Hammill Institute on Disabilities

Journal of Emotional and Behavioral Disorders
2022, Vol. 30(3) 171–184

DOI: 10.1177/10634266211051404
jebd.sagepub.com
SAGE

Behavioral Self-Regulation: A Comparison of Goals and Self-Monitoring for High School Students With Disabilities

Sara Estrapala, PhD[1], Allison Leigh Bruhn, PhD[2], and Ashley Rila, PhD[2]

Abstract

High school students, particularly those with disabilities, often struggle with academic engagement in general education classes. One empirically supported method for improving academic engagement includes self-regulation interventions. However, individual effects of frequently used self-regulation strategies' goal setting and self-monitoring remain unknown. To address this issue, we implemented a multiple-probe single-case design with embedded alternating treatments to evaluate the differential impact of behavioral goal reminders and self-monitoring. The sample consisted of three high school students with emotional or behavioral disorders (EBD) in general education classes who also received special education services. Innovative approaches to intervention design included student-determined behavioral goals and self-monitoring procedures informed by triangulated behavioral assessment data. Although visual analysis indicated only moderate improvements in academic engagement, and no functional relationships were determined for either the multiple-probe or alternating treatments designs, the intervention was rated with high social validity from teachers and students, and the study represents an exemplar of high-quality special education single-case research. Practical implications, limitations, and directions for future research are discussed.

Keywords

behavior, self-regulation, goals, self-monitoring, self-determination, academic engagement, high school, special education

High school students with disabilities and emotional or behavioral disorders (EBD) may lack the self-regulation skills necessary for academic and behavioral success (Gardner et al., 2008). In school, self-regulation includes the ability to control one's actions, thoughts, and emotions, particularly when faced with environmental influences which dissuade academic engagement (AE; Hagler et al., 2016). In addition, high school–age students typically experience nonlinear self-regulation development (e.g., periods of increases and decreases in emotion regulation, coping, anger management; Hagler et al., 2016) compared with early and middle childhood, which makes them particularly vulnerable to problems associated with poor self-regulation (e.g., anxiety, depression, aggression, impulsivity, peer rejection; Trentacosta & Shaw, 2009). In high school, these problems are intensified with increased academic and social pressures, hormonal changes, and the impending transition to adulthood and can result in failure, dropout, substance abuse, and arrest (Gardner et al., 2008). Fortunately, learning and practicing self-regulation through targeted interventions can improve in- and postschool outcomes (Shogren et al., 2017).

Self-regulation interventions frequently include skills such as goal setting and self-monitoring, either in a multicomponent package or target a single self-regulation skill (Carter et al., 2011). Furthermore, interventions targeting behavioral self-regulation may be enhanced with student decision-making, particularly at the high school level (Wehmeyer et al., 2004). That is, providing opportunities for intervention input enhances autonomy while reducing adult dependence—characteristics that are developmentally appropriate and desired by most adolescents (Mager & Nowak, 2011). Goal setting interventions include determining a desired behavior and a criterion or aim for improvement and can involve student input, progress monitoring, feedback, and reinforcement (e.g., Barbrack & Maher, 1984; Kelly & Shogren, 2014). Goals may also be

[1]University of Missouri, Columbia, USA
[2]The University of Iowa, Iowa City, USA

Corresponding Author:
Sara Estrapala, Department of Special Education, University of Missouri, 311 Townsend Hall, Columbia, MO 65211, USA.
Email: sle9bb@missouri.edu

individualized using preintervention behavior data (Bruhn et al., 2016) and are intended to motivate students to engage in contextually appropriate behaviors (Covington, 2000). In a study testing the effects of behavioral goal setting, authors reported significantly higher goal attainment and student satisfaction when students were directly involved in establishing goals compared with students who were uninvolved (Barbrack & Maher, 1984). Unfortunately, the authors did not report behavioral outcome data (attendance, disciplinary contacts, disruptive behaviors, etc.), maintenance, or generalization; instead, they reported student and counselor perceptions of goal attainment.

Self-monitoring interventions require students to think about their behavior at a given moment and record whether a specific behavior is occurring and they may include goals and self-evaluation (i.e., reviewing data and comparing it with goal achievement criteria; Bruhn et al., 2020), although are not essential (e.g., Wills & Mason, 2014). Two studies on self-monitoring with high school students with EBD reported functional relations between self-monitoring and improved on-task behavior (Clemons et al., 2016; Wills & Mason, 2014). However, student input and generalization data were not reported in either study and maintenance data were only collected in Clemons et al. (2016). Although goals were included in Clemons et al. (2016), the motivating impact of using goals was not studied.

Finally, researchers Kelly and Shogren (2014) implemented the Self-Determined Learning Model of Instruction (SDLMI; Wehmeyer et al., 2000), a general curricular overlay where students learn self-determination via (a) setting goals, (b) developing action plans (i.e., self-monitoring), and (c) self-evaluating progress. Through a multiple-baseline-across participants design, authors reported a functional relation between the SDLMI and improved on-task behavior, as well as generalization and maintenance effects. However, as the SDLMI combines multiple self-regulation and self-determination components, it is unclear what component/s led to greater behavioral change. For example, some students in the Kelly and Shogren's (2014) study demonstrated immediate improvements as soon as they set goals (i.e., prior to self-monitoring), whereas others improved gradually and appeared to benefit from extended self-regulation instruction and implementing self-monitoring. Furthermore, the SDLMI is a resource-intensive method for teaching self-determination, which might be less efficient than targeting specific behavioral self-regulation skills.

Beyond the aforementioned studies, self-regulation interventions with students with EBD in high school are scantly researched (Estrapala, 2020) and it remains unknown whether teaching students to set goals, self-monitor, or both result in greater changes in behavior. Although self-monitoring has a robust evidence base across ages and settings (Carter et al., 2011), it is also more intense in terms of frequency and dosage and usually involves more complex procedures, planning, and materials, which may prohibit widespread adoption in schools (Lloyd et al., 2019). Alternatively, if students can improve their engagement through setting their own behavioral goals, students and teachers may find the parsimony more palatable. Thus, the purpose of this study was to compare the effects of student-determined daily goal reminders and self-monitoring on the AE of high school students with EBD who were receiving special education services.

Research questions included the following:

Research Question 1: To what extent does student AE differ when reading daily goal reminders versus self-monitoring?
Research Question 2: To what extent does student engagement generalize to new settings and maintain after the intervention is withdrawn?
Research Question 3: How do teachers and students perceive the social validity of the self-regulation intervention?

Method

Participants

Following institutional review board (IRB) approval, four special education teachers consented to participate. Each consented teacher nominated up to three of their students to participate and sent home parental consents for each nominated student. Students were eligible to be nominated if they (a) attended Grades 9 to 12, (b) received special education services (as "eligible individuals" in a noncategorical state), (c) could read and write simple sentences (as reported by the teacher), and (d) demonstrated frequent and persistent low rates of AE (as reported by the teacher) in an academic class. Four parental consents were signed and returned for four students and all students were nominated for their behavior in general education classes; thus, four additional general education teachers consented. Next, the four students were systematically screened to confirm the presence of low AE. First, teachers completed the Social Skills Intervention System Ratings Scales–Teacher Form (SSIS; Gresham & Elliott, 2011). Students who scored below average on social skills or above average on problem behavior moved to the second phase of screening. Three students met SSIS screening criteria. Second, the research team conducted systematic direct observations (SDO) of AE. Students were required to average 65% or lower AE across three 15-min observations via duration recording (see data collection and recording below). Once students met screening criteria, they provided written assent to participate. At the time of assenting, students were told they

would receive a US$25 gift card to the restaurant of their choice for completing the research study.

Students and teachers. LeeLee was a 15-year-old Black female in Grade 10. She had Individualized Education Program (IEP) goals for reading, writing, and math. LeeLee attended all general education classes, with the exception of one daily period of Learning Support. She did not have a Behavior Intervention Plan (BIP) but was nominated due to behaviors such as singing, putting her head down, and being out of her seat. The primary teacher of LeeLee's target class (i.e., Teacher 1) was a White female special education teacher with 14 years of teaching experience, a master's degree, and was certified in elementary education, Special Education Strategist 1: K–12, and reading: K–12. LeeLee's target class also included a co-teacher (i.e., Co-Teacher 1), who was a White female, had 34 years of teaching experience, a master's degree, and was certified in math (Grades 6–12), computer science, and driver's education. LeeLee's generalization class was taught by a biology teacher.

Ben was an 18-year-old White male in Grade 12. He had IEP goals for reading, writing, financial literacy, health literacy, and behavior. He also had a BIP, which targeted physical and verbal outbursts (e.g., task refusal, inappropriate physical contact, destruction of property). Ben spent the majority of his day in special education classes, with the exception of Personal Finance. The teacher for Ben's target class (i.e., Teacher 2) was a White male, had 13 years of teaching experience, completed all but dissertation toward his PhD, and was certified in business and finance. Ben's generalization class was taught by his special education teacher.

Jared was a 17-year-old Black male in Grade 11. He had IEP goals for reading, writing, and executive functioning and did not have a BIP. His teacher noted persistent behaviors such as off task, inappropriate phone use, and work avoidance as interfering with his AE. He spent the majority of his day in general education classes, except for two sections of Learning Support. The teacher of Jared's target class (i.e., Teacher 3) was a White female, had 7 years of teaching experience and a master's degree, and was certified in high school English and as an elementary teacher. Jared's generalization class was taught by his special education teacher.

Setting

The study took place in a large, public U.S. Midwestern high school. The high school had a total enrollment of 1,622 students, with 4.1% identified as Asian, 18% Black, 16.3% Hispanic, 4.9% Multiracial, and 54% to 58% White (as reported on the district website). Furthermore, 39% received free or reduced-price lunches, 6.6% were English language learners, 12.7% received special education services, and the average class size was 27.6 students. Training activities took place in a secluded area (empty classrooms, office, etc.) during students' Learning Support period. Comparative treatment and superior treatment occurred during the student's targeted academic class where they screened in for low rates of AE. Due to scheduling issues, generalization classes were not restricted to general education or an academic content area. The average duration of each class was 46 min.

LeeLee's target class was Introduction to Algebra, which was taught by Teacher 1 and Co-Teacher 1. The class consisted of 18 students, including 10 students who had IEPs, with one shared paraeducator and a one-to-one paraeducator (not assigned to LeeLee). Typical activities included 5 min to complete a warm-up worksheet, 15 to 20 min of whole-group instruction (e.g., demonstrations with fill-in-the-blank notes), and 25 min of independent work time. Both classroom teachers frequently redirected LeeLee's off-task and disruptive behaviors. LeeLee's generalization class was Biology, a co-taught general education class.

Ben's target class was Personal Finance, which was taught by Teacher 2. The class consisted of 24 students, including five students with IEPs, two shared paraeducators, and a one-to-one paraeducator (not assigned to Ben). Typical activities included 20 min of whole-group instruction (e.g., lecture and slide show), followed by 25 min independent or group work time. Ben sat at a table with two other students with IEPs, and two paraeducators typically helped Ben and his tablemates complete assignments and stay on task. Generalization settings were not restricted to academic classes, and Ben's generalization class was Learning Support taught by his special education teacher.

Jared's target class was U.S. Literature, an English class taught by Teacher 3. This class had 33 students, including two who had IEPs, and no paraeducators. Class routines varied and typically started with 10 min of silent reading or free writing, followed by other activities (e.g., class-wide review games, videos, whole-group instruction, independent work time). Jared frequently slept or looked at his phone during class, but he was rarely redirected by the teacher. Jared's generalization class was Learning Support, taught by his special education teacher.

Data Collection and Recording Procedures

Independent variables. The principal investigator (PI; first author) trained each student, with regard to their targeted class, to (a) identify problem and replacement behaviors, (b) write a goal statement using a replacement behavior, and (c) determine interval-based self-monitoring procedures to promote replacement behavior use. Students then randomly alternated between a carryover baseline and two different treatments: goal reminders and self-monitoring.

The goal-reminder treatment included reading their goal statement at the start of class and initialing a goal sheet; this was their only interaction with their goal for the class period. On self-monitoring treatment days, students self-monitored using a cueing device that vibrated on a fixed interval (e.g., MotivAider). When students were cued, they read a question asking whether they engaged in their replacement behavior during the previous interval and marked their answer on a self-monitoring sheet. Additional intervention details are provided in next sections.

Dependent variable. The dependent variable, AE, was defined as the student engaging in the assigned task. Examples included eyes attending to task or speaker, using appropriate websites, raising hand to participate, discussing content-related topics with teacher or peers, and working on assigned task. Students were considered engaged if they were following classroom norms or expectations during downtime or transitions (e.g., keeping hands to self while standing in line, quietly chatting with neighbors when allowed). Nonexamples of AE included sleeping, using cell phone, staring into space, using inappropriate websites, talking about things other than assigned task, and working on unassigned classwork. Students were also considered disengaged if they were behaving inappropriately during downtime (e.g., talking when they were supposed to be silent, walking around when they were supposed to be in their seat).

Using duration recording during SDO, we collected AE for two consecutive, 15-min observations (i.e., 30 min) each day throughout the study (with one brief 5-min break in between). We averaged the two 15-min observations at the end of the observation session. We collected data daily throughout each phase of the study (e.g., baseline, student training, comparative treatment, superior treatment, and maintenance) during the student's targeted academic class. We reported AE data as the percentage of time during the observation the student was engaged (e.g., number of seconds engaged divided by total seconds of observation, multiplied by 100). We paused observations when the student temporarily left the room (e.g., bathroom break). If the observations totaled less than 25 min because the student left the room, we discarded data and completed observations the following day. Finally, we collected data only when at least one typical classroom teacher was present and no testing was occurring.

Observer training and interobserver agreement (IOA). The PI served as the primary observer and data collector, and two research assistants (RAs) served as secondary observers. Prior to screening observations, we trained by studying the operational definition of AE, including examples and nonexamples. Then, we practiced collecting data using a video example of a student with low AE until we were at least 90% reliable for two consecutive viewings. Next, we observed a student in a classroom using AE data collection procedures until we achieved at least 90% agreement across two consecutive, 15-min sessions.

We calculated IOA using the total agreement method (e.g., smaller percentage of AE divided by larger percentage of AE multiplied by 100; Ledford & Gast, 2018) for each 15-min observation and then averaged IOA for the full 30-min session. We collected IOA data across a minimum of 20% of observations across all conditions (except generalization probes) for each participant. Mean IOA for AE was 98.2% (range = 92.6%–99.9%).

Descriptive measures

Screening measure. We used the SSIS Teacher Form to confirm the nominated student demonstrated problem behavior warranting intervention. The SSIS is a norm-referenced, reliable, and valid instrument that assesses three domains: social skills, problem behaviors, and academic competence (Gresham & Elliott, 2011). The SSIS includes teacher, parent, and student self-report forms. Teacher and parent forms include social skills or behavior statements alongside a 4-point rating scale of *never*, *seldom*, *often*, and *almost always*. Students indicated on a 4-point scale whether the social skills or behavior statement is *not true*, *a little true*, *a lot true*, or *very true*. All reports included overall standard scores and percentile ranks for social skills and problem behavior. Whereas only the teacher form was used as a screening measure, we used all three forms to pinpoint problem behaviors and as a postintervention descriptive measure. We did not use academic competence scores.

Social validity. We assessed teacher perceptions of the intervention using the Intervention Rating Profile for Teachers (IRP-15; Martens et al., 1985), a 15-item rating scale that has evidence of reliability and validity (Cronbach's $\alpha = .98$; Martens et al., 1985). Teachers rate their level of agreement with each item (e.g., 1 = *strongly disagree*, 2 = *disagree*, 3 = *slightly disagree*, 4 = *slightly agree*, 5 = *agree*, 6 = *strongly agree*). Scores can range from 15 to 90, and scores above 60 represent overall acceptability of the intervention. During a postintervention debriefing, the PI met with the teachers of the target classes and discussed the training, presented all student-created intervention materials (e.g., goals and self-monitoring procedures), reviewed their students' data, and asked teachers to reflect on any changes in student behavior.

Students completed the Children's Intervention Rating Profile (CIRP; Turco & Elliott, 1986), both pre-and post-intervention. The CIRP has five statements (e.g., "This intervention is a fair way to help me"), and students rate their agreement on the same scale as the IRP-15. Students completed the preintervention CIRP after the training sequence but prior to using either intervention during class,

and they completed the postintervention CIRP at the end of the study during a debriefing session. During this debriefing session, the PI reviewed the training sequence, both the interventions the students designed and their data. Students also answered five open-ended questions (e.g., "Did you prefer daily goal reminders or self-monitoring and why?").

Treatment fidelity. We collected treatment fidelity data using researcher-created checklists during 100% of classroom observations across comparative treatments and superior phases of the study. Observers, either the PI or RAs, collected treatment fidelity immediately following all observations. Items aligned directly to key components of intervention procedures. We checked "yes" or "no" to indicate whether a component was observed, and a note was made for each item that required a verbal or physical prompt to be completed. To calculate fidelity, we totaled the number of yeses, divided by the total number of items (i.e., yes + no), and multiplied by 100, and the number of prompts required were totaled daily. Goal reminder fidelity averaged 99% (range = 75%–100%) and self-monitoring fidelity was 100%.

Experimental Design and Analysis

We used a multiple-probe design with embedded alternating treatments to evaluate the differential effects of goal reminders versus self-monitoring on AE (Ledford & Gast, 2018). Study phases included baseline, training, comparative treatments, and maintenance.

Baseline. We collected baseline probes during business-as-usual conditions to contextualize any behavior changes that occurred during treatment. The staggered introduction of training across participants (e.g., multiple baseline) allowed for comparisons across participants from baseline to training and intervention (i.e., intersubject replication) while controlling for threats to internal validity (Kratochwill & Levin, 2010). The order in which students entered training and the number of days spent in baseline were randomly selected prior to collecting baseline data. To determine when students would enter training, we implemented an adapted block randomization strategy using a random number generator with Microsoft Excel (Koehler & Levin, 1998). First, we determined a range of school days each student needed to stay in baseline (e.g., between three and six for LeeLee, between 10 and 12 for Ben, and between 16 and 18 for Jared), and randomly selected a day within those ranges that identified the day students would enter training. The range of 3 to 6 days for Jared was determined because it encompassed the recommended three to five baseline data points required per design standards (Horner et al., 2005). We selected subsequent ranges to (a) eliminate the possibility of overlapping start dates, (b) allow a minimum of 3 days of treatment to demonstrate an intervention effect for the prior student, and (c) ensure Ben and Jared received interventions in a reasonable amount of time.

We chose to randomize treatment start days because we predicted there may be no clear initial response pattern once students entered treatment, which could unnecessarily delay treatment for Ben and Jared (Manolov & Onghena, 2018). That is, as treatment started with several days of training, during which students were learning intervention procedures and not implementing them during their target class, we were unsure whether students would show any treatment effects. This method of randomly selecting start dates still allowed the possible demonstration of intervention effects at different points in time across participants while maintaining methodological rigor and enhancing internal validity (Levin et al., 2019). Incorporating randomization into single-case research designs enhances rigor and validity in the same way randomization enhances group designs: randomly assigning participants to treatment or control groups—or randomly assigning the transition from control (i.e., baseline) to treatment—eliminates the influence of extraneous variables on the dependent variable (Kratochwill & Levin, 2010; Levin et al., 2019). In other words, if the participant demonstrates a clear need for intervention (i.e., low AE) and the intervention is powerful enough to change behavior, the transition from baseline to intervention can—and should—demonstrate an effect regardless of when it is introduced.

Training. Student training activities took place after baseline and before the comparative treatments phase. The AE data collection continued throughout training, as training activities took place outside of the student's target academic class (during learning support, study hall, homeroom, advisory, etc.). The PI trained students in goal setting, self-monitoring, and intervention procedures. Student training activities were arranged into three one-on-one lessons, each lasting from 20 to 40 min. All lessons were supplemented with researcher-created worksheets. Students responded to worksheet items verbally and the PI transcribed responses verbatim. During student training, an RA completed a fidelity checklist for 44% of training sessions (i.e., at least one per student) and mean fidelity was 99.2% (range = 96.8%–100%).

Lesson 1: Identifying problem and replacement behaviors. Lesson 1 began by introducing the intervention and teaching the student to identify problem and replacement behaviors using the teacher, parent, and student self-report results of the SSIS. Students were taught to (a) describe their social and behavioral problem areas using the SSIS results, (b) identify and describe replacement behaviors, (c) prioritize the most impactful behaviors, and (d) describe how engaging in the replacement behaviors can help them succeed in

class. Students prioritized problem behaviors and social skills deficits by identifying subscales on the SSIS that multiple raters identified as problematic. Students then provided specific examples from class to define each problem behavior and alternative replacement behaviors.

We were unable to obtain the SSIS Parent Forms from LeeLee prior to starting Lesson 1, so she identified problem and replacement behaviors based on the Student and Teacher Forms (Gresham & Elliott, 2008). LeeLee identified externalizing, hyperactivity/inattention, and self-control as three important problem behaviors and selected follow directions, be engaged, and ignore as the corresponding replacement behaviors. Like LeeLee, we were unable to obtain Ben's SSIS Parent Forms prior to starting Lesson 1, so he identified his problem and replacement behaviors using only the Student and Teacher Forms. Ben's special education teacher expressed concerns with Ben's history of noncompliance and aggressive behaviors, and the sensitive nature of the lesson content, thus a paraeducator assisted with lesson delivery. Ben identified self-control, hyperactivity/inattention, and internalizing as problem areas, and stop and think, pay attention, and be positive as replacement behaviors. Finally, Jared's foster father completed the preintervention SSIS Parent Form prior to starting Lesson 1, so Jared used all three forms to identify problem and replacement behaviors. Jared identified communication, engagement, and self-control as problem behaviors, and asking for help, pay attention, and wait as replacement behaviors.

Lesson 2: Writing specific and measurable goals. Lesson 2 involved teaching the student (a) the importance of setting goals, (b) how goal setting can positively impact their behavior and education, and (c) how to write specific, measurable goals using replacement behaviors identified in Lesson 1. The PI guided students through writing specific, measurable goals using the replacement behaviors generated in Lesson 1. Students were encouraged to write goals that were achievable in a single class period and positively stated. After students wrote their goals, the PI reviewed them to insure they included sufficient detail to be specific and measurable and made sense given the context of the target class. By the end of the lesson, students wrote a goal to use as a prompt during the goal-reminder treatment condition (e.g., Treatment B). LeeLee's self-written goal statement focused on the replacement behavior of following directions and was, "I won't argue 100% of the time during math, and if I feel mad I will count to 10." Although the PI encouraged LeeLee to write her goal statement with positive language (e.g., what she should do, rather than what she should not do), she preferred to keep her goal statement focused on *not* arguing with her teacher. Ben used the replacement behavior "paying attention" to write his goal statement, "I will pay attention at least 80% of the time during Personal Finance." Jared chose to address communication, writing, "During English, I will ask my teacher or classmate a question when I need help, at least twice per class period."

Lesson 3: Self-monitoring. During Lesson 3, students learned the importance of self-monitoring and the PI helped them customize an interval-based self-monitoring system aligned to the replacement behaviors identified in lesson one. Students (a) wrote a question they would ask themselves during self-monitoring, (b) determined examples and nonexamples of their behavior to self-monitor, (c) selected the interval length, and (d) chose between a Likert-type scale (0 = *never*, 1 = *sometimes*, 2 = *all of the time*, etc.) or a dichotomous yes/no to rate their behaviors. The PI reviewed student choices to ensure they fit with the class context, were feasible, and aligned to the student's replacement behaviors identified during Lesson 1. This information was transcribed to a self-monitoring sheet to use during the self-monitoring condition (see Treatment C). Last, students practiced self-monitoring examples and nonexamples of their behaviors using their data tracking sheet and MotivAider (i.e., a small electronic device that vibrates to cue the user to self-monitor on a preset interval schedule).

LeeLee wrote the question, "Am I doing what I'm supposed to be doing?" which she asked herself when prompted by the MotivAider. Examples of behaviors supporting following directions included "doing my work, doing what the teacher tells me to do, count to 10 when I'm upset, and leave the room." She selected a 6-min interval and a 0, 1, or 2 rating scale to record the degree to which she was paying attention. Ben wrote the question, "Am I paying attention?" Examples of paying attention included looking at the teacher, looking at the screen, or looking at the work, and ignoring others. Ben chose a 7-min interval and a yes/no dichotomy to record his behavior when prompted. Jared wrote the question, "Did I ask for help if I needed it?" to self-monitor his behavior. Examples included to raise his hand and wait for the teacher or ask someone near him. Jared chose a 7-min interval and a yes/no response type.

Comparative treatments. Comparative treatments included carryover baseline (Treatment A), goal reminders (Treatment B), and self-monitoring (Treatment C). Treatments were alternated daily during the student's target class period. We randomized the order of Treatments A, B, and C within blocks of conditions (Kratochwill & Levin, 2010). Randomization was necessary to control for sequencing effects (i.e., the order of treatments) by randomly varying the order in which the student received each treatment. Each block contained treatments A, B, and C, and randomizing within blocks was necessary to prevent a student from receiving multiple treatments consecutively (e.g., AAAAABBBBBCCCCC). Treatments were counterbalanced to ensure students experienced each treatment the

same number of times across the comparative phase. Students received each treatment five times for a total of 15 days in comparative (Manolov & Onghena, 2018). Randomization occurred a priori, and neither students nor teachers knew the schedule of treatments prior to the start of class. Instead, research staff prepared folders with the necessary materials (e.g., no materials for Treatment A [baseline], goal sheets for Treatment B, and self-monitoring forms and MotivAider for Treatment C) each day prior to class starting. Students retrieved the materials from a designated location in the classroom each day (e.g., the teacher's desk, storage cabinet).

Treatment A. Treatment A was a continuation of baseline; students received no goal reminder and did not self-monitor, and students followed business as usual procedures. The purpose of continuing baseline through comparative treatments was to help detect carryover effects that occur when "experiencing one treatment enhances the effect of another—making it appear more effective than it would be if used alone" (Holcombe et al., 1994, p. 136). If baseline AE systematically changed in relation to either Treatment B or C during the comparative phase (e.g., baseline AE was always higher after Treatment B), carryover effects were likely present.

Treatment B. In Treatment B, students retrieved a folder at the start of class with their materials (e.g., goal sheets, pencil) and returned to their desk. Then, the student opened the folder, read their goal statement, wrote the date, signed their name, and closed the folder. This process served as their goal reminder and was their only interaction with their goal for the period. At the end of class, they returned the folder to its designated location.

Treatment C. At the start of Treatment C days, students retrieved their materials (e.g., self-monitoring form, pencil, and MotivAider) and began self-monitoring (using procedures developed during training) that continued throughout the entire class period. At the end of class, students returned their self-monitoring materials to their designated location. They did *not* total their scores at the end of each day (e.g., the number of yeses or points circled).

Superior treatment. Superior treatment occurred across five consecutive days, with students implementing the treatment that yielded higher rates of AE during comparative treatments. If there was no clear distinction between treatments as evidenced by no separation of data paths or mean difference of AE, then the student chose which treatment they would prefer to use for the next 5 days.

Maintenance and generalization. We collected at least one generalization probe during each phase of the study (e.g., baseline; training; comparative Treatments A, B, and C; and superior alone) during nontarget classes to determine the extent to which behavioral changes extended to different settings. Generalization probes followed the same procedures for collecting data during the student's target class. With the exception of LeeLee starting on Session 30, interventions were not implemented in generalization settings. Furthermore, to determine the extent to which behavior changes maintained after the intervention was withdrawn, we collected at least three probes within 2 weeks of the conclusion of the superior phase.

Analysis. We visually analyzed graphed data within and across phases for stability, level, and trend from baseline to training, training to comparative treatments, and baseline to training and each treatment individually, as well as the immediacy of effect and overlap between each phase. A functional relation for overall effects of each intervention was determined if all participants demonstrated immediate or gradual increases in AE once entering intervention (i.e., intersubject replication). Within the comparative treatment phase, we compared the level and trend of Treatments B and C to determine whether the student responded more favorably to one treatment versus the other, and how Treatments B and C compared with the continuing baseline, or Treatment A. A functional relation for the alternating treatments design was established if clear differentiation between Treatments B and C was detected at least 5 times across a single participant's data (i.e., within-subject replication; Horner et al., 2005; Manolov & Onghena, 2018). If one treatment was clearly superior to the other after 15 days of intervention, the student entered the superior phase. Finally, we calculated within-phase mean AE for each treatment condition for each participant.

We used an online Tau-U calculator (www.singlecaseresearch.org/calculators/tau-u) to calculate effect sizes. We chose Tau-U because it is well suited for small data sets, combines nonoverlapping data between phases, accounts for trend within phases, and accounts for baseline trends (Vannest et al., 2016). However, Tau-U does not account for magnitude of change when there is no overlap between phases and values may exceed ± 1 (Parker et al., 2011). For each participant, we calculated Tau-U between baseline and goal reminders, and baseline and self-monitoring. We did not include data from the training phase. Next, we calculated combined effects (e.g., across participants) for each intervention. We interpreted effects as small (0.0–0.20), moderate (0.20–0.60), large (0.60–0.80), or very large (0.80 and above; Vannest & Ninci, 2015).

We analyzed social validity data using descriptive statistics (i.e., mean, range). On the open-ended questionnaire, responses were totaled and some responses were quoted in full to provide further details. From the SSIS, we reported overall social skills and problem behavior as percentile ranks provided by the scoring manual.

Results

Academic Engagement

Overall, we detected no functional relations (see Figure 1). First, we observed only two moderate demonstrations of effect from baseline to training and goal reminders (i.e., Ben and Jared). Both Ben and Jared demonstrated reversal in trend (i.e., countertherapeutic to therapeutic; see online supplemental Figures S2 and S3), but exhibited no immediate change in level and intervention data overlapped with baseline. Across participants, overall effect size for goal reminders was moderate (Tau-U = 0.52). Second, only one demonstration of effect emerged from baseline to training and self-monitoring (i.e., LeeLee). Compared with the last three baseline points, LeeLee demonstrated trend reversal (i.e., countertherapeutic to therapeutic; see online supplemental Figure S1); however, those effects were hampered by overlap with baseline. Furthermore, self-monitoring yielded moderate effects (Tau-U = 0.53) across participants. There was no functionally superior treatment based on analysis of the alternating treatments component.

LeeLee. LeeLee's AE during baseline was variable across Sessions 1 through 3, with a stable, decreasing trend across the last three data points. During the training phase of treatment, there was a steep increase in trend from the first to last day of training (M = 77.5%; range = 58.9%–96%). Within the comparative treatments phase, there was no clear separation of data paths given the overlap between carryover baseline, goal reminders, and self-monitoring. Furthermore, data paths for goal reminders (M = 78.5%; range = 58.8%–92.3%) and self-monitoring (M = 78.1%; range = 60.4%–87.5%) were moderately variable throughout the comparative treatments phase, and carryover baseline remained the most stable condition (M = 76.5%; range = 70.7%–83.4%), although all treatments were superior to baseline in terms of average AE. Trend line analysis indicated countertherapeutic trends for both carryover treatment and goal reminders, with a therapeutic trend for self-monitoring. As there was no clear superior treatment as determined by visual analysis and mean difference, LeeLee chose to self-monitor during the superior phase. Her mean AE during superior was 93.3% (range = 88.4%–96.7%). Relative to the comparative treatment phase, there was an increase in level and stability. Statistically, LeeLee demonstrated a moderate improvement in AE with goal reminders (Tau-U = 0.44) and slightly stronger effects with self-monitoring (Tau-U = 0.52).

LeeLee's mean AE during maintenance was 84% (range = 76.2%–89.8%). She demonstrated low, stable levels of AE across generalization probes (M = 27.8%, range = 24.5%–31.1%) during baseline, training, comparative treatments, and superior phases, indicating no generalization effects. When asked by the PI and her special education teacher if she would try self-monitoring during her generalization class (during the maintenance phase), LeeLee agreed. When she began self-monitoring during her generalization class, her mean AE was 87.9% (range = 81.3%–95.4%), with a slight downward trend. On the SSIS, LeeLee's teacher rated improvements in problem behavior and social skills from pre- to postintervention, whereas LeeLee rated herself with improved problem behaviors, but lower social skills (see Table 1).

Ben. Ben's AE data during baseline were moderately stable with an overall decreasing trend. His mean AE was 71.8% (range = 63.7%–83%) during baseline and decreased to 43.9% during training. Within the comparative treatments phase, there was no clear separation of data paths as there was overlap between carryover baseline, goal reminders, and self-monitoring. Data paths for goal reminders (M = 80.2%; range = 65%–93.1%), self-monitoring (M = 82.1%; range = 69.3%–94.5%), and carryover baseline (M = 76%; range = 63.1%–86.5%) were moderately variable throughout the comparative treatments phase, although all superior to baseline. The AE trend improved across comparative conditions relative to baseline. As mean AE for self-monitoring was higher than goal reminders, Ben self-monitored during the superior phase. His mean AE during superior was 82.9% (range = 69.2%–94.5%), and visually, his AE increased for three consecutive days and then began to slightly decrease for the last 2 days. Statistically, Ben demonstrated a moderate improvement in AE with goal reminders (Tau-U = 0.52) and large effects with self-monitoring (Tau-U= 0.64).

Ben's mean AE during maintenance was 85.1% (range = 74.5%–96.6%), with an overall increasing trend, indicating maintenance effects postintervention. Generalization data remained stable and higher than his target class (M = 92%; range = 89.8%–%) across baseline, training, and comparative treatments. Thus, we did not program for generalization. On the SSIS, Ben's teacher reported improvements in problem behavior and social skills from pre- to postintervention, and Ben rated himself with improved social skills but increased problem behavior.

Jared. Jared's baseline data were highly variable with an overall decreasing trend. Mean AE across eight baseline observations was 49.5% (range = 15.7%–76.8%) and increased to 89.9% during the training phase of treatment. During comparative treatments, there was no clear separation of data paths as there was overlap across treatments. Data paths for self-monitoring (M = 65.4%; range = 31.4%–59.8%) and carryover baseline (M = 58.6%; range = 32.9%–98.9%) were variable throughout comparative treatments, with an increasing trend during carryover

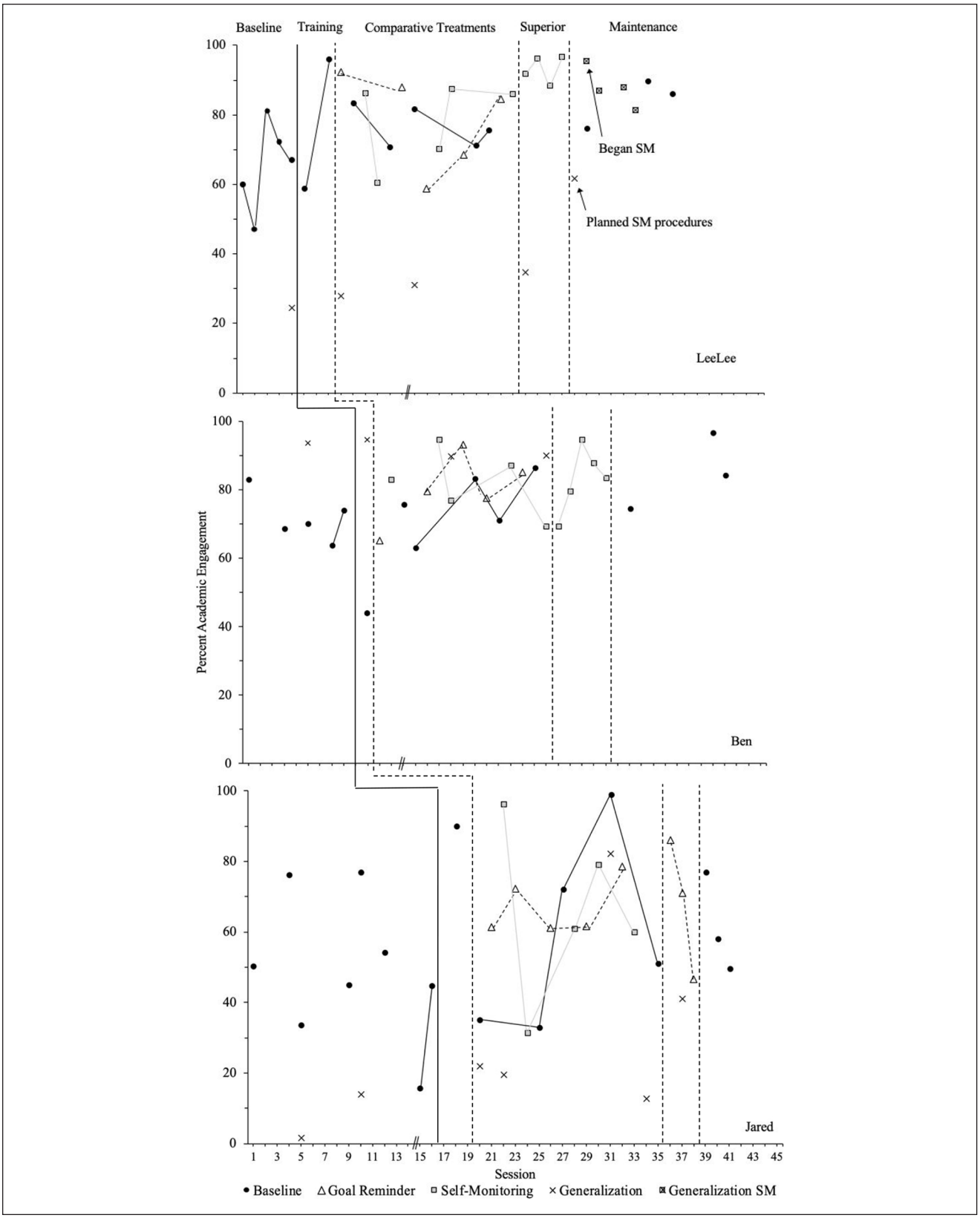

Figure 1. Percentage of academic engagement across participants.
Note. X-axis break represents 2-week recess from school; SM = self-monitoring.

Table 1. SSIS Pre–Post Results.

		Teacher			Student			Parent		
Student	SSIS domain	Pre	Post	Change	Pre	Post	Change	Pre	Post	Change
LeeLee	Social skills	44	71	+27	8	4	−4	4	2	−2
	Problem behavior	92	70	−22	96	89	−7	85	91	+6
Ben	Social skills	12	35	+23	45	75	+30	7	[a]	NA
	Problem behavior	96	74	−22	80	85	+5	96	[a]	NA
Jared	Social skills	8	27	+19	24	[a]	NA	12	8	−4
	Problem behavior	97	76	−21	29	[a]	NA	75	77	+2

Note. Scores reported as percentile rank. Higher scores and increases on social skills preferred. Lower scores and decreases on problem behaviors preferred. Female norms used for LeeLee; male norms used for Ben and Jared. SSIS = Social Skills Intervention System; + = increase from pre to post; − = decrease from pre to post.
[a]Scores not available.

baseline and a decreasing trend during self-monitoring. Goal reminders yielded the most stable AE (M = 67%; range = 61.2%–78.4%). As mean AE for goal reminders was higher than self-monitoring, Jared implemented goal reminders during the superior phase (M = 67.9%; range = 46.7%–86%). Statistically, Jared demonstrated large improvements in AE with goal reminders (Tau-U = 0.60) and moderate effects with self-monitoring (Tau-U = 0.45). No baseline corrections were used in Tau-U calculations. Due to a sharply decreasing trend, the superior phase was stopped after three observations.

Jared's mean AE during maintenance was 61.5% (range = 49.5%–77%) with a steep downward trend. Jared maintained low levels of AE across generalization probes (M = 27.6%, range = 1.6%–82.1%), with the exception of Session 21 (i.e., 82.1%). When asked whether he would like to try either self-monitoring or goal reminders during his generalization class, Jared refused. Thus, we did not implement generalization programming. On the SSIS, Jared's teacher rated improvements in problem behaviors and social skills from pre- to postintervention. Changes in Jared's self-scores were uninterpretable as he did not take the post-SSIS.

Social Validity

In terms of social validity, student mean preintervention rating across students and items was 4.73 (range = 4–6), and mean postintervention rating was 4.67 (range = 1–6). Mean postintervention rating for teachers was 5.73 (range = 4–6).

LeeLee's preimplementation rating of the intervention was favorable (23 of 30) and her postintervention rating increased to 24. During postimplementation debriefing, LeeLee stated self-monitoring made a greater difference with her behavior because it kept her focused, but goal reminders were easier to use. LeeLee's teacher (i.e., Teacher 1) also rated the intervention favorably (e.g., 87 of 90). During postintervention debriefing, Teacher 1 asked for a digital copy of all intervention materials because she wanted to implement the intervention with other students on her caseload.

Ben's preimplementation rating was favorable (27 of 30), although his postintervention rating decreased to 25. During postimplementation debriefing, Ben stated self-monitoring made a greater difference with his behavior because it kept him focused and it was easier to use than goal reminders. Ben's teacher (i.e., Teacher 2) also rated the intervention favorably (e.g., 85 of 90 points).

Jared's preimplementation rating was slightly favorable (21 of 30). His postimplementation rating remained at 21 and he rated *strongly disagree* in response to the statement, "This intervention helped me feel better about myself." He strongly agreed that the interventions were easy to participate in and the process helped him be successful in class. He found self-monitoring the most helpful intervention because it repeatedly prompted him to focus and he found the goal reminders easier because he only had to interact with the intervention at the start of class. Jared's teacher rated the intervention favorably (e.g., 86 of 90 points).

Discussion

Goal setting and self-monitoring are two empirically supported self-regulation strategies for increasing AE (Bruhn et al., 2016, 2020; Carter et al., 2011); however, their differential impact as stand-alone interventions remained unexplored. We hypothesized that simple, low-intensity interventions such as student-set behavioral goals with daily reminders could be practical, feasible means of improving AE, as could the slightly more intense procedures of self-monitoring. To determine which approach was more effective, we implemented a multiple probe with embedded alternating treatments design with three high school students with low AE. Overall, we detected mixed effects across students from baseline comparative treatments, and neither treatment was functionally superior to

the other. We detected no generalization effects, and maintenance effects were apparent in two participants. All teachers felt the interventions were effective and reasonable given the students' behavior. Students felt the interventions were moderately effective and reasonable, and all preferred self-monitoring to goal reminders.

Implications and Directions for Future Research

Research Question 1: Comparing goals and self-monitoring. Although no treatment yielded a functionally superior effect on AE in this study, several variables likely influenced these results. First, there was substantial variability in classroom management practices across participants. Research in tiered prevention models (e.g., Positive Behavioral Interventions and Supports [PBIS]; Horner et al., 2010) suggests first implementing class-wide antecedent strategies (i.e., classroom management) with fidelity prior to any individualized interventions to maximize opportunities for appropriate behavior (Kern & Clemens, 2007). Moreover, emerging evidence indicates that implementing evidence-based classroom management practices (e.g., clear expectations, frequent praise, instructional pacing, consistent routines) improves response to individualized interventions (Van Camp et al., 2020). This was evident with LeeLee, as many strategies were observed in her targeted class, and perhaps LeeLee had the greatest opportunity to demonstrate change due to the supportive classroom environment. Ben's target class, by comparison, exhibited very few classroom management strategies, and Ben's AE appeared to vary with unpredictable class behaviors. Interestingly, Ben's generalization class was predictable and organized and he demonstrated consistent, high rates of AE. Similarly, class context for Jared also appeared to impact his AE, as the difficulty of the curriculum and the pace of instruction may have been beyond his academic abilities (e.g., standardized writing assessment scores were in the 2nd percentile, reading comprehension scores 40% lower than general education peers). In addition, students with low self-regulatory ability struggle to control their goal-oriented behavior when confronted with environmental challenges, which is particularly true for adolescents in social environments (Hagler et al., 2016). Taken together, it is likely that classroom management had a substantially confounding effect on AE, which implies that targeting self-regulation alone may not be powerful enough for students with engagement deficits to overcome environmental influences. Future research should attempt to control these environmental factors to better isolate the effects of self-regulation instruction and practice on student behavior, and measure the level of intensity (e.g., frequency and dosage) needed to produce desired effects.

Relatedly, given that Jared displayed the greatest behavioral deficits of the three participants, based on his SSIS scores and baseline AE, it is possible that neither intervention was sufficiently intense to change his behavior. Additional components, such as self-evaluation, behavior-specific feedback, or contingent reinforcement, could have made a greater impact, as researchers have found that including feedback or reinforcement in self-monitoring interventions significantly improves AE ($p < .001$; Bruhn et al., 2020). Given the empirical evidence supporting these methods of intensifying self-monitoring interventions, Jared may have demonstrated stronger improvements had those components been included. Future researchers could study the additive effects of other self-regulation strategies, feedback, and reinforcement with a range behavioral needs to establish an evidence-based range of intervention intensity.

Research Question 2: Maintenance and generalization. Modest maintenance effects were apparent across all participants, similar to related studies (e.g., Clemons et al., 2016; Kelly & Shogren, 2014). Interestingly, once LeeLee entered the superior phase and used the same intervention (e.g., self-monitoring) continuously, her data stabilized relative to the comparative treatments phase. Perhaps providing consistency (e.g., the same intervention rather than alternating randomly) gave LeeLee the opportunity to develop self-efficacy toward her ability to engage in class content (Covington, 2000). In other words, it is possible that experiencing repeated (rather than intermittent) success helped her behavior and motivated her to continue being highly engaged. This is supported by high AE during maintenance and her substantial improvements in AE when self-monitoring in her generalization class. Unfortunately, no students exhibited generalization unless provided with explicit programming (e.g., LeeLee). In contrast, results from Kelly and Shogren (2014) indicated strong generalization of behaviors across all students without generalization programming. It is possible that the high-intensity intervention (e.g., SDLMI) contributed to these findings. Future research should explore the interaction between intervention intensity, maintenance, and generalization.

Research Question 3: Social validity. Given that both goal reminders and self-monitoring yielded improvements in AE, and teachers and students alike found them socially valid, these interventions are promising options for practitioners. First, both LeeLee and Jared stated that goal reminders were easier to participate in compared with self-monitoring. This is important because the goal reminder intervention was less intense than self-monitoring, as goal reminders only required one interaction with the intervention, whereas self-monitoring required several. Goal reminders did not require any equipment beyond paper and a writing utensil, materials commonly found in classrooms, whereas self-monitoring requires a cuing device (e.g., MotivAider or similar), a tool none of the teachers had readily

available. This ease of use, combined with modest improvements in AE, indicates the goal reminder intervention is a viable option for students who struggle with AE, especially when teachers have limited resources. Next, LeeLee and Jared independently implemented both interventions without prompting with 100% fidelity, and Ben only needed prompting to start the interventions. This finding is important as it reinforces the claim that these interventions are acceptable, feasible, and can be reliably implemented with little to no support from interventionists. Future research should examine the effects of similar student-directed self-regulation interventions with practitioners serving as interventionists.

Methodology, Rigor, and Limitations

Before discussing the limitations of our results, we would like to highlight three critical aspects of the research methodology utilized in this study. First, this study represents a rigorous application of single-case research design, with high levels of internal validity and numerous indicators for quality research. In terms of rigor, there were ample opportunities for demonstrations of effects for both the multiple probe (e.g., three tiers, concurrent baselines, three separate start points) and alternating treatments design (e.g., five data points per condition; reversable dependent variable; Ledford & Gast, 2018). Furthermore, three layers of randomization were incorporated a priori to increase internal validity: (a) order of entry to training, (b) phase changes from baseline to training, and (c) order of treatments during comparative treatments. These randomization strategies increased design rigor, reduced bias (Ledford & Gast, 2018), and addressed threats to internal validity (e.g., multitreatment interference, sequencing effects; Kratochwill & Levin, 2010; Manolov & Onghena, 2018), thus reducing the possibility of Type I errors (Lanovaz et al., 2019). Quality indicators such as systematic measuring and reporting of treatment integrity (e.g., fidelity), measurement reliability (e.g., IOA), maintenance and generalization effects, and social validity were also reported according to scientific guidelines (e.g., *Exceptional Children*, Horner et al., 2005, What Works Clearinghouse [WWC] Standards, Kratochwill et al., 2010). Taken together, these elements enhance the scientific credibility of our findings and highlight the importance of these results in light of the movement to publish rigorous single-case studies with low to moderate effects (see Shadish et al., 2016).

Relatedly, carryover effects were detected because AE systematically increased after training during the carryover baseline condition across all participants. Two possible reasons might explain these carryover effects. First, the self-regulation curriculum may have reduced the reversibility of AE because students learned to self-regulate. Second, the goal reminder and self-monitoring interventions may have been too similar in nature to yield a clear separation of data paths (Manolov & Onghena, 2018). That is, as both interventions were focused on self-regulation strategies and based on the same target and replacement behaviors, perhaps both interventions contributed to behavioral changes. Future researchers may consider studying the overall effects of a cohesive self-regulation intervention package (goal setting, self-monitoring, self-evaluation, etc.) rather than studying the comparative effects of each strategy individually, as these strategies are often used in tandem (Bruhn et al., 2020). This could build evidence for the overall efficacy of self-directed, packaged self-regulation interventions.

Although results of this study are encouraging given the methodological rigor, there are a few key limitations worth noting. First, the decision to incorporate randomized starting points increased the risk of implementing intervention before stable data patterns emerged, as randomization removes data-informed decision-making (Ledford & Gast, 2018). Thus, it is possible that a more traditional, response-guided approach could have yielded more convincing data. Second, given the design and sample size, no causal inferences can be made from SSIS changes. Third, although Tau-U is widely used in single-case research and often correlates with visual analysis (Yucesoy-Ozkan et al., 2019), this method has not been widely studied in the context of combination designs (e.g., the present study; Klingbeil et al., 2019); thus, these statistics should only be interpreted within the context of the present study.

Conclusion

Conducting rigorous behavioral research in high schools is challenging but necessary. Students with disabilities who struggle with behavior in general education classes is common, but the availability of evidence-based, feasible interventions is not (Author, in review). Research on adolescent behavior indicates that self-regulation strategies combined with decision-making are developmentally appropriate and advantageous methods for improving problem behaviors in school. However, self-regulation interventions are typically complex and limit opportunities to engage students in decision-making, thus warranting the investigation of two novel approaches to behavioral self-regulation. In this study, no treatment was clearly superior to the other, yet both goal reminders and self-monitoring provided modest improvements in AE across participants and were rated as highly social valid interventions. With further refinement and testing, both interventions represent promising potential for the future of high school behavior intervention development.

Authors' Note

The authors thank Drs. Seth King, Shawn Datchuk, Pam Ries, and Gerta Bardhoshi from the University of Iowa for their meaningful contributions to the design and execution of this study.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research was supported by Grant H325D130012-17 from the U.S. Department of Education Office of Special Education Programs.

Supplemental Material

Supplementary material for this article is available on the *Journal of Emotional and Behavioral Disorders* website with the online version of this article.

References

Barbrack, C. R., & Maher, C. A. (1984). Effects of involving conduct problem adolescents in the setting of counseling goals. *Child & Family Behavior Therapy*, *6*(2), 33–44. https://doi.org/10.1300/J019v06n02_03

Bruhn, A. L., Gilmour, A., Rila, A., Hancock, E., Van Camp, A., Shaeffer, A., Fernando, J., & Wehby, J. (2020). Treatment components and participant characteristics associated with outcomes in self-monitoring interventions. *Journal of Positive Behavior Interventions*. Advance online publication. https://doi.org/10.1177/1098300720946651

Bruhn, A. L., McDaniel, S. C., Fernando, J., & Troughton, L. (2016). Goal-setting interventions for students with behavior problems: A systematic review. *Behavioral Disorders*, *42*(2), 107–121. https://doi.org/10.17988/0198-7429-41.2.107

Carter, E. W., Lane, K. L., Crnobori, M., Bruhn, A. L., & Oakes, W. P. (2011). Self-Determination interventions for students with and at risk for emotional and behavioral disorders: Mapping the knowledge base. *Behavioral Disorders*, *36*(2), 100–116.

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*, 145–155. https://doi.org/10.1177/1098300715596134

Covington, M. V. (2000). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology*, *51*, 171–200. https://doi.org/10.1146/annurev.psych.51.1.17

Estrapala, S. (2020) Behavioral self-regulation: A comparison of goal reminders and self-monitoring (Publication No. 27961255) [Doctoral Dissertation, University of Iowa]. The University of Iowa, ProQuest Dissertations and Theses Database.

Gardner, T. W., Dishion, T. J., & Connell, A. M. (2008). Adolescent self-regulation as resilience: Resistance to antisocial behavior within the deviant peer context. *Journal of Abnormal Child Psychology*, *36*, 273–284. https://doi.org/10.1007/s10802-007-9176-6

Gresham, F., & Elliott, S. N. (2008). *SSIS: Social skills improvement system-ratings scale manual* (1st ed.). M.N. Pearson.

Gresham, F., & Elliott, S. N. (2011). Test review. *Journal of Psychological Assessment*, *29*, 292–296. https://doi.org/10.1177/0734282910385806

Hagler, M., Banyard, V., Grych, J., & Hamby, S. (2016). The ups and downs of self-regulation: Tracing the patterns of regulatory abilities from adolescence to middle adulthood in a rural sample. *Journal of Rural Mental Health*, *40*, 164–179. https://doi.org/10.1037/rmh0000053

Holcombe, A., Wolery, M., & Gast, D. L. (1994). Comparative single subject research: Description of designs and discussion of problems. *Topics in Early Childhood Special Education*, *16*, 168–190. https://doi.org/10.1177/027112149401400111

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

Horner, R. H., Sugai, G., & Anderson, C. M. (2010). Examining the evidence base for school-wide positive behavior support. *Focus on Exceptional Children*, *42*(8), 1–14.

Kelly, J. R., & Shogren, K. A. (2014). The impact of teaching self-determination skills on the on-task and off-task behaviors of students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders*, *22*(1), 27–40. https://doi.org/10.1177/1063426612470515

Kern, L., & Clemens, N. H. (2007). Antecedent strategies to promote appropriate classroom behavior. *Psychology in the Schools*, *44*, 65–75. https://doi.org/10.1002/pits.20206

Klingbeil, D. A., Van Norman, E. R., McLendon, K. E., Ross, S. G., & Begeny, J. C. (2019). Evaluating Tau-U with oral reading fluency data and the impact of measurement error. *Behavior Modification*, *43*, 413–438. https://doi.org/10.1177/0145445518760174

Koehler, M. J., & Levin, J. R. (1998). Regulated randomization: A potentially sharper analytical tool for the multiple-baseline design. *Psychological Methods*, *3*, 206–217. https://doi.org/10.1037/1082-989X.3.2.206

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. https://files.eric.ed.gov/fulltext/ED510743.pdf

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods*, *15*, 124–144. https://doi.org/10.1037/a0017736

Lanovaz, M. J., Caradinal, P., & Francis, M. (2019). Using a visual structured criterion for the analysis of alternating-treatment designs. *Behavior Modification*, *43*, 115–131. https://doi.org/doi.org/10.1177/0145445517739277

Ledford, J. R., & Gast, D. L. (Eds.). (2018). *Single case research methodology: Applications in special education and behavioral sciences*. Routledge.

Levin, J. R., Kratochwill, T. R., & Ferron, J. M. (2019). Randomization procedures in single-case intervention research contexts: (Some of) "the rest of the story." *Journal of the Experimental Analysis of Behavior*, *112*, 334–348. https://doi.org/10.1002/jeab.558

Lloyd, B. P., Bruhn, A. L., Sutherland, K. S., & Bradshaw, C. P. (2019). Progress and priorities in research to improve outcomes for students with or at risk for emotional and behavioral disorders. *Behavioral Disorders*, *44*, 85–96. https://doi.org/10.1177/0198742918808485

Mager, U., & Nowak, P. (2011). Effects of student participation in decision making at school: A systematic review and synthesis of empirical research. *Educaitonal Research Review*, *7*, 38–61. https://doi.org/10.1016/j.edurev.2011.11.001

Manolov, R., & Onghena, P. (2018). Analyzing data from single-case alternating treatments designs. *Psychological Methods*, *23*, 480–504. https://doi.org/10.1037/met0000133

Martens, B. K., Witt, J. C., Elliott, S. N., & Darveaux, D. X. (1985). Teacher judgements concerning the acceptability of school-based interventions. *Professional Psychology: Research and Practice*, *16*(2), 191–198. https://doi.org/10.1037/0735-7028.16.2.191

Parker, R. I., Vannest, K., Davis, J., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, *42*, 284–299. https://doi.org/10.1016/j.beth.2010.08.006

Shadish, W. R., Zelinsky, N. A., Vevea, J. L., & Kratochwill, T. R. (2016). A survey of publication practices of single-case design researchers when treatments have small or large effects. *Journal of Applied Behavior Analysis*, *49*, 656–673. https://doi.org/10.1002/jaba.308

Shogren, K. A., Villarreal, M. G., Lang, K., & Seo, H. (2017). Mediating role of self-determination constructs in explaining the relationship between school factors and postschool outcomes. *Exceptional Children*, *83*(2), 165–180. https://doi.org/10.1177/0014402916660089

Trentacosta, C. J., & Shaw, D. S. (2009). Emotional self-regulation, peer rejection, and antisocial behavior: Developmental associations from early childhood to early adolescence. *Journal of Applied Developmental Psychology*, *30*, 356–365. https://doi.org/10.1016/j.appdev.2008.12.016

Turco, T. L., & Elliott, S. N. (1986). Students' acceptability ratings of interventions for classroom misbehaviors: A study of well-behaving and misbehaving youth. *Journal of Psychoeducational Assessment*, *4*, 281–289. https://doi.org/10.1177/073428298600400404

Van Camp, A., Wehby, J. H., Copeland, B. A., & Bruhn, A. L. (2020). Building from the bottom up: The importance of Tier 1 supports in the context of Tier 2 interventions. *Journal of Positive Behavior Interventions*, *23*, 53–64. https://doi.org/10.1177/1098300720916716

Vannest, K., & Ninci, J. (2015). Evaluating intervention effects in single-case research designs. *Journal of Counseling & Development*, *93*, 403–411. https://doi.org/10.1002/jcad.12038

Vannest, K. J., Parker, R. I., Gonen, O., & Adiguzel, T. (2016). *Single case research: Web based calculators for SCR analysis* (Version 2.0). Texas A&M University. singlecaseresearch.org

Wehmeyer, M. L., Baker, D. J., Blumberg, R., & Harrison, R. (2004). Self-determination and student involvement in functional assessment: Innovative practices. *Journal of Positive Behavior Interventions*, *6*(1), 29–35. https://doi.org/10.1177/10983007040060010501

Wehmeyer, M. L., Palmer, S. B., Agran, M., Mithaug, D. E., & Martin, J. E. (2000). Promoting causal agency: The self-determined learning model of instruction. *Exceptional Children*, *66*(4), 439–453.

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*, 421–434. https://doi.org/10.1007/s10864-014-9204-x

Yucesoy-Ozkan, S., Rakap, S., & Gulboy, E. (2019). Evaluation of treatment effect estimates in single-case experimental research: Comparison of twelve overlap methods and visual analysis. *British Journal of Special Education*, *47*, 67–87. https://doi.org/10.1111/1467-8578.12294