Meta-Analysis and Quality of Self-Monitoring Behavior Interventions with Children and Adolescents

John W. Maag[1,*]

[1]Deparement of Special Education and Communication Disorders, 202 Barkley Memorial Center, University of Nebraska-Lincoln, Lincoln, NE 68583, U.S.A

*Corresponding author: Tel: 1-(402) 472-5477. E-mail: jmaag1@unl.edu

Abstract

Self-monitoring is an intervention that has been used for decades to improve academic fluency in reading, mathematics, spelling, and promote strategies for solving problems, and increasing attention to task and decreases off-task related behaviors. There have been a few reviews of self-monitoring on the variables listed previously, but only one examined study quality and no meta-analysis of self-monitoring behavior (versus academic tasks) has been undertaken. The purpose of this review was to conduct a meta-analysis of the 20 studies that focused on self-monitoring behavior and apply the Council for Exceptional Children's eight quality indicators. Results from standard mean difference, improvement rate difference, and Tau-*U* effect size calculations were all in the effective range. Implications for practice are presented.

Keywords: Self-monitoring, Quality indicators, Off-task, Self-regulated behavior

1. Introduction

There arguably has not been as ubiquitous method for increasing youngsters' reading and mathematics fluency, use of problem solving strategies, improving attending to tasks, and decreasing undesirable behaviors as self-monitoring. It was advocated as a strategy for modifying academic performance in grade school classrooms 40 years ago (Klein, 1979). Shortly after, Hallahan Lloyd, and Stoller (1982) described self-monitoring attention (SMA) in which students observed their own behavior when cued by randomly taped tones and recorded whether or not they were paying attention or on-task. Self-monitoring has also been used as a psychotherapy assessment technique for clients monitoring their intervention progress and also as a tool for identifying maladaptive or irrational cognitions (Maag, 2018). Regardless of the use, self-monitoring can be performed using either rating scales, checklists, frequency counts of behavior, or journal entries.

Reactivity is the mechanism upon which self-monitoring results in behavior change. It occurs because individuals receive feedback by observing and recording their behavior (Mace & Kratochwill, 1988). That is, the process of observing and recording one's behavior produces a *reaction* which is a change in behavior. The oldest and most widely supported model for explaining self-monitoring is Kanfer's (1970) feedback model. The behavioral feedback individuals obtain from self-monitoring may lead them to self-evaluate their performance in accordance with either established or personal criteria. Information gained through this comparison may lead to self-reinforcement and the continuation of self-regulated behavior. Skinner's (1953) operant model is another mechanism to explain reactivity. This model focuses on the functional relation between behavior and consequences and the process of turning delayed consequences into more immediate consequences. For example, a smoker in her 20s may ignore the prospect of developing lung cancer because it is a distal consequence. However, self-monitoring the number of cigarettes smoked each day provides her with an immediate, proximal consequence that serves as a stimulus to control this behavior. It also creates feelings of guilt that can be alleviated through the behavioral principle of negative reinforcement. Guilt feels bad and the behavior of smoking fewer daily cigarettes removes the aversive feeling and reinforces further reductions in the number of cigarettes smoked. That is why reactivity is sometimes characterized as a phenomenon of "guilt control" (Maag, 2018).

The applications of self-monitoring are almost limitless and has been used to monitor medical problems such as blood glucose, diabetes and blood pressure, psychiatric conditions such as bulimia, depression and anxiety, and other activities such as weight management, physical activity, driving speed, and operation of buses. However, in education, self-monitoring has primarily been used in three areas: (a) improve fluency of reading, mathematics skills, and spelling; (b) activation of academic strategies; and (c) social behaviors such as increasing on-task and attention to activities and decreasing untoward behaviors such as those characterized by off-task, aggression, noncompliance, and tantrums (e.g., Machalicek, O'Reilly, Beretvas, Sigafoos, & Lancioni, 2007; Reid, 1996; Snider, 1987; Webber, Scheuermann, McCall, & Coleman, 1993).

The use of self-monitoring with children and adolescents—both with and without disabilities or psychiatric disorders—has a long history. Forty years ago, Zegiob, Klukas, and Junginger (1978) used self-monitoring procedures with "retarded adolescents" [sic]. Given its long history, it is somewhat surprising that less than 10 reviews of the self-monitoring literature with youngsters with or without disabilities or psychiatric conditions covering both academic and behavioral variables have been conducted. Only five reviews were conducted in or after 2005—the year quality indicators began appearing for single case research design (SCRD) studies (Horner et al., 2005) which is the most common approach for evaluating self-monitoring interventions. Reid, Trout, and Schartz (2005) reviewed self-regulation interventions, including self-monitoring, for children with attention deficit-hyperactivity disorder (ADHD). Sheffield and Waller (2010) reviewed self-monitoring studies to reduce problem classroom behaviors. Joseph and Eveleigh (2011) and Guzman, Goldberg, and Swanson (2018) reviewed the effects of self-monitoring on reading performance of students with and without disabilities. Bruhn, McDaniel, and Kreigh (2015) reviewed self-monitoring interventions for students displaying behavior problems. Briesch and Chafouleas (2009) reviewed various self-management interventions including, but not limited to, self-monitoring. Finally, Maggin, Briesch and Chafouleas (2013) used the data set from the Briesch and Chafouleas review (n = 30) to analyze study quality by applying the Horner et al. (2005) criteria.

There are various methodological limitations to these reviews. For example, Sheffield and Waller (2010) and Bruhn et al. (2015) included self-management procedures such as self-evaluation and self-reinforcement and interventions with multiple components in addition to self-monitoring such as peer mediation, peer tutoring, contingent attention, and function-based supports. Briesch and Chafouleas (2009) also focused on 10 additional self-management components in addition to self-monitoring. Only three reviews used meta-analytic techniques for the variables—two for reading and ADHD (Guzman et al., 2018; Reid et al., 2005)—and one for non-academic content such as on-task and off-task (Briesch & Chafouleas, 2009). However, only one review assessed study quality (Maggin et al., 2013), but focused on 11 components of self-management and used the now dated Horner et al. (2005) quality indicators rather than the Council for Exceptional Children's (CEC) 2014 eight quality indicators comprising 22 items. Finally, none of the reviews addressed, nor accounted for, publication bias which is an important component emphasized in the PRISMA document (Liberati et al., 2009).

The purpose of the present review was to conduct a meta-analysis of SCRD studies using self-monitoring to address prosocial and problem behaviors and to assess their quality using the CEC (2014) criteria. Academic variables such as reading and math were not included because the self-monitoring techniques for those are quite straightforward: Students typically keep frequency counts of either the number of answers written (productivity) or number of correct answers written (accuracy; e.g., Maag, Reid, & DiGangi, 1993). Conversely, self-monitoring social behaviors, such as noncompliance or aggression, can be quite difficult because of the different situations under which they may occur. For example, self-monitoring the number of times a child hits, pushes or throws something at peers during recess or

physical education would be difficult logistically in terms of the recording technique, materials, and storage methods. Similarly, self-monitoring either attention (i.e., on-task/off-task) can also be problematic because those variables usually require a cueing apparatus (recorded beeps or vibrations using a device such as the MotivAider ®). Once the cueing device sounds or vibrates, the student asks himself "Am I paying attention?" and has to stop what he is doing to self-observe the behaviors that constitute paying attention. By stopping the activity to ask himself if he is paying attention necessarily distracts him from paying attention to the task at hand for which he is self-monitoring. Furthermore, there have not been any reviews of self-monitoring behavior as a separate intervention (i.e., apart from other self-management components) that have both calculated effect sizes and examined study quality.

2. Method

A systematic search was performed to identify the extent research regarding the use of self-monitoring behavior interventions. The search methods were consistent with the 12-item PRISMA statement for reporting meta-analyses (Liberati et al., 2009). The purpose was to ensure clarity and transparency of conducting systematic reviews.

Academic Search Premier was the search source with the following selected databases: ERIC, MedLINE, PsycARTICLES, and PsycINFO. The following Boolean terms/phrases were used: ("self-monitoring behavior") AND ("children") OR ("adolescents") OR ("youth") OR ("child") OR ("teenagers") OR ("students") OR ("students with disabilities"). In addition, ancestral searches were conducted of four journals that publish exclusively or primarily SCRD studies: *Journal of Applied Behavior Analysis*, *Journal of Behavioral Education*, *Journal of Positive Behavior Interventions*, and *Behavioral Disorders*. Finally, references of three self-monitoring reviews published on or subsequent to 2005 that focused on non-academic variables were searched (Briesch & Chafouleas, 2009; Bruhn et al., 2015; Sheffield & Waller, 2010).

2.1 Eligibility Criteria and Study Selection

Studies included were those only using SCRDs. Studies had to be in English and published in peer-reviewed journals between January 1, 2005 and December 30, 2018. The date of 2005 was selected because it was the year quality indicators for SCRD studies were developed and published (Horner et al., 2005). In addition, Cook and Tankersley (2007) discussed the problems of trying to "retrofit" present-day quality indicators to studies published years or even decades ago. Participants considered in the present review were students in K-12 settings whose behavioral issues required intervention. Interventions considered were self-monitoring for non-academic content behaviors such as, but not limited to, on-task/off-task, aggression, noncompliance, classroom social skills, walking around, playing with objects, talking without permission, refusal to follow directions. Self-monitoring has become such a ubiquitous technique that it is often incorporated into other interventions such as social skills training or the Check-in Check-out program as a way to enhance treatment

effectiveness. In order for these types of studies to be included, there must be one phase consisting only of self-monitoring that is preceded by a baseline phase.

Self-monitoring traditionally consists of two parts: self-observation and self-recording (Hallahan et al., 1982). Self-graphing has been subsequently considered to be an important final component (DiGangi, Maag, & Rutherford, 1991). However, the addition of components such as self-evaluation, self-reinforcement, or goal setting move the intervention into the more umbrella term of self-management (Maag, 2018). Therefore, studies were excluded that contained these latter three, or similar, techniques. In addition, studies were excluded that contained contingent reinforcement for either displaying target behaviors or for self-monitoring accurately. The reasons were because (a) self-monitoring produces reactivity and negates the need for external contingent reinforcement and (b) students do not have to self-monitor accurately for reactivity to occur (Maag, 2018). The goal was to avoid as many confounding variables that could potentially impact evaluating the effectiveness of self-monitoring.
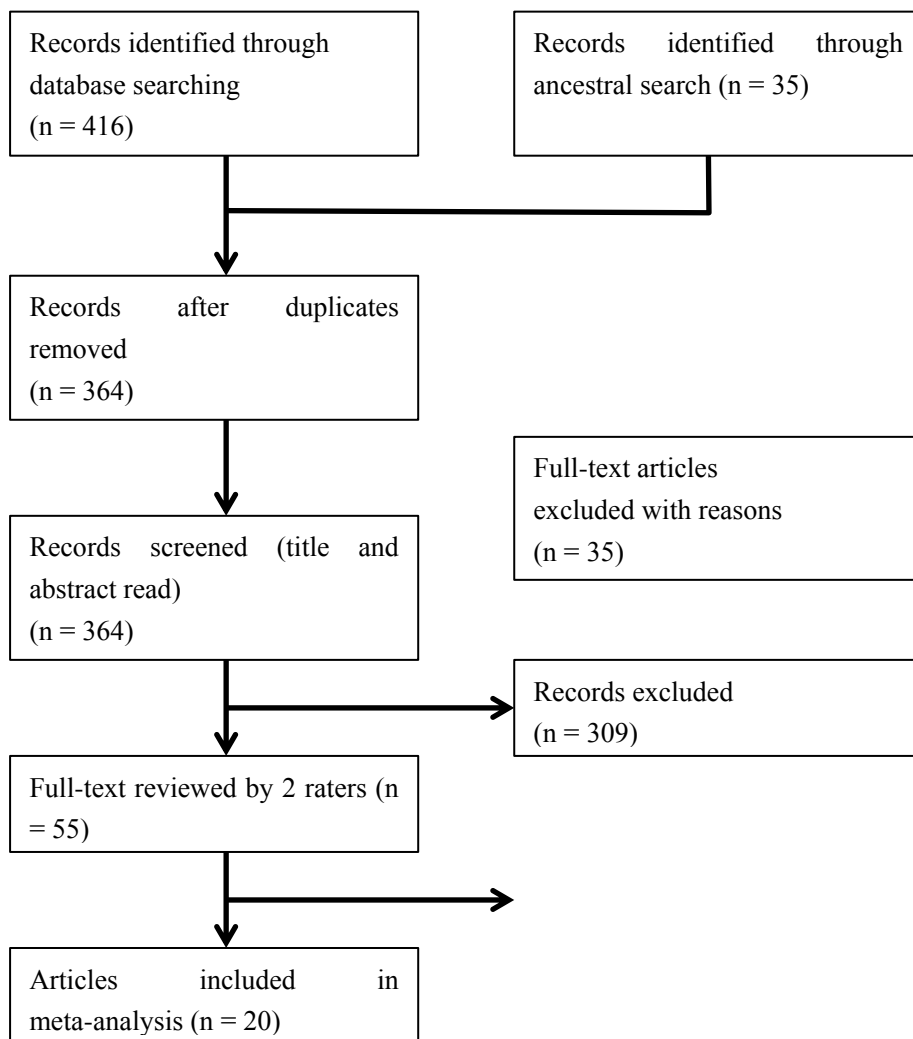
Figure 1. Search Results Using PRISMA Guidelines

Studies were identified and retained at different stages based on PRISMA guidelines, and the results are displayed in Figure 1. There were 451 total records identified that were articles no older than 2005 from peer-reviewed journals. Of those, 55 were read in their entirety (i.e., method sections for inclusion/exclusion criteria). Two graduate students were trained by the researcher how to read each of the 55 studies method sections. After engaging in the flow of information process, there were 20 articles retained for the current review. One graduate student read all 55 studies (i.e., method sections) while the other read 12 randomly selected studies and their interrater agreement was 100%.

2.2 Coding Procedures

Descriptive characteristics. The 20 articles retained from the search were coded along six variables: (a) participant characteristics, (b) diagnosis/educational label, (c) settings, (d) type of design, (e) dependent variables, and (f) type of self-monitoring. Two graduate assistants were trained by the experimenter to code the seven variables. Six studies were randomly selected and the experimenter demonstrated the coding process on three through instructions and modeling. The two graduate assistants coded the remaining three studies with the experimenter providing performance feedback. The two graduate students then each coded the remaining studies independent of each other. Inter-rater reliability (IRR) was calculated for 10 randomly selected studies (50%). This percentage was congruent with other published SCRD meta-analyses (e.g., Gage, Lewis, & Stichter, 2012; Losinski, Maag, Katsiyannis, & Parks-Ennis, 2014; Maggin et al., 2013).

Methodological quality. Two graduate assistants appraised the quality of each article based on the Council for Exceptional Children's (CEC) *Standards for Evidence-Based Practices* (2014) that consisted of 22 component items across eight quality indicators (QIs) for SCRDs. The same training format used for coding descriptive characteristics was used for coding QIs. A binary score of one (met) or zero (not met) were used in the coding scheme (i.e., absolute coding) rather than using weighted coding (Common, Lane, Pustejovsky, Johnson, & Johl, 2017). The reason for using the more stringent approach was because of the extent to which self-monitoring has been used throughout the decades with positive results as typically determined through visual inspection. In essence, it is a proven evidence-based practice, and the absolute quality of those studies is the focus in the present review.

A coding sheet with the 22 components across eight QIs for SCRD studies was created in Excel©. The sheet consisted of three columns. The first column contained the QI, the second column had the description, and the third column consisted of clarification developed by Common et al. (2017).

2.3 Statistical Analysis

Data extraction. Data were extracted from the graph(s) in each study using Enguage Digitizer (Mitchell, 2002)—an open source digitizing software package that converts graphic image files (e.g., .jpg, .bmp) into numerical data. Enguage is a free software package that is comparable to Biosoft's Ungraph 5.0 that was recommended in the manual developed by Nagler, Rindskopf, and Shadish (2008) for conducting SCRD meta-analyses and used in

previous meta-analyses (e.g., Losinski et al., 2014). In addition, Losinski and his colleagues also converted all scores into percentages setting the upper level and lower level of the *y* axis on all students to 100 and 0, respectively, before extraction. Their rational was to address the inherent subjectivity in which target variables were operationally defined, for example "aggression" versus "hitting and pushing" or different behaviors that make up "off-task."

Effect size calculations. Horner, Swaminathan, Sugai, and Smolkowski (2012) noted that currently there is no consensus for the method for quantifying outcomes for use in SCRD meta-analysis, although some effect size calculations may be more robust or appropriate than others depending on data characteristics. Therefore, three types of effect sizes were calculated in the present meta-analysis. Standard mean difference (SMD) was calculated because it is the SCRD analog or variation of Cohen's *d* statistic where the mean of the baseline phase is subtracted from the mean of the intervention phase and divided by the pooled standard deviation (Busk & Serlin, 1992). The similarity to Cohen's *d* makes SMD an important statistic for comparison to non-single-case methods and also for statistically analyzing moderator variables and accounting for publication bias if there are at least three participants in a study to calculate BC-SMD. However, SMD is considered by some unreliable because of small number of observations and floor effects limiting variability and results in overestimates of the parametric treatment effects (Horner et al., 2012; Scruggs & Mastropieri, 2012). Losinski et al. (2014) dealt with this problem by establishing a ceiling values of the 3rd quartile total distribution in order to account for statistical outliers disproportionally affecting the outcomes of the studies when aggregated. This practice (i.e., 3rd quartile ceiling) was also used in the present study resulting in a ceiling of $d = 3.86$. Improvement rate difference (IRD) was also computed because it provides an effect size similar to the *risk difference* used in medical treatment research which has a proven track record in hundreds of studies (Parker, Vannest, & Brown, 2009). Finally Tau-*U* values were computed because it controls for monotonic trend (i.e., increasing data during baseline). The IRD and Tau-*U* were calculated using the www.singlecaseresearch.org/calculators. For studies in which improvement was in the decreasing direction, the correction feature for Tau-*U* was not used (i.e., only Tau). Effect sizes for SMD were computed using a calculator.

Additional analysis. Moderator variables were addressed by computing independent t-tests to compare differences in effectiveness of studies on the following three variables: type of dependent variable (e.g., on-task versus off-task), age of participants, and type of disability/DSM diagnosis. The rationale for type of dependent variable and age was because Losinski et al. (2014) found significant differences in interventions based on contextual variables when examining dependent variable and age. Also, self-monitoring interventions may be more effective as children get older and have a better understanding of the process (e.g., Maag et al., 1993). Age was broken down into two groups young and old (i.e., 3rd quartile and higher). In terms of disability, two groups were formed. The first group consisted of participants who were either at-risk or diagnosed with ADHD. The reason for this grouping is because children with ADHD have been consistently found to display characteristics of inattention and another 5% of children not meeting diagnostic criteria (i.e.,

at-risk) display behaviors indicative of inattention (Sayal, Prasad, Daley, Ford, & Coghill, 2018). The second group consisted of participants who had co-morbid conditions.

These t-tests were computed for all three effect size calculations: SMD, IRD, and Tau values for each of the moderator variables. The reason for using t-tests was because there were not enough studies with three or more participants to use the BC-SMC as the method to assess moderator variables via regression.

2.4 Publication Bias

Publication bias, or the "file drawer" effect refers to presence of potential bias existing because of a greater likelihood that published research shows positive findings (Rosenthal, 1979). In a meta-analysis of group design studies, the Meta-Win's Fail-Safe function (Rosenberg, Adams, & Gurevitch, 2000)—based on Orin's (1983) procedure—can be used to estimate the number of studies with null results sufficient to reduce observed effect size to a minimal level (i.e., < .20). However, there is no comparable formula in SCRD meta-analyses, at least in the absence of calculating BC-SMD. Therefore, to reduce the likelihood of the "file drawer" effect, the number of cases with no effect (i.e., 0) were added to the group of study effect sizes to reduce the overall effect to insignificant or suspect levels (d <.20; IRD < .37; Tau <.20). This process results with the number of participants in potentially "filed" studies (i.e., not submitted for publication for whatever reasons) needed to reduce effect sizes of included studies to insignificant levels (i.e., no observed effect).

2.5 Inter-Rater Reliability

Interrater reliability (IRR) data were conducted on 10 randomly selected articles out of the 20 included studies for a total of 50% of studies on the six coded study characteristics and QIs. Interrater reliability was calculated both for study characteristics and QI components by dividing the total number of agreements by the total number of agreements plus disagreements for each item and averaged for all items. Two graduate research assistants coded the articles for all variables and IRR for study characteristics was 87.1% (range: 72% -100%; SD=8.385) and 89.1% (range 60% - 100%; SD=13.289) for QIs.

3. Results

Results are presented in three sections. The first section addressed descriptive features obtained from the studies including characteristics of participants and settings, design features, dependent variables, and self-monitoring techniques. The second section presents the extent to which studies met each of the 22 component items pertaining to SCRD across CEC's eight QIs. The final section contains effect size and moderator results.

Table 1. Participant Characteristics by Study

Study		Male				Female			
	N	Mean Age	Ethnicity	Disability/Diagnosis	N	Mean Age	Ethnicity	Disability/ Diagnosis
1. Agan et al. (2005)	6	14	Caucasian African-American	Asperger's At-risk				
2. Amato-Zech et al. (2006)	2	11	N/I	SLI & LD	1	11	N/I	ED & SLI
3. Bedesm (2012)	2	13	N/I	OHI				
4. Briere & Simonse (2011)	1	N/I	Caucasian	At-risk	1	N/I	Hispanic	At-risk
5. Coughlin et al. (2012)	1	7	N/I	MID	2	7	N/I	MID
6. Creel et al. (2006)	3	12	African – American Caucasian Hispanic	ADHD	1	11	African-A merican	ADHD
7. Crutchfield et al. (2015)	2	14	Caucasian	Autism Down Syndrome ADHD				
8. Davis et al. (2014)	1	15	Caucasian	At-risk				
9. Graham-Day (2010	2	16	Caucasian	ADHD	1	16	Caucasian	ADHD
10. Gulchak (2008)	1	8	Caucasian	EBD				
11. Gumple (2007)	6	12.7	N/I	EBD Williams Syndrome MID				
12. Gureasko-Moore (2007)	6	11.6	Caucasian	ADHD				
13. Harris et al. (2005)	3	N/I	Caucasian African-American	Tourette Syndrome ADHD At-risk	3	N/I	Caucasian African-A merican	MDD OCD
14. Kartal (2015)					4	5.6	N/I	MID Down Syndrome ModID & ADHD
15. Rafferty (2011)	2	10	Caucasian	ADHD & LD ADHD	1	10	Caucasian	ADHD
16. Rafferty (2012)	2	7.5	African-American	ED	2	7.5	Hispanic Caucasian	ED
17. Rafferty & Raimondi (2009)	1	8.5	African-American	ED	1	8	Hispanic	ED & LD
18. Rock (2005)	7	10.3	Caucasian Asian	Asperger Syndrome Floating Harbor Syndrome SLI	2	8	Caucasian	At-risk
19. Rock & Thread (2007)	4	11.8	African-American	LD & ADHD At-ris	1	14	Caucasian	Autism & ModID
20. Wills (2014)	2	13	Caucasian Native American	LD & ADHD ADHD				

Note. N/I= no information; ADHD=Attention deficit hyperactivity disorder; EBD=emotional and behavioral disorder; ED=emotionally disturbed; LD=learning disability; MDD=major depressive disorder; MID=mild intellectual disability; ModID=moderate intellectual disability; OCD=obsessive compulsive disorder; SLI=speech language impairment

3.1 Descriptive Features of Included Studies

Characteristics of participants and settings. A total of 74 participants (56 male and 16 female) were included in the 20 studies contained in this analysis. One study did not specify the gender of its two participants (Bedesem, 2012). Participants' ages ranged from five years old (Kartal & Ozkan, 2015) to 16 years old (Graham-Day, Gardner, & Hsin, 2010) with an

average of 11.3 for males and 9.8 for females. Two studies that contained eight participants only provide grade level but no age (Briere & Simonsen, 2011; Harris, Friedlander, Saddler, Frizzelle, & Graham, 2005). In terms of diagnosis/label, 13 participants had none but were identified as "at-risk." There were nine participants with ADHD and 19 more that had various co-morbid conditions. There were two studies in which participants had unusual conditions: Tourette Syndrome (Harris et al., 2005) and Floating Harbor Syndrome (Rock, 2005). There were four studies for a total of 17 participants that did not provide ethnicity (Amato-Zech, Hoff, & Doepke, 2006; Bedesem, 2012; Coughlin, McCoy, & Kenzer, 2012; Gumpel, 2007). The remainder of the participants ethnicity was as follows: 34 Caucasian, 17 African American; 4 Hispanic, 1 Asian, and 1 Native American. Table 1 summarized participant characteristics.

The majority of studies were conducted in a general education classroom ($n = 12$). The next most common setting was identified as self-contained classroom ($n = 4$). There was one studies conducted in a resource classroom. Two studies indicated special education classroom but did not specify which type (Crutchfield, Mason, Chambers, Wills, & Mason, 2015; Graham-Day et al., 2010) while another study was conducted in a special school for students with emotional and behavioral disorders (Gumpel, 2007).

Design features. The majority of SCRDs were multiple baseline ($n = 11$) and reversal ($n = 4$). There were five remaining studies that used other designs. Crutchfield et al. (2015) and Rock (2005) embedded reversal phases within a multiple baseline. The rationale provided by Rock was that she only had two participants which typically does not meet experimental rigor for a multiple baseline design. Graham-Day et al. (2010) used an alternating treatments design with three consecutive data points of baseline before alternating self-monitoring, self-monitoring with contingent reinforcement, and baseline (i.e., business as usual). Gumpel (2007) used a changing conditions design in which baseline was followed by self-monitoring with non-contingent reinforcement (i.e., reinforcement for behaviors unrelated to self-monitoring), return to baseline, and followed by self-monitoring with contingent reinforcement. Harris et al. (2005) used a counter balanced multiple baseline design to assess the efficacy of self-monitoring attention versus self-monitoring academic performance, but with no return to baseline between the two independent variables. Rafferty and Raimondi (2009) also used a counter balanced multiple baseline design.

Dependent variables. There were a large variety of dependent variables targeted in the studies reviewed that are presented in Table 2. Some studies had only one dependent variable such as following directions (Agran et al., 2005) to such multi-component variables as those that make up SLANT behaviors: sit up, look at the person talking, activating thinking, noting key information, and tracking the talker (Amato-Zech et al., 2006). Curiously, one study targeted engaging in the self-monitoring process (i.e., looking at the self-monitoring card, tallying self-monitoring card) in addition to looking at a spelling list and hand raising (Rafferty, Arroyo, Ginnane, & Wilczynski, 2011). There were also a number or dependent variables that lacked any specific target behavior such as "engaged in behaviors specified by the teacher," "percentage of classroom preparation skills as determined by the teacher"

(Briere & Simonsen, 2011; Gureasko-Moore, DuPaul, & White, 2007). There were nine studies that targeted both inappropriate and appropriate behaviors

Table 2. Dependent Variables and Self-Monitoring Techniques by Study

Study	Dependent variables	Type of Self-Monitoring
1. Argran et al. (2005)	Following directions; acknowledging person giving directions, beginning and staying engaged in activity.	"+" mark in a box on the self-monitoring sheet each time participant performed a step in the task-analysis, or "-" in the box if a step in the task-analysis was not completed.
2. Amato-Zech et al. (2006)	Off-task behavior: absence of one or more SLANT behaviors (sit up, look at person talking, activate thinking, note key information, track the talker); off task motor (out of seat, ripping pages); off task verbal (verbalizations not relevant or permitted during assigned task; off task passive (disengagement for a period of at least 3 consecutive seconds (looking away from assigned material). On-task behavior: actively or passively attending to instruction or assigned work	Frequency tallies on recording sheet ("yes" or "no") with MotivAider® cuing vibrations to either "I was paying attention" or "No I was not paying attention."
3. Bedsem (2012)	On-task behavior: in seat (buttocks on seat of chair unless given permission to stand); quiet, unless given permission to speak; not disrupting others (passing notes, touching students or their possessions); following teacher directions; eyes on task, teacher, or speaker.	CellF-Monitoring procedure. Participant received four text messages during each session saying "Are you on task" with a choice of "yes" or "no" for the response.
4. Briere & Simonsen (2011)	On-task (engaging in behaviors that are specified and desired by the teacher at that time); disruptive off-task (talking when required to remain quiet, (out of the seat, touching students inappropriately, not following directions. Off-task (student oriented toward anything other than the task at hand, but not negatively impacting the environment	Participant received two self-monitoring sheets with references to one or two of the three school-wide expectations. They rated the extent to which their behavior was appropriate (0=never, 1=occasionally, 2=often, or 3=always. The second sheet asked them to use the same rating scale to the extent to which on-task. Vibration cueing devise cued them every 5 minutes to self-rate.
5. Coughlin et al. (2012)	On-task behavior (engaging in a designated academic task, (staying in designated work area, using designated work materials, working independently). Off-task behavior (leaving designated work area, sitting in designated work area idly, talking to others, engaging in self-stimulatory or destructive behavior).	Participants received self-monitoring sheet and visual cue cards containing one-inch picture of a cartoon character selected by each participant individually in a multiple-stimulus procedure. Placed sticker on self-monitoring sheet after completing a line of work if engaging in target behaviors.

6. Creel et al. (2006)	Pick up folder from crate entering classroom and immediately go to and remains in seat; has writing utensil, and paper/notebook; hands in homework on time, completes daily classwork (e.g., board work, class activities) without being reminded.	Student checklist with the dependent variables listed and a place to check each as they occurred or didn't occur during observation sessions.
7. Crutchfield et al. (2015)	Stereotypy (hand flapping or waving in front of face, hands or objects in mouth, grunts, repetitive laughing, repeating words and phrases); verbalizations not directed to a partner.	I-Connect mobile electronic prompting and monitoring device. The device flashed every 30-s with written question appearing "Quiet hands and mouth?" Participants would check either a "yes" or "no" box that appeared on the screen
8. Davis et al. (2014)	On-task (eyes on teacher, whiteboard, or work, work material on desk; writing on paper)	Participant given VIBRALITE 3 wrist watch, that vibrated every 30 seconds. On a self-monitoring card participant circles "on task" or "off task" when cued.
9. Graham-Day et al. (2010)	On-task (sitting facing forward, both feet on floor, eyes on academic task responding to teacher questions, following directions, writing answers, asking academic questions)	Student checklist with 15 opportunities during a 20-minute period asked "yes" or "no" if on-task at the moment they heard the designated stimulus (i.e., auto chimes every 2-m on average.
10. Gulchak (2008)	On-task (hands away from face, not picking nose, complete work assigned, raise hand to ask questions)	Palm Zire 72 handheld device with HanDBase" self-monitoring software with on-task and the behaviors and marked "yes" or "no" on the screen at 10 minute intervals during 1-hr.
11. Gumple (2007)	No interaction (not engaged in any activity with another person farther than 2m. from peers with no verbal or physical contact between them) Positive interaction (engaged in social interaction with peers such as speaking in a non-confrontational manner no more than 1.5 m. from the other individual and that did not include any sort of cursing, shouting, pushing, name calling, hitting, and making forceful bodily contact with someone else) Inappropriate interaction (defined as any interaction that engaged in aggressive act such as cursing, shouting, pushing, name calling, hitting, and making forceful bodily contact with someone else. Contact with an adult. (being less than 1.5 m. from an adult during recess and engaged in conversation or an activity with that adult)	Frequency tallies on recording sheets with target behaviors listed and buzzing cueing device attached to participants' shirt that went off every 2-m.
12. Gureasko-Moore (2007)	Percentage of classroom preparation skills as determined by the teacher; percentage of homework behaviors completed by students and documented by parents and teachers	Participants had two forms: student log and self-monitoring checklist that contained behaviors from the classroom preparation behavior checklist. Participants wrote behaviors that aided them in or inhibited them from achieving goals and rated them on 5-point scale (0 no effort, 5 best effort)

13. Harris et al. (2005)	On-task behavior was operationally defined as occurring when a student (a) focused her or his eyes on the spelling list, practice paper, or self-monitoring tally sheet; (b) executed any step in the spelling study procedure; or (c) asked for help. Academic performance was operationally defined as the total number of words a student wrote correctly when practicing the items from his or her weekly spelling list during each spelling period (Harris, 1986).	Self-monitoring attention by making mark in "yes" or "no" column to question "Was I paying attention?" Audio tones cued an average of 45-s Self-monitoring performance by participants counting number of times that weekly spelling words were practiced correctly at the end of each spelling period and graphing that number.
14. Kartal (2015)	On-task (attending to tasks during activity and displaying behaviors appropriate for these tasks). Off-task (walking around, crying, screaming, looking around, sitting at table doing nothing, playing with toys or materials inappropriately, leaving the playground)	Participants received forms with pictures showing correct behaviors and would put a smiley face sticker for behaving correctly
15. Rafferty (2011)	On-task (looking at spelling list or self-monitoring card, tallying self-monitoring card, hand raising for assistance Off-task (doodling on the spelling list or self-monitoring card, looking anywhere other than the list or self-monitoring card, talking to peers)	Frequency tallies with taped tones as cueing device heard while wearing headphones on recording sheets with the question "Am I on task?" and two columns with "yes" or "no" that the participant tally mark. On-task was defined using the SLANT strategy: sit up, look at speaker, answer questions, note key information, track talker.
16. Rafferty (2012)	On-task (sitting in seat, eyes on speaker or materials, asking or answering questions, commenting on related topic after raising hand and waiting to be acknowledged, speaking to peers in group when instructed by teacher, writing information related to material) Oral reading fluency rate	Frequency tallies on recording sheet with the question "At this moment, am I on-task?" with "yes" and "no" columns to mark. MotivAider® cuing vibrations every 2-m to either answer the question and mark the corresponding column
17. Rafferty & Raimondi (2009)	On-task (looking at and writing on self-monitoring card or math worksheet, using manipulatives, asking for help). Academic performance (number of math problems completed and completed correctly)	Self-monitoring on-task: Frequency tallies on recording sheet with the question "At this moment, am I on-task?" with "yes" and "no" columns to mark. Tape recorded tones to cue participant to mark corresponding column. Tones occurred every 5-m. Self-monitoring performance: math worksheets participant completed, used answer key at the of session to score answers and graph number.

18. Rock (2005)	Academic disengagement (not in seat or not working quietly on paper and-pencil math task	ACT-REACT self-monitoring strategy: articulate goal, create work plan, take pictures, reflect using self-talk, evaluate progress and ACT again. Combined self-monitoring attention and performance.
	Academic engagement (participating in assignments by being in seat working quietly on paper and-pencil math task)	Frequency tallies on recording sheet with the question "At this moment, am I on-task?" with "yes" and "no" columns to mark. Tape recorded tones to cue participant to mark corresponding column. Math worksheets worksheets participant completed, used answer key at the of session to score answers, and graph number.
19. Rock & Thread (2007)	Academic engagement (participate in math related independent seatwork assignments by being in seat, eyes on papers, working quietly on assigned tasks).	ACT-REACT self-monitoring strategy: articulate goal, create work plan, take pictures, reflect using self-talk, evaluate progress and ACT again. Timing device cuing every 5-m. Recording sheet to mark number of problems completed and check whether or not behavior goal was met.
	Academic disengagement (out of seat, talking to classmates about subjects unrelated to tasks, staring off into distance, head on table, insulting peers, drawing, hitting peers, spitting, playing with objects).	
	Math performance (total number of math problems completed, percentage of the total number of problems completed correctly.	
20. Wills (2014)	On task (eyes on teacher, verbal responses, raising hand, following directions, writing, reading, or otherwise actively completing assigned task such as typing on computer, engaged with materials).	I-Connect self-monitoring. Android application. Every 5-m question appears "Are you on task?" with a "yes" and "no" option to mark.
	Disruptive behavior (engaging in behavior that was contrary to the behavioral expectations of the assigned task that could potentially disrupt instruction).	

Self-monitoring technique. The majority of the techniques were self-monitoring attention (SMA) which involved participants having a recording sheet with a main question at the top such as "Am I paying attention" with two columns—one with "yes" at the top and the other with "no" at the top (n = 13). They were cued by various devices such as hearing tones or feeling vibrations from devices such as the MotivAider® to make a mark in the column as to whether they were paying attention. All participants received instruction in this process and the behaviors that made up "paying attention." Two studies used both SMA and self-monitoring performance (SMP). In the latter case, participants would receive a worksheet mathematics problems and an answer key. Upon completing of the worksheet, they would score their own work and graph the number (Harris et al., 2005; Rafferty & Raimondi, 2009). Three studies used a Likert-type rating scale (e.g., 0= no effort, 5 = best

effort) for participants to appraise their performance (Briere & Simonsen, 2011; Creel, Fore, Boon, & Bender, 2006; Gureasko-Moore et al., 2007).

Some studies used more technologically sophisticated version of SMA. For example, Bedesem (2012) used the CellF-Monitoring procedure in which participants received four text messages on their phone saying "Are you on task" with a place to press either the "yes" or "no" icon. Instead of using the MotivAider® to cue participants, Davis, Dacus, and Bankhead (2014) had them wear VIBRALITE 3 wrist watches that vibrated every 30 seconds. Crutchfield et al. (2015) used the I-Connect mobile electronic prompting and monitoring device that flashed every 30 seconds with questions such as "Quiet hands and mouth?" that participants would check either "yes" or "no box that appeared on the screen. Finally, Gulchak (2008) gave participants a Palm Zire 72 handheld device with HanDBase self-monitoring software that displayed on-task behaviors at 10 minute intervals for students to mark either "yes" or "no." Table 2 presents a summary of the self-monitoring techniques by study.

3.2 Methodological Quality Indicators

CECs *Standards for Evidence-Based Practices* (2014) that consisted of 22 component items across eight quality indicators (QIs) for SCRDs were used to determine methodological quality of reviewed studies. None of the 20 studies met all 22 items, although eight met 21 (Bedesm, 2012; Creel et al., 2006; Gureasko-Moore et al., 2007; Rafferty, 2012; Rafferty et al., 2011; Rock, 2005; Rock & Thread, 2007; Wills & Mason, 2014). The lowest score (16) was obtained for Coughlin et al. (2012). Overall, quality of the 20 studies was relatively high (mean=19.85, SD=1.308, range 16 – 21). The lowest score (7 [35%]) was for item 5.3 which required studies to assess and report implementation fidelity regularly throughout intervention and for each interventionist. However, the other two items under fidelity, 5.1 and 5.2, were relatively high with scores of 20 (91%) and 19 (86%), respectively. Further, only 11 (50%) of the 22 items were met by all 20 studies: 1.1 *Context and setting*, 2.1 *Participant demographics*, 2.2 *Participant disability or risk status*, 4.1 *Detailed intervention procedures*, 4.2 *Description of materials*, 5.1 *Implementation fidelity*, 6.1 *Systematic manipulation of independent variable*, 6.2 *Description of baseline conditions*, 6.4 *Three demonstrations of experimental effects*, 6.5 *Baseline phases include three data points*, and 8.1 *Graphs clearly represent outcome data*. Table 3 displays the results for each component across QIs for all 20 studies. The grey shaded cell are those that received a 0 (N = 41 [9%]). Taken together, 19 out of 20 studies (95%) met 80% or more of the QIs (M=90%, SD = 5.794, range=73% - 95%).

Table 3. Quality Indicators Met by Study

Study	1.1	2.1	2.2	3.1	3.2	4.1	4.2	5.1	5.2	5.3	6.1	6.2	6.3	6.4	6.5	6.6	7.1	7.2	7.3	7.4	7.5	8.1	Total
1. Argan et al. (2005)	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	0	1	1	**19**
2. Amato-Zech et al. (2006)	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	**20**
3. Bedesm (2012)	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	**21**
4. Briere & Simonse (2011)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	**20**
5. Coughlin et al. (2012)	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	0	0	0	1	**16**
6. Creel et al. (2006)	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**21**
7. Crutchfield et al. (2015)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	**20**
8. Davis et al. (2014)	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	**20**
9. Graham-Day (2010)	1	1	1	1	0	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	0	1	**18**
10. Gulchak (2008)	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	**19**
11. Gumple (2007)	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	**19**
12. Gureasko-Moore (2007)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	**21**
13. Harris et al. (2005)	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	**19**
14. Kartal (2015)	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	**19**
15. Rafferty (2011)	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	**21**
16. Rafferty (2012)	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	**21**
17. Rafferty & Raimondi (2009)	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	**20**
18. Rock (2005)	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	**21**
19. Rock & Thread (2007)	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	**21**
20. Wills (2014)	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	**21**
Total	20	20	20	19	14	20	20	20	19	7	20	20	19	20	20	19	15	18	19	12	17	20	

Note: 1.1 Context/setting description; 2.1 Participant description, 2.2 Participant disability/at-risk status; 3.1 Role and description; 3.2 Training description; 4.1 Intervention procedures; 4.2 Materials description; 5.1 Implementation fidelity assessed/reported; 5.2 Fidelity dosage or exposure assessed/reported; 5.3; Fidelity assessed across relevant elements/throughout study; 6.1 Independent variable (IV) systematically manipulated; 6.2 Baseline description; 6.3 No or limited access to IV during baseline; 6.4 Design provides at least demonstrations of experimental effect at three different times; 6.5 Baseline phase contains at least three data points; 6.6 Design controls for common threats to internal validity; 7.1 Socially important goals; 7.2 Description of dependent variable measures; 7.3 Reports effects on the intervention on all measures; 7.4 Minimum of three data points per phase; 7.5 Adequate interobserver agreement; 8.1 Study provide single-case graph clearly representing outcome data across all study phases.

3.3 Statistical Analysis

Effects of studies. Effect sizes were calculated for each AB contrast. Several studies used a multiple baseline design across three behaviors so those participants would have three AB contrasts. Effect sizes were then averaged for each study and appear in Table 4. Overall omnibus effect sizes for each type were as follows: SMD (mean = 2.407, SD = 1.272, range =

2.038 – 3.867); IRD (mean = .805; SD = 0.278, range = 0 – 1); and Tau, mean = .648 , SD = .703, (range = -1 – 3.09).

Table 4. Mean Study Effect Sizes

Study	Effect Sizes					
	SMD		IRD		Tau-*U*	
	Mean	SD	Mean	SD	Mean	SD
1. Agan et al. (2005)	2.914	0.886	1	0	1.00	0.357
2. Amato-Zech et al. (2006)	3.343	0.361	1	0	1	0
3. Bedesm (2012)	2.699	0.764	.92	0	.95	0.035
4. Briere & Simonse (2011)	2.917	0.829	1	0	-1	0
5. Coughlin et al. (2012)	1.393	0.716	.56	0.177	-.70	0.250
6. Creel et al. (2006)	3.861	0	1	0	1.1	0.192
7. Crutchfield et al. (2015)	1.54	0.390	.78	0.035	-.86	0.120
8. Davis et al. (2014)	0.495	0.401	.18	0.318	.16	0.363
9. Graham-Day (2010	1.595	0.789	.68	0.19	.74	0.352
10. Gulchak (2008)[a]	1.356	------	.59	------	.55	------
11. Gumple (2007)	0.631	0.608	.36	0.227	.22	0.404
12. Gureasko-Moore (2007)	1.757	0.997	.83	0.253	.76	0.225
13. Harris et al. (2005)	2.548	0.842	.92	0.092	1.06	0.213
14. Kartal	2.459	1.445	.83	0.252	.83	0.512
15. Rafferty (2011)	3.861	0	1	0	.84	0.173
16. Rafferty (2012)	2.73	1.38	.81	0.401	.94	0.263
17. Rafferty & Raimondi (2009)	2.471	2.354	.78	0.202	0.94	0.123
18. Rock (2005)	3.32	0.837	.98	.98	.55	1.34
19. Rock & Thread (2007)	3.38	0.705	.94	0.085	.97	0.138
20. Wills (2014)	2.16	1.698	.68	0.313	.47	0.932

[a]only one participant

Results of independent samples t-tests with equal variances not assumed displayed were calculated first between age for young (n = 53, <13) and old (n = 19, >13) for each effect size. Statistically significant differences were obtained for SMD (t = 2.475, *p* = .008) and IRD (t = 1.962, *p* = .03), but not for Tau-*U* (t = 1.33, *p* = 09). In terms of differences between on-task (n = 57) and off task (n = 16) behavior, only Tau-*U* (t = 7.602, *p* = 001) was significant with SMD (t = 0.552, *p* = .29) and IRD (t = 0.631, *p* = .27) being insignificant. Finally, the differences between the at-risk and ADHD (n = 44) and comorbid (n = 12) groups were significant for SMD (t = 2.113, *p* = .02) and IRD (t = 1.919, *p* = .03), but not for Tau-*U* (t = 1.372, *p* = .09).

Publication bias. To address the "file drawer effect," the number of studies with results of zero required to reduce the overall effect to insignificant or suspect levels was determined for SMD, IRD and Tau effect sizes. It would take an additional 1,139 cases with an average effect size of 0 to bring the overall SMD to the ineffective range (<.02), 111 cases with an average effect size of 0 to bring the overall IRD into the ineffective range (<.36), and 211 cases with an average effect size of 0 to overall Tau into the small to ineffective range (<.20). The average number of null effect sizes between IRD and Tau was 161. SMD was not included in the average because of overestimation problems addressed previously. Further, the <.02 range as ineffective was determined for Cohen's *d* with group designs which routinely have much smaller effect sizes than those obtained with SMD, even with the 3rd percentile ceiling. Most SCRD studies contain between one and six participants. Using the average of three participants per study, at least 54 unpublished SCRD studies will null effects would be needed to reduce currently obtained effect sizes to small or questionable levels.

4. Discussion

The purpose of this systematic review was threefold: (1) determine the characteristics of the identified literature on self-monitoring behavior, (2) conduct a meta-analysis of the included studies, and (3) determine their methodological quality. Each of these areas are discussed below as well as limitations and implications for practice.

4.1 Characteristics of Studies

This systematic review was interested in self-monitoring behavior (i.e., non-academic content) for children and adolescents in school settings. Participants were predominately Caucasian males who were either identified as at-risk or diagnosed with ADHD. It should not come as a surprise that students with ADHD made up the majority of participants for two reasons. First, the most common dependent variables for self-monitoring behavior are those that make up the response classes of on-task and off-task. Children with ADHD have been consistently found to display characteristics of inattention and another 5% of children not meeting diagnostic criteria (i.e., at-risk) display behaviors indicative of inattention (Sayal et al., 2018). Second, self-monitoring is considered a major technique for improving the self-regulation skills of children with ADHD and those with sluggish cognitive tempo (Capdevila-Brophy et al., 2014; Reid et al., 2005). Relatedly, most of the interventions were self-monitoring attention.

Most previous reviews on self-monitoring behavior included studies that used self-management techniques (e.g., self-evaluation, self-reinforcement) in addition to self-monitoring, and that included an external positive reinforcement component (Briesch & Chafouleas, 2009; Bruhn et al., 2015; Reid et al., 2005; Sheffield & Waller, 2010; Webber et al., 1993). Some individual studies would empirically compare the effectiveness of self-monitoring in isolation to self-monitoring plus positive reinforcement (e.g., Gumpel, 2007) while many studies excluded from the present review would positively reinforce students for either self-monitoring accurately, obtaining a certain number of answers correct,

or both (e.g., Bedesem & Dieker, 2014; Coogan, Kehle, Bray, & Chafouleas, 2007; Vogelgesang, Bruhn, Coghill-Behrends, Kern, & Troughton, 2016). Clearly, the application of various reinforcement techniques (e.g., verbal praise, differential reinforcement) would add to the effectiveness of self-monitoring. However, the issue is that self-monitoring is part of self-management with the goals to give teachers more time to teach than managing students' behaviors, promoting maintenance and generalization, and empowering students (Maag, 2018). Why then would a teacher want to add external components which would, seemingly, contradict the point of using self-monitoring? Further, it has long been demonstrated that self-monitoring produces reactivity (e.g., Mace & Kratochwill, 1985, 1988) in which behavior changes simply and solely by engaging in the self-monitoring process. Finally, it seems redundant and unnecessary for positively reinforcing students for engaging in the self-monitoring process since accuracy is not a prerequisite for reactivity to occur. For example, Bruce, Lloyd, & Kennedy (2012) found meaningful improvements in targeted behaviors when self-monitoring accuracy was quite low.

4.2 Effectiveness of Self-Monitoring

According to PRISMA guidelines (Liberati et al., 2009), it is important for systematic reviews to calculate effect sizes in order to quantify results of interventions. In group design studies, the typical effect size calculations would be either Cohen's d or Hedges g. However, in SCRD studies, there is no clear cut preferable effect size calculation since all have their disadvantages (Parker et al., 2009). Nevertheless, some are more appropriate given the type of data obtained (e.g., expectant increases versus decreases, variability of baseline and intervention phase trends). In the present systematic review SMD, IRD, and Tau-U were calculated and the rationale for using these three were provided.

Only one review on self-monitoring behavior calculated effect sizes—Briesch and Chafouleas (2009) used percentage of nonoverlapping data (PND) and what they called Cohen's d although it is the SCRD equivalent typically referred to as standard mean difference (SMD). In the latter case, there were two problems. First, they used baseline standard deviation in their SMD calculation which is acceptable, although it is arguably more desirable statistically to use the pooled standard deviation (Busk & Serlin, 1992). Furthermore, SMD is, in and of itself, considered unreliable in SCRD studies due to small number of observations and floor or ceiling effects limiting variability (Scruggs & Mastopieri, 2012) and may result in an over-inflation of the parametric effect sizes (Horner et al., 2012). Briesch and Chafouleas did not account for these problems as in the present analysis and that used previously (e.g., Losinski et al., 2014) in which a ceiling was set of the number at the third quartile which would then be the maximum value of d. They found the SMD to be 4.11 which twice as high as that obtained in the current meta-analysis ($d = 2.40$) and that found by Fantuzzo and Polite (1990; [$d = 2.30$]). In terms of PDN, they obtained a mean effect size of 76.30% which is considered a moderate effect. However, there are several known disadvantages of PND including, but not limited to, lack of a known underlying distribution, low ability to discriminate among published studies, low statistical power for small N studies, and open to human error in hand calculations from graphs (Parker et al., 2009). Finally, over 50% of their studies targeted "on-task" which, regardless of how defined, would result in

increases in the target behavior during intervention. That makes it possible that monotonic trends (i.e., increasing trend during baseline) may negatively impact the ability to evaluate the relevance of obtained effect sizes. The proper effect size to use in those cases would be Tau-U that controls for these trends.

An important aspect of the current meta-analysis was that studies that contained contingent reinforcement for either engaging in the self-monitoring process accurately or for improving the quality and quantity of dependent variables were excluded from this review. The reason was to obtain a "true" measure of the effectiveness of self-monitoring. The obtained effect sizes in the "effective" range indicate that forms of positive reinforcement are not necessary to obtain positive results. This finding truly meets the goals of self-management training in general (Maag, 2018).

Results of independent samples t-tests were equivocal. Participants in the young group obtained statistically higher effect sizes than the ones in the old group. However, there were almost three times as many participants in the young group as in the older group. Self-monitoring is typically used for students in elementary grades because of the frequent fluency practice on basic mathematics, reading, and spelling skills whereas older students are more likely to use self-monitoring of strategies for more complex skills (Maag, 2018; Maag et al., 1993). The only significant difference between the effectiveness for on-task versus off-task behaviors was for Tau-U—which makes sense considering this effect size is the only one that corrects for monotonic (i.e., increasing baseline) trends and, consequently, provides a more accurate estimate of true effects when increases in dependent variables are expected. Finally, self-monitoring was more effective for participants with ADHD or at-risk than those with comorbid conditions. This result is also expected for two reasons. First, self-monitoring is used to increase on-task types of behaviors including academic responding—problems typically encountered by students with ADHD or those at-risk who display inattentive behaviors more than their otherwise typically performing peers (Sayal et al., 2018). Second, students with co-morbid conditions, by definition, have more severe involvement that may make understanding and using self-monitoring procedures more problematic.

4.3 Quality of Studies

Only one previous review (Maggin et al., 2013) examined study quality but reviewed studies that had a total of 11 self-management components (as compared to only self-monitoring) and also used the now-dated Horner et al. (2005) QIs rather than the CEC (2014) QIs composed of 22 items as used in the present systematic review. None of the studies in the current review met all 22 items from the eight QIs, and seven (35%) had fewer than three participants (Bedesem, 2012; Briere & Simonsen, 2011; Crutchfield et al., 2015; Davis et al., 2014; Gulchak, 2008; Rafferty & Raimondi, 2009; Wills & Mason, 2014) and, consequently, could not be considered having positive results. However, eight studies met 21 items and the average across studies was relatively high with a mean of 20. Further, all studies met at least one item from all eight quality indicators. Nevertheless, this collective body of literature fell short in terms of participants across studies using absolute coding. However, 19 out of the 20

studies met 80% or more of the 22 items in the eight QIs which is commensurate with guidelines recommended by Lane, Bruhn, Crnobori, and Sewell (2009).

An interesting finding was the extremely low number of studies (N=7) that met QI 5.3 regarding fidelity: *assess and report throughout study*. The two other fidelity items, 5.1 and 5.2 were met by 20 and 19 studies, respectively. The low number of item 5.3 was expected for self-monitoring. Once students learn how to self-monitor, by definition, there typically is no additional external intervention agent. Fidelity at that point becomes moot because engaging in self-monitoring produces reactivity regardless of how accurate students follow the process (Bruce et al., 2012). It also explains why fidelity was so high on the first two items. The steps in teaching a student self-monitoring are straightforward and typically requires only one session: Describe the process to the student, model self-monitoring while the student determines if you are accurate or not, student practices independently, and if feedback is required then the independent practice is repeated (Maag, 2018). Self-monitoring can be learned in one session since it is then practiced regularly as part of the intervention under study. Sometimes brief "booster" sessions (e.g., 1 or 2 minutes in length) are used, but often are not necessary. Hence, another advantage of self-monitoring is that the intervention can be taught (i.e., implemented) with high fidelity.

4.4 Limitations

There are several limitations to the present study. First, the obvious limitation is that only studies that used self-monitoring in isolation were included in the review. Additional components such as positive and differential reinforcement, specifically programming self-evaluation, function-based supports, and adult feedback have been found to improve the effectiveness of self-management (e.g., Bruhn et al., 2015; Sheffield & Waller, 2010). However, the purpose of the present review was to determine the effectiveness of self-monitoring in isolation because that represents the purposes of self-management—a student implemented rather than teacher implemented intervention.

A second limitation was the decision to use the absolute coding approach (0=not met, 1=met) versus the weighted method that gives partial credit (0=not met, .5=partially met, 1=met). The rationale for using this approach was provided previously. Nevertheless, it does represent a limitation because giving studies partial credit helps answer the question as to whether a certain intervention for a particular outcome is effective and thereby avoiding Type II (false negative) errors (Kettler & Lane, 2017). Relatedly, a weighted coding scheme would be helpful in evaluating research conducted before quality indicators were published. That limitation was mitigated in the present review by only including studies published in 2005 or later.

Third, the decision to omit studies published prior to 2005 does limit the scope of conclusions drawn for the synthesis. In addition, studies were only included that appeared in refereed journals, thus excluding those that may appear in book chapters and dissertations. The decision to omit potential studies from book chapters and dissertations was that although they may be of high quality, they did not go through the thorough vetting used in the peer review process. This decision is nevertheless problematic because Pigott, Valentine, Polanin,

Williams, and Canada (2013) found that approximately one-third of null findings come from dissertations. However, in the present review, publication bias was addressed by adding enough cases with null findings to the obtained effect sizes which provides context for how many participants in unpublished SCRD studies would be needed to bring effect sizes down to insignificant levels. Further, as Cooper and Hedges (2009) pointed out, it is easier to correct for publication bias than to detect it with any level of certainty.

Even considering these limitation, self-monitoring is still an effective intervention for improving students attending to task and reducing their off-task behaviors. Further, it is simple for teachers to train students to self-monitor and there are a variety of technological devices (e.g., MotivAider®, CellF-Monitoring, VIBRALITE 3 wrist watch, Palm Zire 72 handheld) that have greatly advanced the ability of cuing students to self-monitor that the old "beep tape" played on a cassette recorder. Self-monitoring also is a socially valid intervention for two reasons. First, it is easy for teachers to learn and train students. Second, after training, the intervention is turned over to the students with minimal to no teacher involvement which frees up instructional time and reduced the need to redirect students.

5. Conclusion

Self-monitoring behavior for children and adolescents in school typically focuses on increasing their on-task behavior while decreasing their off-task behavior. It consists of three parts: self-observation, self-recording, and self-graphing. There have been five previous reviews of self-monitoring but only two focused on self-monitoring to improve students' classroom behaviors (Bruhn et al., 2015; Sheffield & Waller, 2010). However these reviews included other aspects of self-management and external environmental interventions such as peer tutoring and contingent teacher attention. They also did not assess the quality of studies in their reviews. The current review focused solely on self-monitoring as an isolated intervention and only included studies in which reinforcement was not provided for accurately self-monitoring or increasing the amount of work completed. The studies reviewed met 80% or more of the 22 quality indicators which is commensurate with previously reviews and indicates that self-monitoring an emereging evidence-based practice. The current review used a quantitative approach by calculating effect sizes that was absent from the two previous reviews. They indicated thateffectiveness of self-monitoring interventions ranged between large and very large.

References

*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.

*Amato-Zech, N. A., Hoff, K. E., & Doepke, K. J. (2006). Increasing on-task behavior in the classroom: Extension of self-monitoring strategies. *Psychology in the Schools, 43,*

211-221. https://doi.org/10.1002/pits.20137

Bedesem, P. L. (2012). Using cell phone technology for self-monitoring procedures in inclusive settings. *Journal of Special Education Technology, 27,* 33-46. https://doi.org/10.1177/016264341202700403

Bedesem, P. L., & Dieker, L. A. (2014). Self-monitoring with a twist: Using cellphones to CellF-Monitor on-task behavior. *Journal of Positive Behavior Interventions*, *16*, 246-254. https://doi.org/10.1177/1098300713492857

*Briere, D. E., III, & Simonsen, B. (2011). Self-monitoring interventions for at-risk middle school students: The importance of considering function. *Behavioral Disorders, 36,* 129-140.

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. https://doi.org/10.1037/a0016159

Bruce, A., Lloyd, J. W., & Kennedy, M. J. (2012). Targets of self-monitoring: Productivity, accuracy, and attention. In B. G. Cook, M. Tankersley, & T. J. Landrum (Eds.), *Advances in learning and behavioral disabilities: Classroom behavior, contexts, and interventions* (pp. 1-21). Bingley UK: Emerald.

Bruhn, A., McDaniel, S., & Kreigh, C. (2015). Self-monitoring interventions for students with behavior problems: A systematic review of current research. *Behavioral Disorders*, *40*, 102-121. https://doi.org/10.17988/bd-13-45-1

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research designs and analysis: New directions for psychology and education* (pp. 187-212). Hillsdale, NJ: Lawrence Erlbaum.

Capdevila-Brophy, C., Artigas-Pallarés, J., Navarro-Pastor, J. B., Garcia-Nonell, K., Rigau-Ratera, E., & Obiols, J. E. (2014). ADHD predominantly inattentive subtype with high sluggish cognitive tempo: A new clinical entity? *Journal of Attention Disorders*, *18*, 607-616. https://doi.org/10.1177/1087054712445483

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

Coogan, B. A., Kehle, T. J., Bray, M. A., & Chafouleas, S. M. (2007). Group contingencies, randomization of reinforcers, and criteria for reinforcement, self-monitoring, and peer feedback on reducing inappropriate classroom behavior. *School Psychology Quarterly*, *22*, 540-556. https://doi.org/10.1037/1045-3830.22.4.540

Cook, B. G., & Tankersley, M. (2007). A preliminary examination to identify the presence of quality indicators in experimental research in special education. In J. Crockett, M. M.

Gerber, & T. J. Landrum (Eds.), *Achieving the radical reform of special education: Essays in honor of James M. Kauffman* (pp. 189-212). Mahwah, NJ: Lawrence Erlbaum.

Cooper, H., & Hedges, L. V. (2009). Potentials and limitations. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 561-572). New York, NY: Russell Sage Foundation.

*Coughlin, J., McCoy, K. M., & Kenzer, A. (2012). Effects of a self-monitoring strategy on independent behavior of students with mild intellectual disability. *Education and Training in Autism and Developmental Disabilities, 47*, 154-164.

Council for Exceptional Children. (2014). Council for Exceptional Children standards for evidence-based practices in special education. *Exceptional Children*, *80*, 504-511. https://doi.org/10.1177/0040059914531389

*Creel, C., Fore, C., Boon, R. T., & Bender, W. N. (2006). Effects of self-monitoring on classroom preparedness skills of middle school students with attention deficit hyperactivity disorder. *Learning Disabilities: A Multidisciplinary Journal, 14,* 105-113.

*Crutchfield, S., Mason, R., Chambers, A., Wills, H., & Mason, B. (2015). Use of a self-monitoring application to reduce stereotypic behavior in adolescents with autism: A preliminary investigation of I-Connect. *Journal of Autism & Developmental Disorders, 45*, 1146-1155. https://doi.org/10.1007/s10803-014-2272-x

*Davis, T. N., Dacus, S., & Bankhead, J. (2014). A comparison of self-monitoring with and without reinforcement to improve on-task classroom behavior. *Journal of School Counseling, 12*(12). Retrieved from http://www.jsc.montana.edu/articles/v12n12.pdf

DiGangi, S. A., Maag, J. W., & Rutherford, R. B., Jr. (1991). Self-graphing of on-task behavior: Enhancing reactive effects of self-monitoring on on-task behavior and academic performance. *Learning Disability Quarterly*, *14*, 221-230. https://doi.org/10.2307/1510851

Fantuzzo, J. W., & Polite, K. (1990). School-based behavioral self-management: A review and analysis. *School Psychology Quarterly*, *5*, 180-198. https://doi.org/10.1037/h0090612

Gage, N. A., Lewis, T. J., & Stichter, J. P. (2012). Functional behavioral assessment-based interventions for students with or at-risk for emotional and/or behavioral disorders in school: A hierarchical linear modeling meta-analysis. *Behavioral Disorders*, *37*, 55-78. https://doi.org/10.1177/019874291203700201

*Graham-Day, K., Gardner, R., III, & Hsin, Y. (2010). Increasing on-task behaviors of high school students with attention deficit hyperactivity disorder: Is it enough? *Education & Treatment of Children, 33,* 205-221. https://doi.org/10.1353/etc.0.0096

*Gulchak, D. J. (2008). Using a mobile handheld computer to teach a student with an emotional and behavioral disorder to self-monitor attention. *Education & Treatment of Children*, *31,* 567-581. https://doi.org/10.1353/etc.0.0028

*Gumpel, T. P. (2007). Are social competence difficulties caused by performance or acquisition deficits? The importance of self-regulatory mechanisms. *Psychology in the Schools, 44,* 351-372. https://doi.org/10.1002/pits.20229

*Gureasko-Moore, S., DuPaul, G. J., & White, G. P. (2007). Self-management of classroom preparedness and homework: Effects on school functioning of adolescents with attention deficit hyperactivity disorder. *School Psychology Review, 36*, 647-664.

Guzman, G., Goldberg, T. S., & Swanson, H. L. (2018). A meta-analysis of self-monitoring on reading performance of K-12 students. *School Psychology Quarterly*, *33*, 160-168. https://doi.org/10.1037/spq0000199

Hallahan, D. P., Lloyd, J. W., & Stoller, L. (1982). *Improving attention with self-monitoring*. Charlottesville: University of Virginia Learning Disabilities Research Institute.

*Harris, K. R., Friedlander, B. D., Saddler, B., Frizzelle, R., & Graham, S. (2005). Monitoring of attention versus self-monitoring of academic performance: Effects among students with ADHD in the general education classroom. *The Journal of Special Education, 39*, 145-156.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*, 165-179. https://doi.org/10.1177/00140290507100203

Horner, R. H., Swaminathan, H., Sugai, G., & Smolkowski, K. (2012). Considerations for the systematic analysis and use of single-case research. *Education and Treatment of Children, 35*, 269-290. https://doi.org/10.1353/etc.2021.0011

Joseph, L. M., & Eveleigh, E. L. (2011). A review of the effects of self-monitoring on reading performance of students with disabilities. *The Journal of Special Education*, *45*, 43-53. https://doi.org/10.1111/l/drp.12158

Kanfer, F. H. (1970). Self-monitoring: Methodological limitations and clinical applications. *Journal of Consulting and Clinical Psychology*, *35*, 148-152. https://doi.org/10.1037/h0029874

*Kartal, M. S., & Ozkan, S. Y. (2015). Effects of class-wide self-monitoring on on-task behaviors of preschoolers with developmental disabilities. *Education and Training in Autism and Developmental Disabilities, 50*, 418-432.

Kettler, R. J., & Lane, L. (2017). *Methodological foundations of school psychology (MFSP) research and practice*. Manuscript in preparation.

Klein, R. D. (1979). Modifying academic performance in the grade school classroom. In M. Hersen, R. M. Eisler, & P. M. Miller (Eds.), *Progress in behavior modification* (Vol. 8, pp. 293-321). New York, NY: Plenum.

Lane, K. L., Bruhn, A. L., Crnobori, M. L., & Sewell, A. L. (2009). Designing functional assessment-based intervention using a systematic approach: A promising practice for

supporting challenging behavior. In T. E. Scruggs & M. A., Mastropieri (Eds.), *Advances in learning and behavior disabilities: Policy and practice* (Vol. 22, pp. 341-370). Bingley, UK: Emerald.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P. A., . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine, 6*(7), e1000100. https://doi.org/10.1371/journal.pmed.1000100

Losinski, M., Maag, J. W., Katsiyannis, A., & Parks-Ennis, R. (2014). Examining the effects and quality of interventions based on the assessment of contextual variables: A meta-analysis. *Exceptional Children*, *80*, 407-422. https://doi.org/10.1177/0014402914527243

Maag, J. W. (2018). *Behavior management: From theoretical implications to practical applications* (3rd ed.). Boston, MA: Cengage.

Maag, J. W., Reid, R., & DiGangi, S. A. (1993). Differential effects of self-monitoring attention, accuracy, and productivity. *Journal of Applied Behavior Analysis*, *26*, 329-344. https://doi.org/10.1901/jaba.1993.26-329

Mace, F. C., & Kratochwill, T. R. (1985). Theories of reactivity in self-monitoring: A comparison of cognitive-behavioral and operant models. *Behavior Modification*, *9*, 323-343. https://doi.org/10.1177/01454455850093003

Mace, F. C., & Kratochwill, T. R. (1988). Self-monitoring. In J. C. Witt, S. N. Elliott, & F. M. Gresham (Eds.), *Handbook of behavior therapy in education* (pp. 489-522). New York, NY: Plenum.

Machalicek, W., O'Reilly, M. F., Beretvas, N., Sigafoos, J., & Lancioni, G. (2007). A review of interventions strategies to decrease challenging behavior in school settings for students with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *1*, 229-246. https://doi.org/10.1016/j.rasd.2006.10.005

Maggin, D. M., Briesch, A. M., & Chafouleas, S. M. (2013). An application of the What Works Clearinghouse standards for evaluating single-subject research: Synthesis of the self-management literature base. *Remedial and Special Education*, *34*, 44-58. https://doi.org/10.1177/0741932511435176

Mitchell, M. (2002). Engauge Digitizer (version 4.1) [computer software]. Retrieved from http://sourceforge.net/projects/digitizer/

Nagler, E. M., Rindskopf, D. M., & Shadish, W. R. (2008). *Analyzing data from small N designs using multilevel models: A procedural handbook*. Washington, DC: U. S. Department of Education.

Orwin, R. G. (1983). A fail-safe *N* for effect in meta-analysis. *Journal of Educational Statistics*, *8*, 157-159. https://doi.org/10.2307/1164923

Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single-case research. *Exceptional Children*, *75*, 135-150. https://doi.org/10.1177/001440290907500201

Pigott, T. D., Valentine, J. C., Polanin, J. R., Williams, R. T., & Canada, D. D. (2013). Outcome-reporting bias in education research. *Educational Researcher*, *42*, 424-432. https://doi.org/10.3102/0013189x13507104

Rafferty, L. A. (2012). Self-monitoring during whole group reading instruction: Effects among students with emotional and behavioral disabilities during summer school intervention sessions. *Emotional & Behavioural Difficulties, 17,* 157-173. https://doi.org/10.1080/13632752.2012.672866

*Rafferty, L. A., Arroyo, J., Ginnane, S., & Wilczynski, K. (2011). Self-monitoring during spelling practice: Effects on spelling accuracy and on-task behavior of three students diagnosed with attention deficit hyperactivity disorder. *Behavior Analysis in Practice, 4*, 37-45. https://doi.org/10.1007/bf03391773

*Rafferty, L. A., & Raimondi, S. L. (2009). Self-monitoring of attention versus self-monitoring of performance: Examining the differential effects among students with emotional disturbance engaged in independent math practice. *Journal of Behavioral Education, 18,* 279-299. https://doi.org/10.1007/s10864-009-9092-7

Reid, R. (1996). Research in self-monitoring with students with learning disabilities: The present, the prospects, the pitfalls. *Journal of Learning Disabilities*, *29*, 317-331. https://doi.org/10.1177/002221949602900311

Reid, R., Trout, A. L., & Schartz, M. (2005). Self-regulation interventions for children with attention deficit-hyperactivity disorder. *Exceptional Children*, *71*, 361-377.

*Rock, M. L. (2005). Use of strategic self-monitoring to enhance academic engagement, productivity, and accuracy of students with and without exceptionalities. *Journal of Positive Behavior Interventions*, *7*, 3-17. https://doi.org/10.1177/10983007050070010201

*Rock, M. L., & Thread, B. K. (2007). The effects of fading a strategic self-monitoring intervention on students' academic engagement, accuracy, and productivity. *Journal of Behavioral Education, 16,* 389-412. https://doi.org/10.1007/s10864-007-9049-7

Rosenberg, M. S., Adams, D. C., & Gurevitch, J. (2000). *Metawin*. Sunderland, MA: Sinauer.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*, 638-641. https://doi.org/10.1037/0033-2909.86.3.638

Sayal, K., Prasad, V., Daley, D., Ford, T., & Coghill, D. (2018). ADHD in children and young people: Prevalence, care pathways, and service provision. *The Lancet Psychiatry*, *5*, 175-186. https://doi.org/10.1016s/2215-0366(17)30167-0

Scruggs, T. E., & Mastropieri, M. A. (2012). PND at 25: Past, present, and future trends in summarizing single-subject research. *Remedial and Special Education*, *34*, 9-19.

https://doi.org/10.1177/0741932512440730

Sheffield, K., & Waller, R. J. (2010). A review of single-case studies utilizing self-monitoring interventions to reduce problem classroom behaviors. *Beyond Behavior, 19*(2), 7-13.

Skinner, B. F. (1953). *Science and human behavior*. New York, NY: Macmillan.

Snider, V. (1987). Use of self-monitoring of attention with LD students: Research and application. *Learning Disabilities Quarterly*, *10*, 139-151. https://doi.org/10.2307/1510221

Vogelgesang, K. L., Bruhn, A. L., Coghill-Behrends, W. L., Kern, A. M., & Troughton, L. C. W. (2016). A single-subject study of a technology-based self-monitoring intervention. *Journal of Behavioral Education*, *25*, 478-497. https://doi.org/10.17077/etd.zk6r9uac

Webber, J., Scheuermann, B., McCall, C., & Coleman, M. (1993). Research on self-monitoring as a behavior management technique in special education classrooms: A descriptive review. *Remedial and Special Education*, *14*, 38-56. https://doi.org/10.1177/074193259301400206

*Wills, H., & Mason, B. (2014) Implementation of a self-monitoring application to improve on-task behavior: A high-school pilot study. *Journal of Behavioral Education, 23,* 421-434. https://doi.org/10.1007/s10864-014-9204-x

Zegiob, L., Klukas, N., & Junginger, J. (1978). Reactivity of self-monitoring procedures with retarded adolescents. *American Journal of Mental Deficiency*, *83*, 156-163.

*Studies included in the current review.