Review Journal of Autism and Developmental Disorders
https://doi.org/10.1007/s40489-022-00304-y

REVIEW PAPER

Investigating the Effects of Self-Monitoring Interventions with Students with Disabilities on the Maintenance and Generalization of On-Task Behavior: A Systematic Literature Review

Mehmet D. Sulu[1] · Ronald C. Martella[1] · Kharon Grimmet[1] · Amanda M. Borosh[1] · Emine Erden[2]

Received: 23 September 2021 / Accepted: 6 January 2022

Abstract

Although there is a consensus regarding the positive effects of self-monitoring interventions on improving on-task behaviors of students with disabilities, the findings for maintenance and generalization have not been shown to be consistent across studies (i.e., Cook & Sayeski, 2020; Wood et al., 2002). The current study aimed to assess the research analyzing the effects of self-management on students' on-task behavior using the What Works Clearinghouse design standards (Kratochwill et al., 2013) and the resulting maintenance and generalization of on-task behaviors (if any) in specific using standards developed by Neely et al. (2016) and Neely et al. (2018). The findings indicate that programming for maintenance and generalization of on-task behavior in self-monitoring interventions are generally poor in quality and quantity.

Keywords Self-monitoring · Maintenance · Generalization · Students with disabilities · On-task behaviors

The inclusion of students with disabilities into general education settings is an important goal (Individuals with Disabilities Education Improvement Act [IDEIA], 2004; Turnbull et al., 2019). For students with disabilities to be successful in these settings, they need to exhibit academic behaviors that are expected by teachers (Rojewski et al., 2015) such as on-task behavior. Research indicates that a barrier to students exhibiting such academic behaviors include the following: difficulty responding consistently to natural cues to perform a given task, relying on external cues such as teachers and instructional assistants, and failing to maintain previously learned skills over time (Koyama & Wang, 2011; Phillips & Vollmer, 2012). When students with disabilities rely on external cues, this necessitates the continuous prompting and reinforcement from supporting adults such as instructional assistants and teachers in general education settings (Agran et al., 2003; Yucesoy Ozkan & Sonmez, 2011). As a result, students report feelings of helplessness, embarrassment, rejection, and stigmatization when describing their perceived levels of support from instructional assistants (Broer et al., 2005; Richards et al., 2015).

One method of improving students' successful academic behaviors is teaching self-management skills. Self-management has been shown to be a useful approach to improving the on-task behavior and, in turn, the academic engagement behavior of students with disabilities (Briesch & Daniels, 2013; Dalton et al., 1999; McDaugall et al., 2017). Self-management involves several behaviors such as self-evaluation, self-charting, goal setting, self-recording, and self-monitoring (Martella et al., 2012). Self-recording/monitoring interventions, in particular, have been promoted as maintenance and generalization programming approaches and have become prevalent among practitioners (Agran et al., 2003; Aykut, 2020; Maggin et al., 2012; Rosenbaum and Drabman, 1979; Sheffield & Waller, 2010).

Students observe and record occurrence and nonoccurrence of their own behaviors with self-recording and self-monitoring procedures (Cooper et al., 2020). Thus, these two procedures have been used interchangeably by many researchers and practitioners. However, Martella et al. (1993) made a distinction between self-recording and self-monitoring based on use of external prompts. According to their definition, if a student observes and records his or

✉ Mehmet D. Sulu
msulu@purdue.edu

[1] Department of Educational Studies, College of Education, Purdue University, Beering Hall of Liberal Arts and Education, Room 5153, 100 N, University Street, West Lafayette, IN 47907, USA

[2] Faculty of Education, Hasan Kalyoncu University, Gaziantep, Turkey

her own behavior with external prompts, self-recording procedures are in use (Martella et al., 2012). When external prompts are not provided for the student to record his or her behavior, then, the student performs self-monitoring behaviors (Dalton et al., 1999). Therefore, self-monitoring has been identified as more difficult for students with disabilities to use due to lack of external prompts (Dalton et al., 1999; Martella et al., 2012). (Note: Given the similarities of self-recording and self-monitoring, and the fact that researchers use these terms interchangeably, the term self-monitoring will be used for the remainder of this paper.)

Self-monitoring has several benefits. First, self-monitoring has reactivity effects (Cooper et al., 2020) and is obtrusive for undesired behaviors (Cooper et al., 2020) which may result in changes in behavior (Wood et al., 2002). Additionally, self-monitoring is versatile and can be combined with other intervention procedures such as token economies, graphing, and video modeling (Cooper et al., 2020; Knapczyk & Livingston, 1973; Riden et al., 2020). Finally, self-monitoring interventions are advantageous because they can be implemented in a variety of contexts and do not require students to be removed from general education settings (Agran et al., 2003; Cook et al., 2015). Thus, self-monitoring can aid in the inclusion of students with disabilities in general education classrooms.

Although outcomes of self-monitoring interventions are promising for enhancing on-task behaviors and academic engagement of students with disabilities, there is a lack of research supporting the maintenance and generalization of these interventions in students with disabilities (Briesch et al., 2019; Cook & Sayeski, 2020; Dalton et al., 1999; Kolbenschlang & Wunderlich, 2019). This lack of research dates back at least 30 years. For example, Hughes et al. (1991) conducted a review of the self-management research literature on the assessment of maintenance and generalization of behaviors. Of 19 studies reviewed, self-monitoring interventions were used in 11 studies. Among the 11 studies, 3 studies (i.e., Anderson-Inman et al., 1984; Howell et al., 1983; Knapczyk & Livingston, 1973) assessed for only generalization, 1 study (i.e., Albion & Salzberg, 1982) assessed for maintenance, and only 1 study (Szykula et al., 1981) investigated both maintenance and generalization (Hughes et al., 1991). The authors concluded that the majority of self-monitoring interventions did not assess for maintenance and generalization; however, the limited number of investigations suggest that self-monitoring interventions appeared to promote generalization of learned behaviors (Hughes et. al., 1991).

Despite decades of research on self-management in general and self-monitoring, specifically, assessing for maintenance and generalized responding still appears to be limited. Maintenance and generalization of self-monitoring interventions have been neglected in research, with most studies only taking place in school settings without implementation across individuals and behaviors (i.e., Briesch & Chafouleas, 2009; Briesch et al., 2019; Bruhn et al., 2015; Busacca et al., 2015; Yucesoy Ozkan & Sonmez, 2011). In addition, the quality of collected data of these procedures has been shown to be weak in quality (Bruhn et al., 2015; Wood et al., 2002).

Assessing the maintenance of behavior change in self-monitoring studies seems especially important given that demonstrated maintenance across time helps to show that sustainable behavioral change has taken place (Strokes & Osnes, 1989); likewise, the demonstration of the generalization of learned behavior to untrained stimuli that resemble the trained stimulus (Baer et al., 1968; Cooper et al., 2020; Stokes & Baer, 1977; Stokes & Osnes, 1989) also provides evidence of the durability of the behavior change.

One of the important considerations of the maintenance and generalization programming includes the determination of latency to maintenance. Although there is no specific guidance for latency to maintenance in the current literature (Neely et al., 2018), the amount of time a newly acquired behavior needs to be maintained is based on the importance of that behavior for the individual's life (Cooper et al., 2020). Since on-task behaviors are essential skills that students with disabilities need throughout their life in a variety of educational settings (Kartal & Yucesoy Ozkan, 2015), maintenance and generalization assessments for on-task behaviors should occur over a longer period of time (Dalton et al., 1999) and across a range of educational environments such as people (i.e., science teacher versus English teacher), task or activity (i.e., science class versus math class), setting (i.e., special education classroom versus general education classrooms), or condition/context with specific teaching strategies (i.e., gradual fading procedures) (Neely et al., 2016, 2018).

Therefore, given that the most commonly used self-management practices among researchers and practitioners are self-monitoring interventions (Cooper et al., 2020; Hughes et al., 1991), this review of self-monitoring interventions seeks to assess the quality and length of maintenance and quality and dimension of generalization of self-monitoring interventions focused on teaching on-task behaviors to students with disabilities. The specific research questions were as follows:

1. What is the overall quality of self-monitoring intervention studies based on What Works Clearinghouse (WWC) standards?
2. What is the quality of self-monitoring intervention studies for students with disabilities relating to maintenance?
3. What is the quality of self-monitoring intervention studies for students with disabilities relating to generalization?

4. To what extent were maintenance and generalization assessed (i.e., length, dimension) and what were the outcomes?

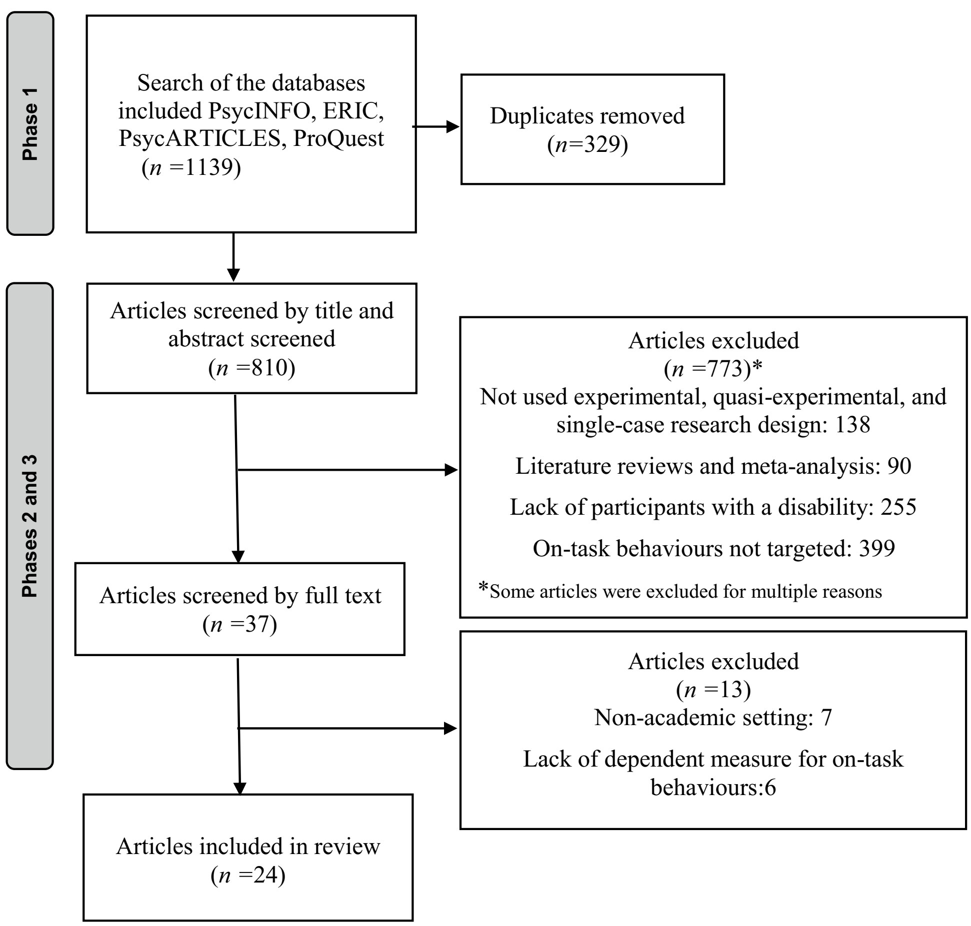
Method

The systematic literature review consisted of three phases (see Fig. 1). The first phase involved searching electronic databases: PsycINFO, Education Resources Information Center (ERIC), PsycARTICLES, and ProQuest databases. Key terms with truncation and Boolean operators were used for the search: *self-monitoring*, *self-recording*, *self-evaluation*, *self-regulation*, *self-reporting*, *self-management*, *self-reinforcement*, *self-observation*, *self-graphing*, *self-control* and *disability*, *intellectual disability*, *learning disability*, *developmental disability*, *neurodevelopmental disability*, *autism*, *autism spectrum disorders* (*ASD*), *Asperger's syndrome*, *pervasive developmental disorder* (*PDD*), *pervasive developmental disorder or not otherwise specified* (*PDD-NOS*), *cerebral palsy*, *emotional and behavioral disabilities* (*EBD*). The initial search included studies only written in English and published in peer-reviewed journals in the last 10 years (2011 to March 2021). We selected only articles from the last 10 years because we wanted to restrict this research review to the most current research. Unfortunately, there is no consensus on how "current" is defined. Therefore, we defined current research for this paper as anything published within the last 10 years.

An ancestral search was also conducted using the reference sections of articles found during the electronic review were examined to ensure articles were not missed in the electronic search. Initially, a total of 1139 articles were found. There were 810 studies remaining after the removal of duplications.

In the second phase, the first and fourth authors used an additional set of criteria to narrow the number of research articles to those that were empirical and experimental in nature and involved participants with disabilities. These criteria included: (a) used a(n) experimental, quasi-experimental, or single-case design; (b) involved at least one participant with a disability; (c) was not a meta-analysis or review article; and (d) intervention aimed to improve on-task/off-task/task-engagement of participants. The first author reviewed all 810 articles and the fourth author reviewed 20% ($n = 162$) of studies. The interrater agreement on the title/abstract review was 100%. A total of 37

Fig. 1 PRISMA flow diagram of literature review process



articles were included for further review. All of the 37 articles were single-case designs.

In the final phase of this review, the first and the second authors reviewed all of the remaining 37 articles using the following criteria: (a) a self-monitoring intervention/program/strategy was used to increase on-task/off-task/task-engagement behaviors in academic settings, and (b) a dependent measure was used for on-task/off-task/task-engagement behavior. If the two authors disagreed on an article, they discussed until they reached a consensus. The interrater agreement in this phase was 91.9%. A total of 24 studies are included for the current study. Note: Because the purpose of this review was to determine if and to what extent maintenance and generalization were assessed in the literature, maintenance and generalization were not a requirement for inclusion.

Article Analysis

The method of analysis of included articles was replicated from Neely et al. (2016) and Neely et al. (2018) with some exceptions described below. Some alterations were made in the descriptive synthesis (e.g., participant[s] characteristic[s]) in order to give more detailed information about included studies to better understand specifics about maintenance and generalization probes.

Descriptive Synthesis

The first and the fourth authors summarized the 24 studies within following categories: (a) participant(s) characteristics and setting(s); (b) interventionist(s) characteristics; (c) research design; (d) maintenance assessment design; (e) latency to maintenance and maintenance length; (f) generalization dimension and assessment design; (g) maintenance and/or generalization teaching strategy; and (h) maintenance and generalization results. While Neely et al. (2018) included target behavior(s) in the participant characteristics column, we excluded "target behaviors" since *on task/task engagement* behaviors were part of our inclusion criterion for this study. Instead, we added *settings* in this column to make potential comparisons between intervention and generalization settings across studies.

Participant Characteristic(s) and Setting(s)

In addition to the WWC standards (Kratochwill et al., 2013), we included additional information on participants' characteristics compared to Neely et al. (2016) and Neely et al. (2018). We added the number of participants along with age, race, gender, IQ, and diagnosis. Settings included information on where studies took place (i.e., general education classroom, special education resource room).

Interventionist Characteristics

Three main categories were used for interventionist characteristics: experimenter, educator, or family members. Note that more than one category of interventionists can occur in a study (i.e., researchers and educators). For the present study, *experimenter*(*s*) were researchers who conducted any of phases of implementation process (i.e., baseline, intervention, maintenance, generalization); *educator*(*s*) were general or special education teachers, instructional assistants, paraprofessionals, or group home members; and *family member*(*s*) were caregivers who participated in any phase of implementation process. If the interventionist was not clearly stated, it was coded as *researcher implemented*.

Research Design

The research designs utilized for each study were coded descriptively (e.g., withdrawal, ABAB design).

Maintenance Assessment Design

Two different maintenance assessment designs were defined: *single probe* and *multiple probes*. If a study included only one data point for maintenance for at least one of the participants, it was coded as *single maintenance probe*. When a study included more than one data point for at least one of the participants, it was coded *multiple probes*.

Latency to Maintenance and Maintenance Length

Studies were coded descriptively (e.g., 2 weeks, 6 weeks) depending on the elapsed time between the cessation of final intervention phase to the initial maintenance data collection. Differing from Neely et al. (2016) and Neely et al. (2018), the elapsed time between initial and final maintenance data collection was also coded descriptively to define the total length of time for maintenance data collection.

Generalization Dimension and Assessment Design

The generalization dimension included four different categories: people, task or activity, setting, or condition/context. The generalization assessment design included three categories: *single probe*, *multiple probes*, and *continuous probes* (Note: For purposes of this review, we were concerned with determining if and how generalization was assessed as opposed to the experimental demonstration [i.e., study quality] of generalized responding. Thus, there was not a requirement for generalization to occur before, during, and/or after the intervention in these categories). A study was coded as *single probe* when it included only one generalization data point for one of the participants. To be coded as *multiple probes*, studies had to have more than one data point for at least for one of the participants. Finally, studies were coded *continuous probes* when generalization probes were collected during baseline, intervention, and maintenance or generalization phases.

Maintenance and/or Generalization Teaching Strategy

Differing from Neely et al. (2016) and Neely et al. (2018) who categorized teaching strategies such as train and hope, program common stimuli, or train sufficient exemplars, we simply included whatever teaching strategy/program or procedures were used during maintenance and generalization of studies due to the use of self-monitoring interventions (i.e., a maintenance and generalization strategy, Chafouleas et al., 2012; Cooper et al., 2020).

Maintenance and Generalization Results

Maintenance and generalization results were categorized into three categories: *positive*, *negative*, or *mixed*. *Positive* results were coded when on-task/task engagement behaviors were maintained and/or generalized above baseline for all participants; *negative* results were coded when on-task/task engagement behavior were maintained and/or generalized at or below baseline condition for all participants; *mixed* results were coded when on-task/task engagement behaviors were maintained and/or generalized above baseline for some participants and at or below for other participants. Note that the reversal of this coding was used for off-task behaviors. For instance, *positive* results were coded when off-task behaviors were maintained and/or generalized below baseline for all participants; *negative* results were coded when off -task behavior were maintained and/or generalized at or above baseline condition for all participants; *mixed* results were coded when off-task behaviors were maintained and/or generalized below baseline for some participants at or above baseline for other participants.

Design Standards

A single-case design was used across all studies included in this review. Four design standards (1, 2[2A, 2B, 2C], 3, 4) from the WWC (Kratochwill et al., 2013) and a rubric based on these standards (2AM, 2AG; 2BM, 2BG, 2CM, 2CG; 4 M, 4G) developed by Neely et al. (2018) were used to investigate the quality, maintenance, and generalization of self-monitoring interventions to enhance on-task/off-task/task engagement behaviors for students with disabilities (see Table 1).

Interrater Agreement for Descriptive Table and Design Standards

The first and the fourth authors independently coded included studies for design standards and descriptive analysis. The 24 studies were coded by the first author. The randomly selected 15 (62.5%) studies were coded by the fourth author for WWC design standards and maintenance and generalization rubric standards that yielded 95.3% interrater agreement. The randomly selected 13 (54%) studies were coded by the fourth author for descriptive table. The interrater agreement was 97.4% for the descriptive table where there was a total of eight components used per study (participant characteristic(s) and setting(s), interventionist(s) characteristics, research design, maintenance assessment design, latency to maintenance and maintenance length, generalization dimension and assessment design, maintenance and/or generalization teaching strategy, maintenance, and generalization results). The disagreement s across design standards and descriptive table were discussed until the authors came to 100% consensus.

Results

Descriptive Data

Table 2 depicts the results for descriptive data for the 24 studies.

Participant Characteristic(s) and Setting(s)

A total of 58 participants with disabilities were included across the 24 studies. Seven participants were diagnosed with an intellectual disability (ID), 15 participants diagnosed with autism spectrum disorder (ASD), and eight participants were diagnosed with ID and ASD. The remaining disability categories included attention deficit hyperactivity disorder ($n=7$); emotional disorders ($n=3$); language impairment ($n=5$); other health impairments ($n=11$); and pervasive developmental disorder ($n=2$). The ethnicities were reported in 12 studies and included 29 participants—White ($n=14$), Black ($n=4$), Turkish ($n=4$), Latinx ($n=3$), Black/Biracial ($n=1$), Chinese ($n=1$), Multiracial ($n=1$), and Native American ($n=1$). Excluding Bruhn et al. (2016), all the studies reported the age of participants; the age range was from 5.5 to 19 years. Participants' IQs were reported in 10 studies and ranged from 49 (Boswell et al., 2013) to 106 (Romans et al., 2020).

The majority of studies took place in school settings in addition to one juvenile facility and research centers. Studies were conducted in general education classroom settings ($n=7$), special education classrooms (non-specified)

Table 1 WWC design standards and rubric standards for maintenance and generalization developed by Neely et al. (2016) and Neely et al. (2018)

WWC design standards
Design standard 1: When the independent variable was manipulated systematically, it is coded with "1." A rating of "0" was given when the independent variable was not manipulated systematically
Design standards 2A: When interobserver agreement was reported, it is coded with "1." A rating of "0" was provided when interobserver agreement was not reported
Design standard 2B: When interobserver agreement was reported at least 20% of all sessions along with at least 20% of the data points for each condition, it was coded with "2." A rating of "1" was marked when interobserver agreement was reported for at least 20% of all sessions but was not reported at least 20% of the data points for each condition. Studies were coded with "0" if interobserver agreement was reported less than 20% of all sessions
Design standard 3: When three attempts at three different times for ABAB design, multiple baseline designs, and changing criterion design or five attempts at five different times for alternating treatment design included to demonstrate intervention effect, it was coded with "2." A coding of "1" defined when an alternative treatment design had four attempts and the other designs had three attempts for one participant at least. If alternative treatment design studies included less than four data points and the all other designs provided less than three data points at least for one participant, the study was coded "0."
Design standard 4: When minimum five data points were reported for each condition, it was coded with "2." A rating for "1" was coded when each condition had at least 3 data points. Studies were coded with "0" any condition having less than 3 data points
Overall rating
When all design standards 1–4 were met with the highest possible rating, the study was coded "meet standards" with a coding "2." If at least one of the design standards 1–4 were coded with "1" where the highest score was 2 and then the study was coded with "1" meaning "meet standards with reservations." An overall rating of "0" indicated that at least one design standard 1–4 was coded with a "0" and considered "did not meet standards." Note: Ratings for rubric standards 2AM, 2AG, 2BM, 2BG, 4 M, and 4G below were not taken into account for overall rating
Rubric standards for maintenance and generalization developed by Neely et al. (2016) **and** Neely et al. (2018)
Rubric standard 2AM: When interobserver agreement was collected during maintenance and identified specifically, it is coded with "1." A rating of "0" was given when interobserver agreement was not specifically reported for maintenance
Rubric standard 2AG: When interobserver agreement was collected during generalization and identified specifically, it is coded with "1." A rating of "0" was documented when interobserver agreement was not specifically reported for generalization
Rubric standard 2BM: When interobserver agreement was specifically reported for at least 20% of maintenance, it was coded with "1." A rating of "0" was coded when interobserver agreement was not collected for at least 20% of maintenance or was not specifically defined
Rubric standard 2BG: When interobserver agreement was specifically reported for at least 20% of generalization, it was coded with "1." A rating of "0" was given when interobserver agreement was not collected for at least 20% of generalization or was not specifically defined
Design standard 4 M: When a maintenance phase included 3 or more data points, it was coded with "2." A rating of "1" was provided when 1 or 2 data point were collected for maintenance. Studies were coded with "0" if the study was not reported any point for maintenance
Design standard 4G: When generalization phase included 3 or more data points, it was coded with "2." A coding of "1" was provided when 1 or 2 data points were collected for generalization. Studies were coded with "0" if the study was not reported any data points for generalization

($n = 3$), special education resource rooms ($n = 2$), special education self-contained classrooms ($n = 3$), a maximum security juvenile facility ($n = 1$), a general education and self-contained classroom ($n = 1$), a large public university ($n = 1$), an autism classroom ($n = 1$), a university-based Applied Behavior Analysis (ABA) center ($n = 1$), and an autism research center affiliated with a university ($n = 1$) and a school setting (non-specified, $n = 3$).

Interventionist Characteristic(s)

The interventionist(s) were researcher(s) in 12 studies, educators in 7 studies, and a combination of researcher(s) and educator(s) in 5 studies.

Research Design

The following single case research designs were used in the studies: ABAB/withdrawal ($n = 12$), multiple probe ($n = 8$), alternating treatments ($n = 1$), modified reversal ($n = 1$), non-concurrent multiple baseline ($n = 1$), and changing criterion ($n = 1$).

Maintenance Assessment Design

A total of 12 studies assessed maintenance; 2 of these (Cook & Sayeski, 2020; Kolbenschlag & Wunderlich, 2019) also assessed generalization. Of these 12 studies, 10 used multiple-probe assessment designs and 2 studies used a single-probe design.

Table 2 Descriptive table for included studies

Article	Participant(s) characteristics and setting(s)	Interventionist(s) characteristic(s)	Research design	Maintenance assessment design	Latency to maintenance and maintenance length	Generalization dimension and assessment design	M/G teaching strategy	M/G results
Aykut (2020)	$n=3$; 13-year-old female with mild/moderate ID; IQ = 46; 12-year-old male with mild/moderate ID; IQ = 54; 13-year-old female with mild/moderate ID; IQ = 54 Special education self-contained classroom	Researcher(s)	Multiple probe design	Single probe	Upon completion of the intervention phase for the first participant	—	—	Positive
Beckman et al. (2019)	$n=2$; 12-year-old white male with fragile X syndrome, PDD, and ASD; IQ = 61; 10-year-old white male with ASD Special education self-contained classroom in an upper elementary school (5th and 6th grade)	Educator(s)	ABAB withdrawal design	—	—	—	—	—
Bedesem (2012)	$n=2$; 13-year-old with OHI; 14-year-old with SLD General education classroom in a public middle school	Researcher(s)	Multiple probe design	—	—	—	—	—
Boswell et al. (2013)	$n=1$; 11-year-old male with moderate ID and seizure disorders, IQ = 49 Special education resource room in a public middle school	Educator(s)	ABAB withdrawal design	—	—	—	—	—

Table 2 (continued)

Article	Participant(s) characteristics and setting(s)	Interventionist(s) characteristic(s)	Research design	Maintenance assessment design	Latency to maintenance and maintenance length	Generalization dimension and assessment design	M/G teaching strategy	M/G results
Bruhn et al. (2016)	$n=2$; 7th grade white male with ADHD; 6th grade white male with EI General education classroom a middle school	Researcher(s)	ABAB withdrawal design plus maintenance	Multiple probe	Following master criterion	—	—	Positive
Caldwell and Joseph (2012)	$n=3$;17-year-old Latinx female with Conduct Disorder, Bipolar Disorder without psychotic features, ADHD, Cannabis abuse, and alcohol abuse; 18-year-old white female with mood disorder (not otherwise specified), polysubstance dependence, and conduct disorder; and 14-year-old Black/Biracial female with emotional disturbance and diagnosed with major depressive disorder, post-traumatic stress disorder, conduct disorder, and ADHD Maximum-security juvenile facility in a large metropolitan city	Researcher(s)	ABABC withdrawal design	Multiple probe	—	—	Gradual fading	Mixed

Table 2 (continued)

Article	Participant(s) characteristics and setting(s)	Interventionist(s) characteristic(s)	Research design	Maintenance assessment design	Latency to maintenance and maintenance length	Generalization dimension and assessment design	M/G teaching strategy	M/G results
Clemons et al. (2016)	$n = 3$; 17-year-old male with SLD; 17-year-old male with ASD; 15-year-old-female with ID General education and self-contained classroom in a public high school	Researcher(s)	ABAB withdrawal design	Multiple probe	LM: 2–4 weeks ML: 2 weeks	—	Gradual fading	Positive
Cook and Sayeski (2020)	$n = 4$;16-year-old-Black male with OHI; 16-year-old white female with SLD; 15-year-old-multiracial male with SLD; 15-year-old-white-female with SLD General education classroom in a public high school	Researcher(s)	Multiple probe design	Multiple probe	LM: 2–7 weeks ML: 3–7 weeks No maintenance data for P2	Across settings Continuous probes	One-time reminder	M-potential Durability G-Negative
Ennis et al. (2018)	$n = 3$; 11.8-year-old Black male with OHI; 11.8-year-old Black male with OHI; 11.1-year-old white female with SLD Special education resource room in an elementary school	Educator(s)	ABAB withdrawal design and visual analysis	—	—	—	—	—

Table 2 (continued)

Article	Participant(s) characteristics and setting(s)	Interventionist(s) characteristic(s)	Research design	Maintenance assessment design	Latency to maintenance and maintenance length	Generalization dimension and assessment design	M/G teaching strategy	M/G results
Finn et al. (2015)	$n=4$; 8.10-year-old-white male with ASD and LI, IQ = N/A; 8.7 year-old white male with ASD and OHI, IQ = 97; 9.10 year-old Latinx male with ASD and LI, IQ = 71; 8.8 year-old-Latinx male with ASD and LI, IQ = 101 The first author's classroom in an elementary school	Researcher(s) and educator(s)	Multiple probe design	Multiple probe	Following mastery criterion	—	Gradual fading and booster retraining sessions	Positive
Huffman et al. (2019)	$n=1$; 19-year-old male with ASD and no ID Large lecture hall with stadium-style multi-level seating class in a large public university	Educator(s)	Alternating treatment design	—	—	—	—	—
Imasaka et al. (2020)	$n=2$; 8-year-old male with Asperger's Disorder; 8-year-old male with ADHD General education classrooms in a public school	Researcher(s)	Multiple probe design	Single probes	1 week to 10 days	—	Gradual fading	Positive

Table 2 (continued)

Article	Participant(s) characteristics and setting(s)	Interventionist(s) characteristic(s)	Research design	Maintenance assessment design	Latency to maintenance and maintenance length	Generalization dimension and assessment design	M/G teaching strategy	M/G results
Kartal and Yucesoy Ozkan (2015)	$n=4$; 5.5-year-old Turkish female with ID; 5.8-year-old-Turkish female with down syn and mild ID; 6.3 year-old-Turkish female with ID; 5.9-year-old female with moderate ID and ADHD General education classrooms in public schools	Researcher(s)	Multiple probe design	Multiple probe	Following mastery criterion	—	—	Mixed
Kolbenschlag and Wunderlich (2019)	$n=3$; 9-year-old with ASD and low-average range IQ; 11-year-old with ASD, low-average range IQ; 11-year-old with ASD, low-average range IQ Autism classroom in an elementary school	Educator(s)	Multiple probe design	Multiple probe	Following mastery criterion	Across settings Continuous probe	—	M-positive G-negative
Roberts et al. (2019)	$n=2$;18-year-old white male with ASD, IQ=99; 17-year-old white male with ASD, IQ=105 Special education classroom in a public school	Researcher(s) and educator(s)	ABAB withdrawal design	—	—	—	—	—

Table 2 (continued)

Article	Participant(s) characteristics and setting(s)	Interventionist(s) characteristic(s)	Research design	Maintenance assessment design	Latency to maintenance and maintenance length	Generalization dimension and assessment design	M/G teaching strategy	M/G results
Romans et al. (2020)	$n=2$; 17-year-old male with OHI for PDD-NOS, ADHD, and a mild ID, IQ = 69; 15-year-old male with OHI for ADHD and ASD, IQ = 106 Special education classroom in a public high school	Educator(s)	ABAB withdrawal design	—	—	—	—	—
Rosenbloom et al. (2016)	$n=1$; 9-year-old Black male with ASD no ID General education classroom in an elementary school	Researcher(s)	ABAB withdrawal design	—	—	—	—	—
Scalzo et al. (2015)	$n=1$;12-year-old-white male with ASD and severe ID Therapy rooms in university-based ABA center	Researcher(s)	Modified reversal design	—	—	—	—	—
Stasolla et al. (2014)	$n=2$; 7.5 male with ASD no ID; 8.5 male with ADHD no ID During academic activities in a school setting	Researcher(s) and educator(s)	Non-concurrent multiple baseline design	Multiple probe	LM: 1 month ML: 1 month	—	Fading out strategy	Positive
Stasolla et al. (2017)	$n=3$; 9-year-old male with CP, IQ = 72; 8-year-old male with CP, IQ = 70; 10-year-old male with CP, IQ = 68 During classroom activities in school settings	Researchers and educator(s)	Multiple probe design	Multiple probe	LM: 3 months ML: 1 month	—	—	Positive

Table 2 (continued)

Article	Participant(s) characteristics and setting(s)	Interventionist(s) characteristic(s)	Research design	Maintenance assessment design	Latency to maintenance and maintenance length	Generalization dimension and assessment design	M/G teaching strategy	M/G results
Szwed and Bouck (2013)	*n* = 3; 8-year-old-male with ADHD; 7-year-old-male with ADHD; 7 year-old-male with ED General education classroom in an elementary school	Researcher(s) and teacher (s)	ABAB withdrawal design	Multiple probe	LM: 3 weeks ML: 3 days	—	—	Negative
Wills and Mason (2014)	*n* = 2; 15.7-year-old Native American with SLD; 14.9-year-old white male with OHI for ADHD General education classroom in a high school	Researcher(s)	ABAB withdrawal design	—	—	—	—	—
Xin et al. (2017)	*n* = 4; 11-year-old female with ASD, IQ = 50; 10-year-old female with ASD, IQ = 57;10-year-old female with ASD, IQ = 59; 12-year-old male with ASD, IQ = N/A Special education classroom	Educator(s)	ABAB withdrawal design	—	—	—	—	—
Xu et al. (2017)	*n* = 1; 9-year-old Chinese male with mild ASD, IQ = 69 Autism research center affiliated with a university and a public elementary school	Researcher(s)	Changing criterion design	Multiple probe	1 week	—	—	Positive

M/G maintenance/generalization, *ABA* applied behavior analysis, *ASD* autism spectrum disorders, *ADHD* attention deficit hyperactivity disorder, *ED* emotional disorders, *ID* intellectual disability, *LI* language impairment, *LM* latency to maintenance, *ML* maintenance length, *OHI* other health impairments, *PDD* pervasive developmental disorder, *PDD-NOS* pervasive developmental disorder not otherwise specified, *SLD* specific learning disability

Latency to Maintenance and Maintenance Length

Of the 12 studies that assessed maintenance, 5 studies collected maintenance data immediately after meeting a mastery criterion. Stasolla et al. (2017) had the longest maintenance period at 3 months. Cook and Sayeski (2020) collected maintenance data up to 7 weeks. Caldwell and Joseph (2012) did not specify the latency to maintenance. The remaining 4 studies collected maintenance data between 1 week and 1 month. The length of maintenance data collection procedures was reported in only 5 studies and ranged between 3 days (Szwed & Bouck, 2013) to 7 weeks (Cook & Sayeski, 2020).

Generalization Dimension and Assessment Design

Only two studies assessed for generalization of self-monitoring interventions. Cook and Sayeski (2020) assessed generalization effects weekly in another classroom. A one-time reminder was provided during training. Kolbenschlag and Wunderlich (2019) assessed generalization across settings within a general education classroom once or twice in a week. There was no additional instruction provided for students to monitor their own behavior. Multiple baseline designs were used in both studies.

Maintenance and/or Generalization Teaching Strategy

An additional teaching strategy was used to program for maintenance and generalization in 6 studies. These strategies included gradually fading self-monitoring interventions ($n=4$), adding a one-time reminder ($n=1$), and gradually fading self-monitoring interventions with booster training ($n=1$).

Maintenance/Generalization Results

Among the 10 studies that assessed maintenance only, 7 demonstrated positive results—the participants' on-task behaviors maintained after withdrawal of the self-monitoring intervention. Caldwell and Joseph (2012) and Kartal and Yucesoy Ozkan (2015) revealed mixed results for maintenance in which on-task behaviors returned to near baseline levels for at least one participant in the studies. Szwed and Bouck (2013) found negative results for maintenance where participants' off-task behaviors reduced during the self-monitoring intervention but returned to baseline levels after the self-monitoring intervention was removed. The 2 studies (e.g., Cook & Sayeski, 2020; Kolbenschlag & Wunderlich, 2019) that assessed for maintenance and generalization indicated that maintenance results were positive, but participants' on-task behaviors did not generalize to other settings in either study.

Quality Review

Table 3 depicts the results for quality review.

Design Standards

Of the 24 studies reviewed, only 3 met all design standards, 15 studies met standards with some reservations, and 6 studies did not meet standards. Of the 15 studies that met standards with some reservations, design standard 4 (three to five data points provided for each condition) and design standard 2B (interobserver agreement data collected 20% of each condition) were the most frequent limitations. Of the 6 studies that did not meet the design standards, 2 studies did not meet design standard 4 due to the lack of a minimum of three data points for each condition. Interobserver agreement was not reported in 2 studies that did not meet design standard 2A. Two studies did not meet design standard 3 due to lacking a minimum of three conditions or opportunities to observe an intervention effect.

Maintenance and Generalization Rubric Standards Synthesized with Quality Results

Overall, of the 12 studies that assessed for maintenance and/or generalization, only 1 study (Xu et al., 2017) met all WWC design standards (Kratochwill et al., 2013) and maintenance rubric standards. One study (Kartal & Yucesoy Ozkan, 2015) met WWC design standards with some reservations and met all maintenance rubric standards. These 2 studies were not evaluated for generalization rubric standards because they did not include generalization phases. Of the remaining 10 studies, 1 met the WWC standards, 6 met the WWC design standards with some reservations, and 3 did not meet the WWC design standards. None of the 10 met the generalization and maintenance rubric standards.

For these 12 studies that assessed maintenance and generalization, interobserver agreement during maintenance and/or generalization conditions (rubric standards 2AM and 2AG) were collected in only 3 studies. All 3 of these studies assessed at least 20% of maintenance and/or generalization probe sessions (rubric standards 2BM and 2BG) with 80% agreement (rubric standards 2CM and 2CG) and collected a minimum three data points (rubric standards 2CM and 2CG).

Discussion

The current study aimed to evaluate the quality of self-monitoring studies and to assess the maintenance and generalization for the on-task behaviors of students with disabilities. Problem behaviors that occur in the classroom settings such

Table 3 Design and rubric standard evaluation

WWC standards								Neely et al. (2016) and Neely et al. (2018) standards							
Article citation	Overall	DS#1	DS# 2A	DS# 2B	DS# 2C	DS# 3	DS# 4	RS# 2AM	RS# 2AG	RS# 2BM	RS# 2BG	RS# 2CM	RS# 2CG	RS# 4 M	RS# 4G
Meets standards															
Clemons et al. (2016)	2	1	1	2	2	2	2	0	—	0	—	0	—	1	—
Rosenbloom et al. (2016)	2	1	1	2	2	2	2	—	—	—	—	—	—	—	—
Xu et al. (2017)	2	1	1	2	2	2	2	1	—	1	—	1	—	2	—
Meets standards with reservations															
Aykut (2020)	1	1	1	1	2	2	2	0	—	0	—	—	—	1	—
Beckman et al. (2019)	1	1	1	1	2	2	1	—	—	—	—	—	—	—	—
Boswell et al. (2013)	1	1	1	1	2	2	1	—	—	—	—	—	—	—	—
Bruhn et al. (2016)	1	1	1	1	2	1	2	0	—	0	—	0	—	—	—
Cook and Sayeski (2020)	1	1	1	2	2	2	1	1	0	1	0	1	0	2	2
Ennis et al. (2018)	1	1	1	2	2	2	1	—	—	—	—	—	—	—	—
Finn et al. (2015)	1	1	1	1	1	2	1	0	—	0	—	0	—	2	—
Huffman et al. (2019)	1	1	1	1	1	2	1	—	—	—	—	—	—	—	—
Imasaka et al. (2020)	1	1	1	2	1	2	2	0	—	0	—	0	—	1	—
Kartal and Yucesoy Ozkan (2015)	1	1	1	2	2	2	1	1	—	1	—	1	—	2	—
Kolbenschlag and Wunderlich (2019)	1	1	1	2	2	2	1	0	0	0	0	0	0	2	2
Romans et al. (2020)	1	1	1	2	2	2	1	—	—	—	—	—	—	—	—
Scalzo et al. (2015)	1	1	1	2	2	1	1	—	—	—	—	—	—	—	—
Szwed and Bouck (2013)	1	1	1	1	2	2	2	0	—	0	—	0	—	2	—
Wills and Mason (2014)	1	1	1	1	2	2	2	—	—	—	—	—	—	—	—
Does not Meet Standards															
Bedesem et al. (2012)	0	1	1	1	1	0	2	—	—	—	—	—	—	—	—
Caldwell and Joseph (2012)	0	1	0	0	2	2	1	0	—	0	—	0	—	2	—
Roberts et al. (2019)	0	1	1	2	2	2	0	—	—	—	—	—	—	—	—
Stasolla et al. (2014)	0	1	1	0	1	0	1	0	—	0	—	0	—	2	—
Stasolla et al. (2017)	0	1	1	0	1	2	0	0	—	0	—	0	—	2	—
Xin et al. (2017)	0	1	0	0	1	2	2	—	—	—	—	—	—	—	—

as negative statements and disruptive behaviors tend to be incompatible with on-task behaviors (i.e., Xu et al., 2017). Therefore, improving on-task behaviors can result in fewer disruptive behaviors (i.e., Martella et al., 1993; Xu et al., 2017) and provide more time for teachers to teach instead of continuing to prompt for appropriate behaviors (Agran et al., 2003). Additionally, improved on-task behaviors can also improve the academic performance of students (Dalton et al., 1999; Mattson & Pinkelman, 2020). In turn, the improvement of on-task behaviors can help educators create positive classroom climates where the inclusion of students with disabilities occurs effectively (Huang et al., 2013).

An advantage of teaching students with disabilities to self-monitor is that students learn to observe and record their own behavior instead of relying on adults (e.g., parents, teachers, educational assistants) (Wehmayer et al., 1998; Wehmayer et al., 2003; Yucesoy Ozkan & Sonmez, 2011). Previous research has shown that these interventions are inclusive in nature and their effective implementations can yield improved academic and behavioral outcomes (Ardoin & Martens, 2004; Xin et al., 2017) along with social acceptance of students with disabilities (Sheffield & Waller, 2010). As a result, self-monitoring interventions have been commonly used among practitioners to enhance on-task behaviors of students with disabilities (Cooper et al., 2020). Therefore, an evaluation of the self-monitoring research is particularly important.

To determine the quality of self-monitoring research conducted with the students with disabilities to improve on-task behaviors, we conducted a systematic search of the literature and found a total of 24 studies. The quality of included studies was assessed with WWC design standards (Kratochwill et al., 2013) and the rubric standards outlined and modified by Neely et al. (2016) and Neely et al. (2018). Information from the studies included the following: (a) participant characteristic(s) and setting(s), (b) interventionist(s) characteristics, (c) research design, (d) maintenance assessment design, (e) latency to maintenance and maintenance length, (f) generalization dimension and assessment design, (g) maintenance and/or generalization teaching strategy, (h) maintenance and generalization results.

The overall quality of these studies appeared to be weak based on WWC design standards (Kratochwill et al., 2013). Our findings revealed that only 3 studies (Clemons et al., 2016; Rosenbloom et al., 2016; Xu et al., 2017) met all design standards. One main issue was the limited amount of data collected. WWC recommends the collection of a minimum of five data points for each condition of the intervention in case attrition occurs (Kratochwill et al., 2013). This data collection helps to not only establish the stability of data during baseline but allow for the evaluation of the treatment effect on behavior over time (Dixon et al., 2016). Given that we found an inadequate number of data points in baseline and intervention conditions, future studies should include multiple data points (i.e., five or more) to demonstrate meaningful experimental control with rigorous internal validity (Martella et al., 1999; Horner et al., 2005).

Another area of weakness found in this review was a lack of interobserver agreement during maintenance and generalization conditions. Acceptable interobserver agreement requires that these assessments occur for at least 20% of sessions in each condition (Kratochwill et al., 2013). However, it appears the majority of studies in the current review did not provide adequate information with respect to interobserver agreement across *all* conditions (i.e., baseline, intervention, maintenance, and generalization conditions). Future researchers should conduct adequate interobserver agreement across all conditions including maintenance and generalization.

Several other areas of needed improvement in the research literature were found based on the article analyses. First, Wood et al. (1998) indicated that self-monitoring interventions were generally implemented in special education settings; they recommended that future studies be conducted in more inclusive settings. However, our findings revealed that despite over 20 years of research since Wood et al. (1998) recommendation, the majority of self-monitoring studies were still conducted in special education classrooms or other restrictive settings (i.e., juvenile facilities and therapy rooms). This finding is particularly important if the goal is the inclusion of students with disabilities in general education classrooms. One way to overcome this challenge can be accomplished through encouraging general education classroom teachers to participate in self-monitoring interventions. An initial step can be taken through training general education teachers to implement and incorporate self-monitoring intervention into their daily practices. In addition, parents and caregivers should also be trained in self-monitoring procedures to provide continuity and consistency from school to home.

Second, there is a question about the extent to which behavior changes maintain into the future after the self-monitoring interventions are withdrawn (Martella et al., 1993). Of the 24 studies included in the review, 12 did not assess for maintenance and generalization. A total of 10 studies assessed maintenance and only 2 assessed both maintenance and generalization. When maintenance data were collected, the latency to maintenance was either very short (e.g., less than a week or month) or not present at all (e.g., maintenance probes were collected immediately [i.e., following mastery criterion and upon completion of the phase by the first participant] following the intervention). Additionally, the maintenance length was reported in only 5 studies with the longest maintenance length being 7 weeks. Although there is no consensus in the field about the length of latency before starting to collect maintenance data (Neely et al., 2018), increased length of latency and long-term maintenance with multiple probes should be considered as a quality criterion in future investigations.

Third, even though the results were generally positive for maintenance, 2 studies reviewed for generalization indicated the generalization of self-monitoring interventions were negative. Negative generalization results were associated with minimal generalization programming in Cook and Sayeski (2020) and with a lack of fading procedures prior to maintenance and generalization phases in Kolbenschlang and Wunderlich (2019). Given there was very limited systematic planning for both maintenance and generalization across these studies, further studies with explicit maintenance and generalization programming are warranted. For instance, Wood et al. (2002) implemented self-monitoring interventions for students at-risk for school failure and drop-out by sequentially introducing the intervention into six different settings to promote generalization. Outcomes of this study revealed that the participants' on-task behaviors maintained after 87 days of the intervention and generalized to other settings and individuals. Wood et al. (2002) concluded that self-monitoring effects could be generalized when sequentially introduced into multiple settings (i.e., three).

Interestingly, the majority (i.e., $n=13$) of the 24 studies included an ABAB/withdrawal design ($n=12$) or a modified reversal ($n=1$). Given that these designs require a return to near baseline levels (Martella et al.,2013), it is not surprising that maintenance and generalization were not addressed given they would compromise the experimental control of these designs. Thus, unless maintenance and generalization are explicitly programmed *after* the demonstration of experimental control is made, it is not surprising that maintenance and generalization data were not integrated into these designs.

Fourth, there is a need to determine how effective educators (and parents) are at implementing self-monitoring interventions. Additionally, there is a need to train educators to fade intervention components (Bruhn et al., 2015). Unfortunately, our findings indicate that the majority of self-monitoring interventions were implemented by researchers; the absence of educators as interventionists was even more common across studies that assessed maintenance and/or generalization. Therefore, researchers should integrate teachers (and parents) into the implementation of self-monitoring programs to increase the external validity of this intervention.

A positive finding of this review was that research on self-monitoring interventions is not only focused on one group of disability (e.g., individuals with ID), but instead shows positive results with a diverse group of disabilities (e.g., autism, ADHD, students with behavioral challenges). This outcome with respect to demonstrating the effects of self-monitoring with a wide range of disabilities is important when considering increased educational opportunities in less restrictive educational environments. Additionally, the diversity in ethnic backgrounds (e.g., Asian, Black, Latinx, Middle Eastern, Native American, and White) across participants was positive and further demonstrates the external validity of self-monitoring. These positive outcomes are important when considering the need to promote access to equitable and inclusive educational opportunities for students with disabilities for diverse backgrounds.

Although several interesting and important findings were identified in this literature review, there were several limitations. First, we included studies up to March 2021. There is a possibility of other self-monitoring studies published after our review was completed. Secondly, some studies might have been omitted due to our inclusion criteria that at least one of the participants had to be diagnosed with a disability; thus, studies might have been excluded due to lack of participant diagnosis being identified. Finally, we coded studies if they *specifically* indicated they collected interrater agreement for each condition including maintenance and generalization; therefore, some studies might have collected interrater agreement for each condition but did not specify this in the manuscript and were thereby excluded.

Overall, there is a continued weakness in self-monitoring studies targeting on-task behaviors of students with disabilities with regard to assessing for maintenance and generalization of on-task behaviors. Half of the studies did not assess either maintenance or generalization. Strokes and Bear (1977) indicated limitations in generalization assessments in the field of ABA over 40 years ago. Although the need for generalization and maintenance was recognized decades later, particularly with regard to self-monitoring interventions (i.e., Wood et al., 2002), our findings indicate that this problem is still ongoing. Therefore, there is a continued need to establish the maintenance and generalization effects of self-monitoring on the on-task behavior of students with disabilities.

Declarations

Conflict of Interest The authors declare no competing interests.

References

*Articles Included in Review

Agran, M., King-Sears, M. E., Wehmeyer, M. L., & Copeland, S. R. (2003). *Teachers' guide to inclusive practices: Student-directed learning*. Paul H. Brookes.

Albion, F. M., & Salzberg, C. L. (1982). The effect of self-instruction on the rate of correct addition problems with mentally retarded children. *Education and Treatment of Children*, *5*(2), 121–131. https://www.jstor.org/stable/42900380

Anderson-Inman, L., Paine, S. C., & Deutchman, L. (1984). Neatness counts: Effects on direct instruction and self-monitoring on the transfer of neat-paper skills to non-training settings. *Analysis and Intervention in Developmental Disabilities, 4*(2), 137–155. https://doi.org/10.1016/0270-4684(84)90037-5

Ardoin, S. P., & Martens, B. K. (2004). Training children to make accurate self-evaluations: Effects on behavior and the quality of self-ratings. *Journal of Behavioral Education, 13*(1), 1–23. https://www.jstor.org/stable/41824317

*Aykut, C. (2020). Increasing self-evaluation use through video feedback to improve academic engagement among students with intellectual disabilities. *International Journal of Progressive Education, 16*(1), 111–124. https://doi.org/10.29329/ijpe.2020.228.9

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*(1), 91–97. https://doi.org/10.1901/jaba.1968.1-91

*Beckman, A., Mason, B., Wills, H. P., Garrison-Kane, L., & Huffman, J. M. (2019). Improving behavioral and academic outcomes for students with autism spectrum disorder: Testing an app-based self-monitoring intervention. *Education and Treatment of Children*, *42*(2), 225–244. https://doi.org/10.1353/etc.2019.0011

*Bedesem, P. L. (2012). Using cell phone technology for self-monitoring procedures in inclusive settings. *Journal of Special Education Technology*, *27*(4), 33–46. https://doi.org/10.1177/016264341202700403

*Boswell, M. A., Knight, V., & Spriggs, A. D. (2013). Self-monitoring of on-task behaviors using the MotivAider by a middle school student with a moderate intellectual disability. *Rural Special Education Quarterly*, *32*(2), 23–30. https://doi.org/10.1177/875687051303200205

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*(2), 106–118. https://doi.org/10.1037/a0016159

Briesch, A. M., & Daniels, B. (2013). Using self-management interventions to address general education behavioral needs: Assessment of effectiveness and feasibility. *Psychology in the Schools, 50*(4), 366–381. https://doi.org/10.1002/pits.21679

Briesch, A. M., Daniels, B., & Beneville, M. (2019). Unpacking the term "self-management": Understanding intervention applications within the school-based literature. *Journal of Behavioral Education, 28*(1), 54–77. https://doi.org/10.1007/s10864-018-9303-1

Broer, S. M., Doyle, M. B., & Giangreco, M. F. (2005). Perspectives of students with intellectual disabilities about their experiences with paraprofessional support. *Exceptional Children, 71*, 415–430.

Bruhn, A. L., McDaniel, S. M., & Kreigh, C. (2015). Self-monitoring interventions for students with behavior problems: A review of current research. *Behavioral Disorders, 40*(2), 102–121. 10.17988%2FBD-13–45.1

*Bruhn, A. L., Vogelgesang, K., Fernando, J., & Lugo, W. (2016). Using data to individualize a multicomponent, technology-based self-monitoring intervention. *Journal of Special Education Technology, 31*(2), 64–76. https://doi.org/10.1177/0162643416650024

Busacca, M., Anderson, A., & Moore, D. W. (2015). Self-management for primary school students demonstrating problem behavior in regular classrooms: Evidence review of single-case design research. *Journal of Behavioral Education, 24*(4), 373–401. https://doi.org/10.1007/s10864-015-9230-3

*Caldwell, S., & Joseph, L. M. (2012). Helping female juveniles improve their on-task behavior and academic performance using a self-management procedure in a correctional facility. *Contemporary School Psychology, 16*, 61–74. https://doi.org/10.1007/BF03340976

Chafouleas, S., Sanetti, L., Jaffery, R., & Fallon, L. (2012). An evaluation of a classwide intervention package involving self-management and a group contingency on classroom behavior of middle school students. *Journal of Behavioral Education, 21*(1), 34–57. https://doi.org/10.1007/s10864-011-9135-8

*Clemons, L. L., Mason, B. A., Garrison-Kane, L., & Wills, H. P. (2016). Self-monitoring for high school students with disabilities: A cross-categorical investigation of I-Connect. *Journal of Positive Behavior Interventions, 18*(3), 145–155. https://doi.org/10.1177/1098300715596134

Cook, B. G., Buysse, V., Klingner, J., Landrum, T. J., McWilliam, R. A., Tankersley, M., & Test, D. (2015). CEC's standards for classifying the evidence base of practices in special education. *Remedial and Special Education, 36*(4), 220–234. https://doi.org/10.1177/0741932514557271

*Cook K. B., & Sayeski K. L. (2020). High-school students with high-incidence disabilities 'use of smartphones for self-monitoring, *Exceptionality.* https://doi.org/10.1080/09362835.2020.1772064

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis (3rd Edition).* Pearson.

Dalton, T., Martella, R. C., & Marchand-Martella, N. E. (1999). The effects of a self-management program in reducing off-task behavior. *Journal of Behavioral Education*, 9(3), 157–176. https://doi.org/10.1023/A:1022183430622

Dixon, M. R., Belisle, J., Stanley, C., Speelman, R., Rowsey, K. E., & Daar, J. H. (2016). Establishing derived categorical responding in children with disabilities using the PEAK-E curriculum. *Journal of Applied Behavior Analysis, 50*, 134–145. https://doi.org/10.1002/jaba.355

*Ennis, R. P., Lane, K. L., & Oakes W. P., (2018) Empowering teachers with low-intensity strategies to support instruction: Self-monitoring in an elementary resource classroom. *Preventing School Failure: Alternative Education for Children and Youth, 62*(3), 176-189. https://doi.org/10.1080/1045988X.2017.1408055

*Finn, L., Ramasamy, R., Dukes, C., & Scott, J. (2015). Using Watch Minder to increase the on-task behavior of students with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 45*, 1408–1418. https://doi.org/10.1007/s10803-014-2300-x

Horner, R. H., Carr, E. G., Halle, J., McGee, G. G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165–179. https://doi.org/10.1177/001440290507100203

Howell, K. W., Rueda, R., & Rutherford, R. B. (1983). A procedure for teaching self-recording to moderately retarded students. *Psychology in the Schools, 20*, 202–220.

Huang, A. X., Jia, M., & Wheeler, J. J. (2013). Children with autism in the People's Republic of China: Diagnosis, legal issues, and educational services. *Journal of Autism and Developmental Disorders, 43*, 1991–2001. https://doi.org/10.1007/s10803-012-1722-6

*Huffman, J. M., Bross, L.A., Watson, E. K., Wills, H. P., & Mason, R. A. (2019). Preliminary investigation of a self-monitoring application for a postsecondary student with autism. *Advanced Neurodevelopmental Disorders, 3***,** 423–433. https://doi.org/10.1007/s41252-019-00124-y

Hughes, C. A., Korinek, L., & Gorman, J. (1991). Self-management for students with mental retardation in public school settings: A research review. *Education and Training in Mental Retardation, 26*(3), 271–291. http://www.jstor.org/stable/23878616

*Imasaka, T., Lee, P. L., Anderson, A., Wong, C. W. R., Moore, D. W., Furlonger, B., & Bussaca, M. (2020). Improving compliance in primary school students with autism spectrum disorder. *Journal of Behavioral Education, 29*(4)**,** 763–786. https://doi.org/10.1007/s10864-019-09346-5

Individuals with Disabilities Education Improvement Act, 20 U.S.C. § 1400 (2004).

*Kartal, M. S., & Yucesoy Ozkan, S. (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*(4), 418–432. https://www.researchgate.net › publication › 289373874

Knapczyk, D. R., & Livingston, G. (1973). Self-recording and student teacher supervision: Variables within a token economy structure. *Journal of Applied Behavior Analysis, 6*, 481–486.

*Kolbenschlag, C. M., & Wunderlich, K. L. (2019). The effects of self-monitoring on on-task behaviors in individuals with autism spectrum disorders. *Journal of Behavioral Education, 30***,** 1-12. https://doi.org/10.1007/s10864-019-09352-7

Koyama, T., & Wang, H. T. (2011). Use of activity schedule to promote independent performance of individuals with autism and other intellectual disabilities: A review. *Research in Developmental Disabilities, 32*(6), 2235–2242. https://doi.org/10.1016/j.ridd.2011.05.003

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34*(1), 26–38. https://doi.org/10.1177/0741932512452794

Maggin, D. M., Briesch, A. M., & Chafouleas, S. M. (2012). 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

Martella, R. C., Leonard, I. J., Marchand-Martella, N. E., & Agran, M. (1993). Self-monitoring negative statements. *Journal of Behavioral Education*, 3(1), 77–86. https://doi.org/10.1007/BF00947146

Martella, R. C., Nelson, J. R., & Marchand-Martella, N. E. (1999). *Research methods: Learning to become a critical research consumer*. Allyn and Bacon.

Martella, R. C., Nelson, J. R., Marchand-Martella, N. E., & O'Reilly, M. (2012). *Comprehensive behavior management: Individualized, classroom, and schoolwide approaches* (2nd ed.). Sage.

Martella, R. C., Nelson, J. R., Morgan, R. L., & Marchand-Martella, N. E. (2013). *Understanding and interpreting educational research*. Guilford.

Mattson, S. L., & Pinkelman, S. E. (2020). Improving on-task behavior in middle school students with disabilities using activity schedules. 13(1), 104–113. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7070130/

McDaugall, D., Heine, R. C., Wiley, L. A., Sheehey, M. D., Sakanashi, K. K., Cook, B. G., & Cook, L. (2017). Meta-analysis of behavioral self-management techniques used by students with disabilities in inclusive settings. *Behavioral Interventions, 32*(4), 399–417. https://onlinelibrary.wiley.com/doi/abs/https://doi.org/10.1002/bin.1491

Neely, L. C., Ganz, J. B., Davis, J. L., Boles, M. B., Hong, E. R., Ninci, J., & Gilliland, W. D. (2016). Generalization and maintenance of functional living skills for individuals with autism spectrum disorder: A review and meta-analysis. *Review Journal of Autism and Developmental Disorders, 3*, 37–47. https://doi.org/10.1007/s40489-015-0064-7

Neely, L., Garcia, E., Bankston, B., & Green, A. (2018). Generalization and maintenance of functional communication training for individuals with developmental disabilities: A systematic and quality review. *Research in Developmental Disabilities, 79*, 116–129. https://doi.org/10.1016/j.ridd.2018.02.002

Phillips, C. L., & Vollmer, T. R. (2012). Generalized instruction following with pictorial prompts. *Journal of Applied Behavior Analysis, 45*(1), 37–54. https://doi.org/10.1901/jaba.2012.45-37

Richards, S., Brady, M., & Taylor, R. (2015). *Cognitive and intellectual disabilities* (2nd ed.). Routledge.

Riden, B. S., Taylor, J. C., Ruiz, S., Lee, D. L., & Scheeler, M. C. (2020). Using a daily report card to reduce off-task behaviors for a student with autism spectrum disorder. *Journal of Behavioral Education*. https://doi.org/10.1007/s10864-020-09382-6

*Roberts, G. J., Mize, M., Reutebuch, C. K., Falcomata, T., Capin, P., & Steelman, B. L. (2019). Effects of self-management with peer training intervention on academic engagement for high school students with autism spectrum disorder. *Journal of Behavioral Education, 28*(1), 1–23. https://doi.org/10.1007/s10864-018-09317-2

Rojewski, J. W., Lee, I. H., & Gregg, N. (2015). Causal effects of inclusion on postsecondary education outcomes of individuals with high-incidence disabilities. *Journal of Disability Policy Studies, 25*(4), 210–219. 10.1177%2F1044207313505648

*Romans, S. K., Wills, H. P., Huffman, J. M., & Garrison-Kane, L. (2020). The effect of web- based self-monitoring to increase on-task behavior and academic accuracy of high school students with autism. *Preventing School Failure: Alternative Education for Children and Youth, 64*(3), 249-260. https://doi.org/10.1080/1045988X.2020.1732282

Rosenbaum, M. S., & Drabman, R. S. (1979). Self-control training in the classroom: A review and critique. *Journal of Applied Behavior Analysis, 12*(3), 467–485. https://doi.org/10.1901/jaba.1979.12-467

*Rosenbloom, R., Mason, R. A., Wills, P. H., & Mason, B. A. (2016). Technology delivered self-monitoring application to promote successful inclusion of an elementary student with autism. *Assistive Technology*, *28*(1), 9-16. https://doi.org/10.1080/10400435.2015.1059384

*Scalzo R, Henry K., Davis T. N., Amos, K., Zoch, T., Turchan, S., & Wagner, T. (2015) Evaluation of interventions to reduce multiply controlled vocal stereotypy. *Behavior Modification, 39*(4), 496–509. 10.1177%2F0145445515573986

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.

*Stasolla, F., Caffò, A. O., Perilli, V., Boccasini, A., Damiani, R., Albano, V., & Albano, A. (2017). Comparing self-monitoring and differential reinforcement of an alternative behavior to promote on-task behavior by three children with cerebral palsy: A pilot study. *Life Span and Disability, 20*(1), 63–92. https://www.researchgate.net/publication/318084792

*Stasolla, F., Perilli, V., & Damiani, R. (2014). Self-monitoring to promote on-task behavior by two high functioning boys with autism spectrum disorders and symptoms of ADHD. *Research in Autism Spectrum Disorders, 8*(5), 472-479. https://doi.org/10.1016/j.rasd.2014.01.007

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*(2), 349–367. https://doi.org/10.1901/jaba.1977.10-349

Stokes, T. F., & Osnes, P. G. (1989). An operant pursuit of generalization. *Behavior Therapy, 20*(3), 337–355. https://doi.org/10.1016/S0005-7894(89)80054-1

*Szwed, K., & Bouck, E. C. (2013). Clicking away: Repurposing student response systems to lessen off-task behavior. *Journal of Special Education Technology, 28*(2), 1–12. 10.1177%2F016264341302800201

Szykula, S. A., Saudargas, R. A., & Whaler, R. G. (1981). The generality of self-control procedures following a change in the classroom teacher. *Education and Treatment of Children, 4*(3), 253–263. https://www.jstor.org/stable/42900345

Turnbull, A. Turnbull H. R, Wehmeyer, M., & Shogren, K. (2019). *Exceptional lives: Practice, progress, & dignity in today's schools* (9th ed.). Pearson.

Wehmeyer, M. L., Agran, M., & Hughes, C. (1998). *Teaching self-determination to students with disabilities: Basic skills for successful transition.* Paul H. Brookes.

Wehmeyer, M. L., Yeager, D., Bolding, N., Agran, M., & Hughes, C. (2003). The effects of self-regulation strategies on goal attainment for students with developmental disabilities in general education classrooms. *Journal of Developmental & Physical Disabilities, 15*(1), 79–91. https://doi.org/10.1023/A:1021408405270

*Wills, H. P., & Mason, B. A. (2014). Implementation of a self-monitoring application to improve on-task behavior: A high school pilot study. *Journal of Behavioral Education, 23*(4), 421-434. https://doi.org/10.1007/s10864-014-9204-x

Wood, S. J., Murdock, J. Y., & Cronin, M. E. (2002). Self-monitoring and at-risk middle school students: Academic performance improves, maintains, and generalizes. *Behavior Modification*, *26*(5), 605–626. 10.1177%2F014544502236653

Wood, S. J., Murdock, J. Y., Cronin, M. E., Dawson, N. M., & Kirby, P. C. (1998). Effects of self-monitoring on on-task behaviors of at-risk middle school students. *Journal of Behavioral Education, 8*(2), 263–279. https://doi.org/10.1023/A:1022891725732

*Xin, J. F., Sheppard, M. E., & Brown, M. (2017). Brief report: Using ipads for self-monitoring of students with autism. *Journal of Autism and Developmental Disorders, 47*(5), 1559-1567. https://doi.org/10.1007/s10803-017-3055-y

*Xu, S., Wang, J., Lee, G. T., & Luke, N. (2017). Using self-monitoring with guided goal setting to increase academic engagement for a student with autism in an inclusive classroom in China. *The Journal of Special Education*, *51*(2),106–114. 10.1177%2F0022466916679980

Yucesoy Ozkan, S., & Sonmez, M. (2011). Examination of single subject studies conducted on individuals with disabilities by using self-management strategies: A meta-analysis study. *Educational Sciences*: *Theory and Practice*, *11*(2), 809–821. https://eric.ed.gov/?id=EJ927378

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.