysis yields scores between −1.0 and 1.0, with a score of 0 indicating no difference between phases. Scores above 0 indicate improved performance across phases. Conversely, scores below 0 indicate deterioration in performance (Parker et al. 2011b). Tau-U scores from individual phase contrasts can be aggregated to provide a single omnibus ES for a variety of SCR designs, individual phase contrasts. Tau-U is also useful for a range of simple to complex designs.

Effect Size Aggregation ESs from available studies were combined to determine omnibus effects, in addition to differences between intervention component sets and moderators. Tau-U was aggregated using similar methods and presented separately. The Tau-U effect size is particularly innovative because multiple phase contrasts can be easily aggregated. Tau-U uses the S distribution to determine the variance score (Var_s). Tau-U effects were averaged after weighting each ES by the inverse of the variance score (Var_s).

Comparing Effects Analysis of intervention components and moderators followed standard practice for analyzing categorical variables (Agresti 2010; Siegel and Castellan 1988). Statistical significance for moderator variables with two groups was calculated using the Wilcoxon signed-ranks statistic (Wilcoxon 1945). Moderator variables with three or more groups were evaluated with the Kruskal-Wallis one-way analysis of variance (Kruskal and Wallis 1952). In cases where the Kruskal-Wallis showed significant differences within groups of variables, the Dunn post-hoc test (Dunn 1964) was used to determine significance between each pairwise combination of groups. The Dunn post-hoc test is a nonparametric method for comparing pairwise differences between groups. As such, it is the recommended method for evaluating data that (a) do not meet the normal distribution assumption and (b) have unequal samples sizes (Hollander and Wolfe 1999).

Effect Size Calculation Effect size calculation and aggregation were analyzed using original software developed by the first author using the Maple platform (Maplesoft 2012). The Kruskal-Wallis and Dunn post-hoc test were analyzed with SAS (Version 9.3) statistical software.

Results

Descriptive Summary of Results

Data from this study yielded 72 separate effect sizes from 16 unique studies with 28 participants. The omnibus Tau-U across all SM studies was .83 CI_{95} [.78, .78]. Within these studies, a broad range of Tau-U values were identified (from −.20 to 1.00). Given the broad range of ES across studies, additional analyses were conducted to answer questions that are critical to the implementation of SM

The current analysis found 11 unique intervention packages among the 17 published studies that targeted ASD with SM (see Table 3). Tau-U ESs ranged from −.08 CI_{95} [−.54, .38] to 1.00 CI_{95} [.39, 1.00]. A Kruskal-Wallis analysis showed significant differences among treatment packages within these groups ($p = <.0001$). Examination of statistical significance following the Dunn post-hoc procedure indicated statistically significant differences between SM package 1 (Selection of the DV, Defining the DV, Observation, and Recording) and SM package 8 (Selection of the DV, Defining the DV, Observation, Recording, Selecting the Reinforcer, and Administering the Primary Reinforcer) at $p < .05$.

Elaborating on the previous analysis, studies were also analyzed based on student responsibility for component implementation (see Table 4). Descriptive results show 16 different intervention arrangements were present in the 17 total studies when this coding scheme was applied. Tau-U effect size (ES) ranged from .05 CI_{95} [−.25 to .35] to 1.00 CI_{95} [.49–1.0]. Six of the intervention packages showed no variation in implementation responsibility. Statistical significance testing was applied between interventions that had the same components present rather than across all potential intervention arrangements. Therefore, SM packages 2, 4, 8, 9, 10, and 11 were not examined further. Within SM package 1, two variations were present in the research literature. Within this package, the intervention components varied specifically on the responsibility of completing the recording of behavior. There was no significant difference (Wilcoxon $p = 1.00$) between these interventions given higher student involvement in SM package 1A vs. 1B. Intervention package 3 showed variation in implementation based on student involvement in selecting Reinforcement. Statistical significance testing showed a significant difference (Wilcoxon $p = .01$) between package 3A versus 3B favoring intervention packages where students are involved in the selection of reinforcement (e.g., package 3B). Intervention package 5 showed variation in implementation based on student involvement in delivering instructional prompts, evaluation of performance goal, and administration of the primary reinforce. Statistical significance testing showed a significant difference (Wilcoxon $p = .01$) between these packages indicating packages with higher levels of student involvement in instructional prompt delivery yielded greater effects. Evaluation of performance goals and administration of the primary reinforcer relates to higher effects. Within SM package 6, two variations were present in the research literature. Within this package, the intervention components varied specifically on the responsibility of completing the recording of behavior, selecting the reinforcer, and evaluating the performance goal. Treatment package 6A placed responsibility for evaluating the performance goal on the student, whereas package 6B placed more responsibility on the student for recording and selecting the reinforcement. There was no significant difference (Wilcoxon $p = 1.00$) between these interventions. Finally, SM package 7 showed differences between studies based on Determining and evaluating the performance goal. Statistical significance testing showed no difference effect between these interventions (Wilcoxon $p = 1.00$).

Participant Characteristics

This study further sought to identify if participant characteristics, specifically age and gender, impacted the magnitude of change on targeted outcomes when self-monitoring was implemented (see Table 5). The age variable was sorted based on three levels: primary (EC-2), elementary (3–5), and secondary (6–12). Results yielded Tau-U ES's ranging from .56 (CI_{95} [.44–.68]) for *Primary*-aged participants, .96 (CI_{95} [.90–1]) for *Elementary*-aged participants, and .75 (CI_{95} [.66–.84]) for secondary-aged students. Kruskal-Wallis test showed statistically significant difference between participants on this variable ($p < .001$). The Dunn post-hoc procedure indicated statistically significant differences between elementary and primary age students ($p < .05$). In addition, statistically significant differences were found between secondary and primary age students ($p < .05$). These differences indicate significantly higher effect sizes for elementary and secondary age students in comparison to primary age students. In regards to gender, the Tau-U ES obtained for *males* was .82 (CI_{95} [.76–.87]) whereas the Tau-U ES obtained for *females* was .99 (CI_{95} [.89–1]). Statistical significance testing showed no significant differences between groups on this variable (Wilcoxon $p = .14$).

Studies were analyzed to identify differential effects of self-monitoring based on targeted outcomes. Five distinct categories of outcome variables (see Table 5) were identified. Tau-U effect size (ES) ranged from .25 CI_{95} [−.05 to .45] for *Disruptive behavior* to 1 CI_{95} [.39–1.0] for *Following Rules*. Statistical significance testing showed no significant differences between studies based on the dependent variable (Kruskal-Wallis $p = .81$)

Setting

Four setting categories, as defined in the methods section, were analyzed to determine if setting moderated the

Table 3 Aggregated effects by component presence

Int. package marker	Components											Tau-U	95 % CI	# of studies	# of subjects	# of ESs
	Select DV	Def DV	Obs.	Rec.	Inst. pro.	Det. goal	Eval. goal	Sel. rein	Admin sec. rein	Admin pri. rein	Graph/Chart					
1	X	X	X	X	–	–	–	–	–	–	–	.98	.84 to 1.00	3	5	14
2	X	X	X	X	X	–	–	–	–	–	–	1.00	.39 to 1.00	1	1	1
3	X	X	X	X	–	X	X	X	X	–	–	.97	.88 to 1.00	2	7	23
4	X	X	X	X	X	–	–	X	X	X	–	.34	.18 to .51	1	1	2
5	X	X	X	X	X	X	X	X	X	X	–	.86	.74 to .97	2	3	10
6	X	X	X	X	–	X	X	X	X	X	–	.81	.59 to 1.00	2	4	5
7	X	X	X	X	X	X	X	–	–	–	–	.98	.77 to 1.00	2	2	6
8	X	X	X	X	–	–	–	X	X	–	–	.32	.13 to .52	1	2	4
9	X	X	X	X	–	X	X	X	X	X	X	.91	.72 to 1.00	1	2	4
10	X	X	X	X	X	–	–	X	–	X	–	–.08	–.54 to .38	1	1	1
11	X	X	X	X	X	X	X	X	X	X	X	.72	.46 to .99	1	1	1

Select DV selection of target behavior, *Def DV* definition of target behavior, *Obs* observation of target behavior, *Rec* recording of behavior, *Inst. pro.* instructional prompts, *Det. goal* determination of performance goal, *Eval goal* evaluation of performance goal, *Admin sec. rein* administration of secondary reinforcer, *Admin pri rein* administration of primary reinforcer, *CI* confidence interval

magnitude of change that occurs on targeted outcomes (see Table 6). Results for these analyses appear in Table 6. SM implemented in *Intervention Pullout* programs generated the largest Tau-U effect size of .99 CI$_{95}$ [.91–1] whereas SM implemented in a *General Education* setting yielded the smallest ES of .74 CI$_{95}$ [.65–.82]. Despite differences in effect within this analysis, no statistically significant differences were detected between the setting variables (Kruskal-Wallis $p = .15$).

Discussion

This review assessed the effects of self-monitoring interventions on behavioral outcomes for children with ASD. Focusing solely on single case studies delivered in public school settings that met WWC quality criteria, analyses of overlap were conducted using a distribution-free metric (Tau-U). The review was undertaken to answer two central questions: (1) Does the current analysis support the use of self-monitoring with and without other self-management components as a behavioral intervention for students with ASD in school settings? and (2) Do mean Tau-U effect sizes differ by student characteristics (e.g., grade, gender, and outcome), setting characteristics (e.g., self-contained vs inclusive classrooms), or intervention characteristics (e.g., number of components, adult vs student responsibility)?

In response to the first research question, obtained effect sizes generally support the use of school-based self-monitoring interventions in addressing behavioral challenges for students with ASD. The answer to the second research question is more complex. Component analysis was undertaken to allow for the "apple to apples" aggregation of studies, as well as the "apples to oranges" comparisons that allow researchers and practitioners to make informed intervention decisions.

With reference to student characteristics, the current study found no differences in intervention effects based on participant gender. As is commonly the case within ASD literature, males represented over 90 % of the intervention sample. This finding is consistent with previous findings evaluating SM intervention effects based on gender (Briesch and Chafouleas 2009). The current study also examined grade level difference among study participants. Previous studies had found no differences between participants based on age/grade both in broad examinations of SM interventions across disability categories (Bresch and Chafoules 2009; Fantuzzo and Polite 1990) and in reviews specific to students with autism (Lee et al. 2007). The current study found significantly weaker intervention effects for participants in the primary age category (e.g., 3–7 years old) in comparison to older age participants. In general, these results are in line with those reported by Carr et al. (2014). However, more direct comparisons between the reviews are precluded by Carr and colleagues inclusion of community and home settings and broader intervention packages that

Table 4 Aggregated effects by student participation

Int. package marker	Components											Tau-U	95 % CI	# of studies	# of subjects	# of ESs
	Select DV	Def DV	Obs.	Rec.	Inst. pro.	Det. goal	Eval. goal	Sel. rein	Admin sec. rein	Admin pri. rein	Graph/chart					
1A	R	R	S	S	–	–	–	–	–	–	–	.98	.84 to 1.00	2	4	13
1B	R	R	S	R	–	–	–	–	–	–	–	1.00	.49 to 1.00	1	1	1
2	R	R	S	S	S	–	–	–	–	–	–	1.00	.39 to 1.00	1	1	1
3A	R	R	S	S	–	R	R	R	–	R	–	.91	.77 to 1.00	1	3	6
3B	R	R	S	S	–	R	R	S	–	R	–	.99	.90 to 1.00	1	4	17
4	R	R	S	S	S	–	–	S	S	R	–	.34	.17 to .51	1	1	2
5A	R	R	S	S	S	R	S	R	S	S	–	.98	.86 to 1.00	1	1	6
5B	R	R	S	S	R	R	R	R	S	R	–	.05	–.25 to .35	1	2	4
6A	R	R	S	S	–	R	S	R	S	R	–	.89	.60 to 1.00	1	3	3
6B	R	R	S	S	–	R	R	S	S	R	–	.71	.38 to 1.00	1	1	2
7A	R	R	S	S	S	R	R	–	–	–	–	.98	.70 to 1.00	1	1	4
7B	R	R	S	S	S	R	R	–	–	–	–	.98	.67 to 1.00	1	1	2
8	R	R	S	S	–	–	–	S	–	R	–	.32	.13 to .52	1	2	4
9	R	R	S	S	–	R	R	R	R	R	S	.91	.72 to 1.00	1	2	4
10	R	R	S	S	R	–	–	R	–	R	–	–.08	–.54 to .38	1	1	1
11	R	R	S	S	S	R	S	S	R	R	S	.72	.46 to .99	1	1	1

Select DV selection of target behavior, *Def DV* definition of target behavior, *Obs* observation of target behavior, *Rec.* recording of behavior, *Inst. pro.* instructional prompts, *Det. goal* determination of performance goal, *Eval goal* evaluation of performance goal, *Sel. rein* administration of secondary reinforcer, *Admin sec. rein* administration of secondary reinforcer, *Admin pri rein* administration of primary reinforcer, *CI* confidence interval

Table 5 Aggregated results by participant characteristics

Variable	Tau-U	95 % CI	# of studies	# of subjects	# of ESs
Age variable					
Primary	.55	[.44, .67]	3	6	13
Elementary	.94	[.88, 1.0]	7	12	39
Secondary	.75	[.66, .84]	7	7	17
Gender variable					
Male	.78	[.73, .84]	15	26	62
Female	.98	[.88, 1.0]	2	2	10
Outcome variable					
Disruptive behavior	.25	[−.05, .45]	2	2	3
Following rules	1.0	[.39, 1.0]	1	1	1
On task	.81	[.73, .88]	7	12	30
Social Communication/ Interactions	.89	[.82, .97]	6	13	31
Stereotypy	.91	[.80, 1.0]	2	2	7
Disruptive behavior	.25	[−.05, .45]	2	2	3

included SM in conjunction with other interventions (e.g., token economies). Finally, measured effects were aggregated and compared across study outcomes. Findings were generally positive for outcomes associated with ASD (e.g., social behaviors and stereotypy) as well as broader engagement behaviors (e.g., on task). Conversely, weaker findings were obtained when disruptive behaviors were targeted. While the current examination found no statistically significant differences between outcomes for students with ASD, more research is necessary to determine if the lower effect size is simply an artifact of low study numbers or if true differences exist when this behavioral outcome is targeted.

At the component level, findings suggest few positive benefits for self-management interventions that include components beyond the basic self-monitoring task itself (e.g., Self-assessment and Self-Recording). Specifically, a basic four-component self-monitoring intervention is generally as effective as a more elaborate 11-component intervention. The current study found a statistically significant difference between studies that employed a basic four-component self-monitoring intervention versus a more elaborate six-component intervention. These results should be interpreted with caution given the small number of studies within each category. However, these results do support previous findings that a more streamlined version of the self-management intervention (e.g., self-monitoring) is as effective as more elaborate versions of this intervention that include components beyond assessing and

recording behavior (Briesch and Chafouleas 2009; Fantuzzo and Polite 1990). With reference to student involvement, the current study found preliminary evidence that higher levels of student involvement resulted in stronger intervention effects. This effect was most prominent in studies that involved the students in recording versus studies with identical components that had a person other than the student as the primary recorder of behavior. These results diverged from previous findings that have found no differences between studies based on student involvement in the intervention (Briesh and Chafouleas 2009). This divergence may be due, in part, to the specificity of the sample in the current analysis. The Briesch and Chafouleas analysis was a more broad analysis of self-management interventions whereas the current analysis targeted students with ASD. Higher levels of student involvement in self-management interventions may be particularly important for students with ASD compared to other student populations in schools.

Substantial care was taken to identify setting-level variables that impacted measured results. Elementary schools were the most frequently employed site for self-monitoring implementation, representing approximately half of all unique effect sizes. Additionally, elementary schools reported the strongest results, demonstrating significant differences from primary setting results. Furthermore, study results in secondary settings were weaker than elementary (though not significantly so), yet still yielded significant differences from primary settings.

Table 6 Aggregated results by setting

Setting variable	Tau-U	95 % CI	# of studies	# of subjects	# of ESs
Self-contained	.81	[.73–.89]	5	9	18
Gen Ed	.73	[.64–.83]	5	9	27
Gen Ed w/supports	.86	[.71–1.0]	3	4	7
Intervention pullout	.93	[.84–1.0]	3	6	20

While the data cannot yet yield meaningful answers to the differences in results across grade level settings related to the source of these differences, the current review suggests self-monitoring is more efficacious for students with ASD in elementary or secondary settings than in primary settings. Determining why this is the case is more difficult. Students requiring selected intervention in older grades may represent a group that is developmentally ready for interventions that benefit from more mature executive functions. A broader review comparing measured effects for early childhood participants with those obtained from older students but without constraining the sample of subjects to one disability category may better serve this question.

Noteworthy results may also be found within the setting-level variables. The four intervention settings coded for the review were as follows: General Education, General Education with Supports, Intervention Pullout, and Self-Contained settings. These four settings codes resulted in a high level of interrater agreement that was supported by distinct (and stable) Tau-U results. The best results were found in studies using Intervention Pullout as the intervention setting. With an overall effect size that was virtually indistinguishable from 1.0 for this intervention setting (improvement for every possible intervention datapoint compared to every baseline datapoint in adjacent phases), the use of neurotypical peers or adult implementers as collaborative agents in settings that were within the school but separate from the classroom was highly effective. General Education with Supports was also highly efficacious for students with autism and was followed closely by Self-contained settings in measured effects. General Education was the site of the lowest measured results, resulting in only moderate effects. These results are interesting, particularly given the role of (a) typically developing peers as social and behavioral models, and (b) academic and behavioral supports in assisting children and adolescents with autism. General education settings often have an abundance of neurotypical models but few behavioral supports. Conversely, self-contained settings typically contain multiple supports with few neurotypical models. General education settings with supports have both neurotypical models and classroom supports, and intervention pullout studies have both supports and models but with even greater intensity and focus on study outcomes, conceptually supporting the continuum of obtained differences in Tau-U effect sizes.

Limitations and Directions for Future Research

There are two main limitations of the current review. First, information related to subject level variables was difficult to extract, as relevant assessment data (e.g., cognitive assessments and/or ASD-specific rating scale data) were provided for fewer than half of studies. The idiosyncratic reporting of subject information allowed for no clear conclusions to be drawn regarding subject characteristics beyond general interventional efficacy. Second, the lack of obtained ES differences for certain comparisons are likely the result of insufficient study numbers. Larger numbers of conducted studies within secondary settings and inclusive classrooms are sorely needed and published research has unfortunately failed to keep pace with shifts in instructional settings for students with ASD.

Future research should continue to explore the support for self-monitoring in school setting across additional setting-, student-, and intervention-level variables. Setting-level variables of interest include finer-grained exploration of the intervention setting to determine if the results of this study will generalize to other settings. Examples of important student-level variables include exploration of self-monitoring effects for students with disabilities other than autism as well as for at-risk youth in general education settings. Additionally, student age should be investigated as a principal variable of interest to determine if results from this review are anomalous or if SM interventions are more efficacious for older children and adolescents in general. At the intervention level, further exploration should extend beyond broad investigation of SM component-presence to more refined analysis of SM implementation. These include variables such as cueing (e.g., frequency, method of delivery) and reinforcers associated with the intervention as well as if those reinforcers are delivered for simply engaging in the self-monitoring act, self-monitoring accurately, or if the reinforcement is conferred for the behavioral improvement resulting from the SM intervention. Going forward, answers to these question are likely to yield meaningful support for school-based self-monitoring as an empirically based intervention for behavior that matches problem type, intervention features, and setting to maximize intervention effects.

References

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

*Agran, M., Sinclair, T., Alper, S., Cavin, M., 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.

*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*(2), 117–122.

*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 Developmental Disabilities, 45*(1), 136–149.

*Coyle, C., & Cole, P. (2004). A videotaped self-modeling and self-monitoring treatment program to decrease off-task behavior in children with autism. *Journal of Intellectual & Developmental Disability, 29,* 3–15.

*Fritz, J. N., Iwata, B. A., Rolider, N. U., Camp, E. M., & Neidert, P. L. (2012). Analysis of self-recording in self-management intervention for stereotypy. *Journal of applied behavior analysis 45*(1), 55–68.

*Hughes, C., Copeland, S. R., Agran, M., Wehmeyer, M., L., Rodi, M. S., & Presley, J. A. et al. (2002). Using self-monitoring to improve performance in general education high school classes. *Education and Training in Mental Retardation Developmental Disabilities, 37,* 262–272.

*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.

*Morrison, L., Kamps, D., Garcia, J., Parker, D., & Dunlap, G. (2001). Peer mediation and monitoring strategies to improve initiation and social skills for students with autism. *Journal of Positive Behavior Interventions, 3,* 237.

*Newman, B., Buffington, D. M., & Hemmes, N. S. (1996). Self-reinforcement used to increase the appropriate conversation of autistic teenagers. *Education and Training in Mental Retardation and Developmental Disabilities, 31,* 304–309.

*Parker, D., & Kamps, D. (2011). Effects of task analysis and self-monitoring for children with autism in multiple social settings. *Focus on Autism and Other Developmental Disabilities, 26*(3), 131-142.

*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.

*Reynolds, B. M., Gast, D. L., & Luscre, D. (2014). Self-management of social initiations by kindergarten students with disabilities in the general education classroom. *Journal of Positive Behavior Interventions 16*(3), 137–148.

*Rock, M. L. (2005). Use of strategic self-monitoring to enhance academic engagement, productivity, and accuracy of students with and without disabilities. *Journal of Positive Behavioral Interventions, 7,* 3–17.

*Rock, M. L., & Thead, B. K. (2007). The effects of fading a strategic self-monitoring intervention on students' academic engagement, accuracy, and productivity. *Journal of Behavioral Education, 16*(4), 389–412.

*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*(6), 540–549.

*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*(3), 205–220.

Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). Hoboken: Wiley.

Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman & Hall/CRC.

Ary, D., & Suen, H. K. (1989). *Analyzing qualitative behavioral observational data*. Mahwah: Lawrence Erlbaum Associates.

Barling, J., & Patz, M. (1980). Differences following self-and external-reinforcement as a function of locus of control and age: a social learning analysis. *Personality and Individual Differences, 1*(1), 79–85.

Bornstein, P. H., & Quevillon, R. P. (1976). The effects of a self-instructional package on overactive preschool boys. *Journal of Applied Behavior Analysis, 9*(2), 179–188.

Briesch, A. M., & Chafouleas, S. M. (2009). Review and analysis of literature on self-management interventions to promote appropriate classroom behaviors (1988-2008). *School Psychology Quarterly, 24,* 106–118.

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, 81*(1), 28–44.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*(1), 7–18.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: controversies and evidence. *Annual Review of Psychology, 52,* 685–716.

Cliff, N. (1993). Dominance statistics: ordinal analyses to answer ordinal questions. *Psychological Bulletin, 114,* 494–509.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis (2nd edl.)*. Upper Saddle River: Prentice Hall.

Delaney, H. D., & Vargha, A. (2002). Comparing two groups by simple graphs. *Psychological Bulletin, 79,* 110–116.

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics, 6*(3), 241–252.

Fantuzzo, J. W., & Polite, K. (1990). School-based, behavioral self-management: a review and analysis. *School Psychology Quarterly, 5,* 180–198.

Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature (FMHI publication no. 231)*. Tampa: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.

Hollander, M., & Wolfe, D. A. (1999). *Nonparametric statistical methods*. New York: Wiley.

Horner, R. H., & Kratochwill, T. R. (2012). Synthesizing single-case research to identify evidencebased practices: some brief reflections. *Journal of Behavioral Education, 21*(3), 266–272.

Horner, R. H., Carr, E. G., Halle, J., McGee, J., Odom, S., & Wolery, M. (2005). The use of singlesubject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165–179.

Kratochwill, T. R., & Stoiber, K. C. (2002). Evidence-based interventions in school psychology: conceptual foundations of the Procedural and Coding Manual of Division 16 and the Society for the Study of School Psychology Task Force. *School Psychology Quarterly, 17*(4), 341–389.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M. et al. (2010). Single-case designs technical documentation. Retrieved from: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association, 47*(260), 583–621.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks: Sage.

Mace, F. C., Belfiore, P. J., & Hutchinson, J. M. (2001). Operant theory and research on selfregulation. In B. Zimmerman & D. Schunk (Eds.), *Self regulated learning and academic achievement* (pp. 39–66). New York: Guilford.

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*(3), 229–246.

McLeskey, J., Landers, E. J., Hoppy, D., & Williams, P. (2012). Are we moving toward educating students with disabilities in less restrictive settings? *The Journal of Special Education, 46*(3), 131–140.

Mooney, P., Ryan, J. B., Uhing, B. M., Reid, R., & Epstein, M. H. (2005). A review of self-management interventions targeting academic outcomes for students with emotional and behavioral disorders. *Journal of Behavioral Education, 14,* 203–221.

Morris, S., & Messer, S. B. (1978). The effect of locus of control and locus of reinforcement on academic task persistence. *The Journal of Genetic Psychology, 132*(1), 3–9.

Sainato, D. M., Goldstein, H., & Strain, P. S. (1992). Effects of self-evaluation on preschool children's use of social interaction strategies with their classmates with autism. *Journal of Applied Behavior Analysis, 25*(1), 127–141.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioural sciences*. New York: McGraw-Hill.

Snider, V. (1987). Use of self-monitoring of attention with LD students: research and application. *Learning Disability Quarterly, 10.*

Wilcox, R. R. (2010). *Fundamentals of modern statistical methods: Substantially improving power and accuracy*. New York: Springer.