Hammill Institute on Disabilities

Journal of Emotional and Behavioral Disorders
2020, Vol. 28(1) 29–42

DOI: 10.1177/1063426618806281
jebd.sagepub.com
SAGE

Treatment Sensitivity of Direct Behavior Rating–Multi-Item Scales in the Context of a Daily Report Card Intervention

Chelsea L. Hustus, MS[1], Julie Sarno Owens, PhD[1], Robert J. Volpe, PhD[2], Amy M. Briesch, PhD[2], and Brian Daniels, PhD[3]

Abstract

The primary goal of this study was to assess the treatment sensitivity of four newly developed Direct Behavior Rating–Multi-Item Scales (DBR-MIS) that assess the domains of academic engagement, disruptive behavior, organizational skills, and oppositional behavior in the context of a Daily Report Card (DRC) intervention. To achieve this goal, we first evaluated the integrity and effectiveness of the DRC intervention in this sample. Participants included six elementary school teachers, each of whom delivered a DRC intervention with one student from their classroom, while completing DBR-MIS ratings on a daily basis for 2 months. Results confirmed the effectiveness of the DRC intervention (all DRC target behaviors demonstrated improvement, with at least half demonstrating improvement that was moderate to large in magnitude) and revealed a positive relationship between DRC implementation integrity and student outcomes. We found strong evidence for the treatment sensitivity of the DBR-MIS assessing academic engagement, disruptive behavior, and organizational skills. Results for the treatment sensitivity of the DBR-MIS oppositional scale were inconclusive. Implications for progress monitoring using the recently developed DBR-MIS are discussed.

Keywords

progress monitoring, Direct Behavior Rating, treatment sensitivity

In recent decades, elementary school personnel have been addressing students' academic, social-emotional, and behavioral needs through the use of multitiered systems of supports (MTSS; Benner, Kutash, Nelson, & Fisher, 2013). MTSS are proactive models of service delivery in which all students receive the level of support that they need. In successful MTSS models, school personnel provide (a) primary prevention efforts (Tier 1) to support all students' academic, social-emotional, and behavioral functioning; (b) engage in universal screening to identify students in need of secondary, targeted (Tier 2) supports; (c) collect formative assessment data to monitor student progress over time; and (d) use these data to determine student needs and the effectiveness of the given level of support.

A central tenet of MTSS is that the level of intervention intensity should be matched to student need and can be reduced or intensified based on the student's response to a given level of intervention. For example, those students scoring above a particular threshold on a behavioral screening measure might receive a Tier 2 intervention, such as a Daily Report Card (DRC; see Vujnovic, Holdaway, Owens, & Fabiano, 2014, for example). Once the DRC intervention is in place, school personnel engage in progress monitoring to determine if the student needs are adequately supported or if additional, more intensive Tier 3 supports are needed. To effectively determine a student's response to intervention, it is necessary to employ progress monitoring, or the use of repeated assessments, to determine if student needs are adequately supported at a given level. Desirable psychometric characteristics of progress monitoring tools include reliability, validity, and treatment sensitivity (Gresham, 2005). Treatment sensitivity refers to the ability of a measure to detect small changes in behavior as a function of an intervention. Although reliability and validity are relevant across all assessment purposes (e.g., screening, diagnostic assessment), treatment sensitivity is particularly important within a progress monitoring context because the primary question of interest is whether the student is responding to the provided level of support.

[1]Ohio University, Athens, USA
[2]Northeastern University, Boston, MA, USA
[3]University of Massachusetts Boston, USA

Corresponding Author:
Chelsea L. Hustus, Ohio University, Porter Hall 200, Athens, OH 45701, USA.
Email: ch051414@ohio.edu

Unfortunately, there are few behavioral progress monitoring tools that have demonstrated sufficient levels of psychometric adequacy (Chafouleas, Volpe, Gresham, & Cook, 2010). Without psychometrically sound tools for assessing student response to intervention, there will likely be errors in decision making that could result in costly outcomes for schools, such as needless resource expenditures and student failure. Thus, the goal of this study was to advance the science of behavioral progress monitoring tools by assessing the treatment sensitivity of four newly developed teacher-completed behavior ratings in the context of an evidence-based classroom intervention.

School-Based Progress Monitoring of Social, Emotional, and Behavioral Functioning

To date, three primary methods of school-based behavioral progress monitoring have garnered attention in both research and practice: Systematic Direct Observation (SDO; see Briesch, Volpe, & Floyd, 2018), Norm-Referenced Brief Behavior Rating Scales (e.g., Reynolds & Kamphaus, 2004), and, most recently, Direct Behavior Rating (DBR; see Briesch, Chafouleas, & Riley-Tillman, 2016). Although both SDO and brief rating scales have a large body of evidence in support of their reliability and validity in diagnostic decision making (Whitcomb & Merrell, 2013), there are concerns about their feasibility and treatment sensitivity (Briesch & Volpe, 2007; National Center on Intensive Intervention, 2014).

Although some SDO protocols have adequate treatment sensitivity, most require 15 to 20 min per observation from an independent observer and multiple observations are needed to obtain a reliable estimate of the student's behavior (Hintze, 2005; Volpe, McConaughy, & Hintze, 2009). Repeated, extended, observations are simply not feasible for the ongoing progress monitoring required within the MTSS context. Similarly, many Norm-Referenced Brief Behavior Rating Scales still may include up to 30 items (e.g., Behavior Assessment System for Children [BASC]–Progress Monitor; Reynolds & Kamphaus, 2004). Although these brief rating scales are certainly shorter than the full scales from which they were derived (e.g., BASC), the time required for completion may make them less acceptable and feasible for use as a progress monitoring tool (Volpe, Briesch, & Gadow, 2011; Volpe & Gadow, 2010). Furthermore, most Norm-Referenced Brief Rating Scales primarily assess symptoms (e.g., inattention, hyperactivity) that contribute to the problems teachers witness, as opposed to the actual referral concerns (e.g., limited academic engagement or work productivity), which likely diminishes teachers' perceptions of the value of existing rating scales (Owens & Evans, 2018).

In contrast, DBR was designed to integrate the strengths of both SDO and brief behavior rating scales in a way that is defensible, flexible, repeatable, and efficient (Chafouleas, Riley-Tillman, & Christ, 2009; Christ, Riley-Tillman, & Chafouleas, 2009). DBR is characterized by three primary principles (Christ et al., 2009). First, the behavior is rated as it naturally occurs (e.g., in a classroom or on the playground) by the individual who is working with the child in that environment (e.g., a teacher). Second, the behavior being rated must be observable, operationally defined, and related to the teacher's primary concern about the student. Third, DBR is short and provides a means of quantifying the frequency or severity of a given target behavior in a period of time (e.g., class period, day, or week). Thus, DBR is an assessment method that relies on the completion of very brief ratings (one to six items) of specific behaviors (e.g., argues with the teacher; out of seat), directly following an observation period, by an individual who is already present in the context. In addition, the observer can select the DBR (e.g., disruptive behavior DBR, academic engagement DBR) to match the child's most problematic behaviors to increase the efficiency and meaningfulness of the rating. Finally, DBR is one of only two methods that has "Convincing Evidence" for treatment sensitivity, according to the National Center on Intensive Intervention. Thus, in comparison with SDO and brief behavior rating scales, DBR is a progress monitoring tool that maximizes treatment sensitivity, efficiency, and value to the teacher, and minimizes resource utilization (training, independent personnel time, purchased materials).

State of the Science of DBR

Most studies of DBR have focused on Single-Item Scales (DBR-SIS), wherein one behavior of interest is operationally defined and the informant typically rates the percentage of time the behavior was present in a given time period (e.g., one school day or one period of the day). More than 25 published studies with elementary school samples have helped to establish convincing evidence that the data obtained from DBR-SIS demonstrate acceptable reliability when completed by the same rater across time points (Chafouleas et al., 2010), adequate sensitivity to small changes in behavior with and without intervention implementation (Chafouleas, Sanetti, Kilgus, & Maggin, 2012; Fabiano, Pyle, Kelty, & Parham, 2017; Miller, Crovello, & Chafouleas, 2017), and are considered to be feasible and acceptable by teachers (Miller et al., 2017; Sims, Riley-Tillman, & Cohen, 2017).

However, this body of research is not without limitations. One concern is that most studies of DBR-SIS have focused primarily on three domains of behavior: disruptive behavior, respectful behavior, and academic engagement (Fabiano et al., 2017; Miller et al., 2017; Sims et al., 2017).

Although these domains represent some of the most frequently identified behavioral concerns by teachers, there are many other behaviors that can put students at risk for academic, social, and behavioral failure. For example, in a study by Owens and colleagues (2012), behaviors targeted for a DRC intervention covered concerns including aggression/touching others, off-task behaviors, rule violations, and temper tantrums. In addition, reports by school psychologists regarding common referral concerns (e.g., Bramlett, Murphy, Johnson, Wallingsford, & Hall, 2002) and reviews of skills targeted on individualized education plans (e.g., Spiel, Evans, & Langberg, 2014) highlight the need for more nuanced items related to academic engagement (e.g., initiation of work, class participation, task completion) and organizational skills (e.g., prepared for lesson, keeps track of materials), as well as items that differentiate hyperactive/impulsive disruptions (e.g., interrupts, out of seat) and oppositional and aggressive disruptions (e.g., argues) as each may differentially affect student–teacher relations and/or peer relations. To ensure that educators can select a DBR that is well matched to a variety of referral behaviors, it is necessary to test and establish the psychometric properties of a wider array of constructs and DBR domains. We chose to build upon the well-established domains but add items that address the nuances mentioned above, as well as to assess additional domains (i.e., organizational skills).

A second concern regarding DBR-SIS is the number of ratings needed to produce a reliable indicator of student behavior within a progress monitoring context. Multiple studies have found that between seven and 10 DBR-SIS ratings of a student's behavior would be necessary to achieve an adequate level of dependability (Chafouleas, Christ, & Riley-Tillman, 2009). However, recent studies have found that fewer rating occasions are necessary when using Multi-Item DBR (i.e., DBR-MIS) as opposed to DBR-SIS (Volpe & Briesch, 2012) and that fewer assessments were necessary to reach adequate dependability as the number of items increased from three to six items (Daniels, Volpe, Briesch, & Gadow, 2017). These data demonstrate the possible strengths of DBR-MIS for progress monitoring, wherein dependable data are desired in a short period of time (e.g., weekly, biweekly).

Although evidence for adequate dependability of DBR-MIS is emerging, we are aware of only two studies that have examined the treatment sensitivity of DBR-MIS. Volpe and Gadow (2010) demonstrated evidence for the treatment sensitivity of abbreviated teacher ratings of inattention-overactivity, aggression, and peer conflict (three items for each construct) in the context of 6-week double-blind placebo controlled methylphenidate trial (2 weeks per each of the three doses of medication). The abbreviated scales demonstrated adequate internal and temporal reliability, convergent validity, and were sensitive to change in behavior as a function of the lowest dose of medication. Similarly, Daniels et al. (2017) found that a six-item DBR-MIS assessing peer conflict demonstrated acceptable treatment sensitivity across 3 days of baseline and 3 days of pharmacological treatment. Although these studies provide support for the use of DBR-MIS in progress monitoring, additional studies are needed to examine treatment sensitivity over longer periods of time (i.e., greater than 6 weeks), across additional DBR domains, and in the context of behavioral interventions that mirror typical school contexts.

The Intervention Context: DRC

The primary goal of this study was to assess the treatment sensitivity of four newly developed DBR-MIS in the context of a classroom intervention. We chose the DRC as the intervention because it is one of the most widely studied and effective classroom interventions for inattentive and disruptive behavior (Vannest, Davis, Davis, Mason, & Burke, 2010). It is flexible enough to address a wide variety of student behaviors, is viewed as acceptable and feasible by teachers, and there are empirical benchmarks for expected rates of success across 4 months of use (Girio & Owens, 2009; Owens et al., 2012). Because the DRC involves the documentation of daily behavior (i.e., frequency counts, percent correct) that is compared with individualized goals (e.g., five or fewer interruptions), it provides an optimal context for evaluating alignment between daily intervention outcome data and daily DBR-MIS ratings.

Conceptually, the daily data from the DRC represents proximal behaviors targeted by the intervention, whereas DBR-MIS ratings represent broader classroom performance objectives. This is an important distinction for two reasons. First, when evaluating a student's response to intervention, it is recommended that the method for assessing progress be independent from the intervention data (Suhr, 2015). Second, because intervention targets may change as a student masters specific behaviors, the DBR-MIS allows for consistency in the progress monitoring over the course of an intervention that flexibly addresses multiple narrow target behaviors.

Finally, it is important to note that, like with any intervention, teacher implementation of the DRC as recommended (i.e., intervention integrity) is variable (e.g., Fabiano et al., 2010; Owens et al., 2002). Because intervention integrity is associated with student outcomes, it is critically important, both in research and in the context of MTSS, to assess integrity simultaneous to progress monitoring student intervention response as lack of student progress could be a function of continued unmet student need and/or low quality intervention implementation. Thus, we first assessed change in DRC target behaviors and DBR-MIS ratings in the context of intervention integrity.

Current Study

The primary goal of this study was to assess the treatment sensitivity of four newly developed DBR MIS in the context of a DRC intervention. To achieve this goal, we first evaluated the effectiveness of the DRC intervention and the integrity with which it was implemented in this sample. This study advances the literature on DBR by (a) expanding the behaviors evaluated, (b) assessing the treatment sensitivity of DBR-MIS in the context of an evidence-based classroom intervention with consideration of the integrity of implementation, and (c) assessing the treatment sensitivity of DBRs over a longer time period than previously studied. We selected a 2-month window as this aligns well with the typical school timeline for making intervention decisions and is the duration necessary to determine if a child has a high or low likelihood of positive response to a DRC (Owens et al., 2012). This study advances the DRC literature by offering a replication of the monthly benchmarks identified by Owens et al. (2012) and does so using a Tau effect size that corrects for possible baseline trends while attending to implementation integrity.

Method

Participants

Data were collected during the 2016-2017 school year. Participants included six kindergarten through fourth-grade teachers, one of whom was a special education teacher and five of whom were general education teachers. Five general education teachers each referred one student; however, one student (Child E) was referred by both his general and special education teachers. Thus, the sample included a total of five students.

Participants were recruited from two elementary schools in Southeast Ohio. All teacher participants were non-Hispanic Caucasian, between 25 and 50 years old (M = 42). Student participants were between 6 and 10 years old (M = 8) and all were non-Hispanic Caucasian; 67.7% were receiving special education services, one of whom (Child E) received at least 50% of instruction in a special education classroom. Although the students did not undergo comprehensive assessments as part of the study, all were referred for academic and behavioral concerns consistent with the symptoms and impairment associated with attention deficit hyperactivity disorder (ADHD) and all had elevated scores on the screening measure (see Integrated Screening and Intervention System Teacher Report Form [ITRF]).

Measures

ITRF. The ITRF is a 43-item screening instrument (Volpe & Fabiano, 2013), which was used to confirm the severity of child behaviors. The ITRF focuses on specific observable and malleable behaviors (rather than diagnostic symptoms) that inform the development of DRC target behaviors, are viewed as acceptable by teachers (Daniels et al., 2016), and have demonstrated high internal consistency (α = .97), strong 2- to 4-week stability (r = .84) and evidence for convergent validity (r >.81) with scores from a measure of overall problem behavior (Daniels, Volpe, Briesch, & Fabiano, 2014). Each item is rated on a 4-point scale, ranging from 0 (*no concern*) to 3 (*strong concern*). To be eligible for the study, students had to have a total score of 30 or higher as this is predictive (area under the curve = .90) of students demonstrating problematic behavior (Daniels et al., 2016).

DBR–Multi-Item Scales (DBR-MIS). The DBR-MIS used in this study were developed through an iterative, three-stage process. First, an initial pool of items within each scale was developed following a review of (a) extant measures of academic enablers and disruptive behavior and (b) databases of DRC target behaviors from prior intervention studies. Second, a Consumer Advisory Panel, comprised of teachers, parents, school psychologists, and principals reviewed items and provided feedback, rating the degree to which each item was believed to be observable, malleable, and important to change. Third, a Scientific Advisory Panel, comprised of researchers with expertise in the constructs of interest and scale development, reviewed items and provided feedback, rating the degree to which each item assessed the intended construct, was observable, malleable, and important to change. Finally, an exploratory factor analysis (N = 307 students) was individually conducted for each DBR-MIS item pool to identify items most representative of each construct. Results indicated a one-factor solution for each of the four DBR-MIS (Daniels et al., manuscript under review), with acceptable factor loadings for all retained items (ranging from .75 to .92) and acceptable internal consistency (>.92).

For this study, the four DBR-MIS were Academic Engagement DBR-MIS (e.g., *starts tasks promptly; actively participates in class; stays on task*), Organization Skills DBR-MIS (e.g., *prepared for lesson, follows instructions for assignments*), Disruptive Behavior (e.g., *out of seat/area, interrupts teacher*), and Oppositional Behavior (e.g., *loses temper; argues with teacher*). The DBR-MIS contained six items with the exception of the Study Skills DBR-MIS that contained seven items. For the Academic Engagement and Organization Skills scales, the teacher is asked to rate how often the positive behavior is exhibited during the day using a 7-point scale ranging from (0) *never* to (6) *almost always*. For the Disruptive and Oppositional scales, the teacher is asked to rate the degree to which each behavior was a problem (e.g., interfered with the student's functioning or functioning of others) using a 7-point scale with response options ranging from (0) *not a problem* to (6) *a serious problem*.

Intervention integrity. DRC integrity was assessed through two methods. First, teachers were asked to give students feedback when a DRC rule violation occurred (e.g., *Carlos, that's an interruption*) and make a tally on the DRC. Teachers were asked to submit these data (either into a website that produced graphs of student performance or to the consultant). DRC integrity was defined as the number of days in which DRC data were submitted by the teacher divided by the total eligible school days (e.g., excluding teacher and student absences, holidays, snow days). This metric has been used in previous studies (e.g., Owens et al., 2012; Owens, Murphy, Richerson, Girio, & Himawan, 2008).

Second, a project consultant completed weekly 30-min classroom observations for the duration of implementation. Following each observation, the observer (a) completed an integrity checklist (found in Volpe & Fabiano, 2013), indicating adherence (yes/no) to nine DRC implementation behaviors (e.g., *teacher reviewed DRC goals with the child, teacher informed the child of behaviors that violated DRC goals*), and (b) rated the quality of four teacher implementation behaviors (e.g., *used an appropriate tone of voice when provided feedback*) on a 7-point scale ranging from 1 (*not at all appropriate*) to 7 (*very appropriate*). The total percent adherence was calculated for each teacher/child pair and the quality indicators were averaged for each teacher (see Table 1). For the five cases in which the teacher implemented the DRC for the required 8 weeks, an average of 7.2 observations per teacher were completed. In one case (Child B), the general education teacher discontinued the DRC intervention after 5 weeks because the student qualified for special education services and began to spend the majority of the day in a special education classroom. Four observations were completed with this teacher. The case was retained because this type of placement change represents typical school practice.

Procedures

All procedures were approved by the university institutional review board. Information regarding the project procedures, risks, and benefits was provided to teachers in the participating schools via email and at a staff meeting. Interested teachers signed consent forms. To identify students who would likely benefit from a DRC intervention, teachers completed the ITRF for up to five students in their classroom, who were demonstrating behavior that interfered with academic performance. Children were eligible for participation if their total ITRF scores were 30 or higher. If more than one student met this criterion, the teacher was instructed to rank order the eligible students and send a parent-friendly project description and parent consent form to the parent of the top-ranked student. If a parent declined to consent, the teacher selected the next highest ranked student, and the process continued until parent consent was obtained for one student. Parents were encouraged to contact the investigators to ask questions before signing the consent form. After obtaining parent consent, the project consultant (graduate student in clinical psychology supervised by a licensed clinical psychologist) obtained child assent.

The consultant conducted an initial target behavior interview (TBI; available at http://oucirs.org/daily-report-card/ http://oucirs.org/daily-report-card/ website) to learn more about the teacher's classroom management style and identify student target behaviors for intervention. Each student's individualized target behaviors were operationally defined and evaluated for periods when the student was with the participating teacher. The teacher and the consultant selected the two DBR-MIS that best matched the child's DRC targets. Prior to recruitment for this project, six project team members identified the two DBR-MIS that best matched each ITRF item. There was high agreement on most items (all six raters identified the same two items). In cases where the agreement was lower, matches were selected as long as three or more members selected the DBR-MIS as a match for the ITRF item.

Once the DRCs were developed and DBR-MIS selected, teachers were randomly assigned to one of three intervention start dates, resulting in two teachers for each of three start points (Child A and Child B are in Cohort 1, Child C and Child D are in Cohort 2, and Child E is represented in Cohort 3 in two separate classes, once with his general education teacher and once with his special education teacher). For the baseline period, all teachers were instructed to begin daily tracking of the DRC target behaviors and complete the DBR-MIS at the end of each day using the Qualtrics survey platform. The project consultant checked in with each teacher on a weekly basis to encourage implementation and was available via email for additional support.

Once at least five data points were collected and a stable baseline was observed in the target behaviors of students in the first cohort, the teachers in this cohort were instructed to launch the DRC intervention. Once students in the first cohort demonstrated a positive response to the DRC, teachers in the second cohort were instructed to launch the DRC (as long as a stable baseline had been achieved). A positive response to the DRC was defined as the student having met the goals on at least 70% of intervention days for at least two out of three DRC target behaviors. Once the second cohort of students demonstrated a response to the DRC, teachers in the third cohort were instructed to launch the DRC. Teachers were asked to implement the DRC and complete the daily DBR-MIS for 8 weeks.

Data Analysis

Intervention effectiveness. To assess treatment sensitivity, we first had to assess the effectiveness and integrity of the DRC intervention. To assess effectiveness, we employed a

Table 1. Summary of DRC Target Frequency and DBR-MIS Scores During Baseline, Month 1, and Month 2 of Intervention.

Child (class, cohort)	ITRF score	Data during baseline *M* (*SD*)	Data during Month 1 *M* (*SD*)	Tau_{novlap} Month 1 effect size	Data during Month 2 *M* (*SD*)	Tau_{novlap} Month 2 effect size	Observed integrity: Quality rating (0 to 7)	Observed integrity:% adherence (0% to 100%)	Integrity % days DRC data complete
Child A (Cohort 1; G)	58						6.47 (0.57)	97.05	95.30
% attention check Qs correct		21.6 (12.56)	45.56 (28.71)	0.75*	66.0 (27.25)	0.92*			
% returns to class on time		9.0 (12.45)	65.61 (37.71)	0.84*	92.55 (24.44)	0.93*			
Leaves seat		8.8 (4.76)	7.06 (3.19)	0.24	6.5 (2.81)	0.31			
DBR-MIS engagement		1.07 (0.51)	2.98 (1.30)	0.88*	4.31 (0.67)	1.0*			
DBR-MIS org skills[a]		1.20 (0.76)	2.88 (1.21)	0.71*	4.17 (0.65)	1.0*			
Child B (Cohort 1; S)	86						3.5 (0.35)	4.35	93.00
% a.m. routine completed		39.5 (12.56)	69.44 (17.66)	0.88*	80.0 (18.03)	1.00*			
Disobeys		7.83 (1.60)	4.83 (1.65)	0.80*	5.33 (2.09)	0.67			
% work complete in ELA[b]		16.67 (25.82)	31.39 (19.39)	0.30	41.67 (14.43)	0.11			
DBR-MIS engagement		1.72 (0.51)	1.80 (0.37	0.09	1.42 (0.44)	0.42			
DBR-MIS org skills		1.50 (0.44)	1.62 (0.47)	0.24	1.29 (0.25)	0.42			
Child C (Cohort 2; S)	93						6.79 (0.30)	100	86.21
% a.m. routine		76.36 (20.33)	97.5 (7.69)	0.66*	100 (0.0)	0.69*			
Leaves seat		7.71 (2.70)	6.7 (2.66)	0.15	5.13 (1.02)	0.56*			
Interrupts		11.43 (4.40)	8.2 (3.49)	0.49*	6.44 (1.15)	0.80*			
DBR-MIS disruptive		4.95 (1.90)	3.44 (1.68)	0.56*	1.96 (0.21)	0.78*			
DBR-MIS engagement		0.87 (0.51)	1.54 (0.57)	0.63*	2.0 (0.0)	1.0*			
Child D (Cohort 2; G)	48						3.39 (1.04)	36.67	84.21
Disobeys		1.87 (0.92)	1.2 (0.95)	0.36	1.5 (1.16)	0.15			
Interrupts		2.07 (1.44)	1.3 (0.92)	0.14	0.57 (0.79)	0.62*			
Leaves seat		1.87 (0.92)	0.85 (0.67)	0.33	1.14 (0.90)	0.23			
DBR-MIS disruptive[b]		2.18 (0.62)	1.12 (0.80)	0.46	1.61 (0.27)	0.22			
DBR-MIS oppositional		1.70 (0.77)	1.14 (1.32)	0.41	0.64 (0.71)	0.69			
Child E[c] (Cohort 3; G)	78						4.08 (1.17)	50.00	79.41
% a.m. routine		67.05 (23.64)	85 (29.58)	0.62*	81.94 (23.96)	0.45*			
% p.m. routine		75.91 (20.35)	96.67 (12.91)	0.38*	94.44 (16.17)	0.46*			
DBR-MIS org skills		2.10 (0.56)	2.09 (0.73)	0.04	2.36 (0.61)	0.26			
DBR-MIS oppositional[b]		1.65 (0.91)	0.98 (1.11)	0.11	0.78 (1.07)	0.28			
Child E[d] (Cohort 3; S)	74						4.94 (0.65)	65.23	84.85
Interrupts		5.16 (2.95)	0.83 (0.99)	0.91*	0.42 (0.90)	0.86*			
Leaves seat		4.58 (3.53)	1.0 (1.03)	0.75*	0.84 (0.83)	0.70*			
DBR-MIS disruptive[b]		2.79 (0.92)	1.35 (0.72)	0.62*	1.21 (0.78)	0.58*			
DBR-MIS engagement		1.89 (0.36)	1.47 (1.34)	0.20	2.89 (1.11)	0.66*			

Note. Positive effect sizes indicate that the behavior is moving in the desired direction (i.e., positive behaviors improving; negative behaviors declining). DRC = Daily Report Card; DBR-MIS = Direct Behavior Rating–Multi-Item Scales; ITRF = Integrated Screening and Intervention System Teacher Report Form (total scores of 30 or higher indicate that the student is at risk for demonstrating problematic behavior); G = general education class; org = organizational; S = special education class; ELA = English language arts.

[a]The teacher associated with Child A indicated that the child was frequently getting out of his seat, rather than raising his hand when needed help with classwork. The behavior of asking for help inappropriately was identified as an organizational skill. [b]Denotes a variable that violates stable baseline criterion; effect sizes for these target behaviors are Tau-U instead of Taunovlap. [c]Child E with general education teacher. [d]Child E with special education teacher.

*$p < .05$.

multiple baseline design across participants. We evaluated change in DRC target behaviors via examination of average levels of behavior during baseline, Month 1, and Month 2 (see Table 1). We also calculated Tau_{novlap} and Tau-U effect sizes (ES; Parker, Vannest, Davis, & Sauber, 2011) to quantify treatment outcome at the end of each month of intervention (Month 1 and Month 2). Tau_{novlap} represents the number of days during a given month that represent improvement (i.e., nonoverlap) from the baseline phase minus the number of days not improved from the baseline phase (i.e., overlap) divided by the total number of data pairs compared between baseline and follow-up (Parker et al., 2011). Thus, Tau provides information regarding the *consistency* of improvement. When baseline trends are present (i.e., Tau values greater than or equal to .10; Tau-U is calculated to correct for baseline trend (Vannest & Ninci, 2015). The following standards were applied to evaluate the magnitude of Tau-U and Tau_{novlap} ESs: ⩽ .20 = small, .21–.60 = moderate; .61–.80 = large; > .80 = very large (Vannest & Ninci, 2015). Finally, we also used visual analyses to assess changes in level and variability from baseline to intervention, in the context of integrity.

Intervention integrity. Given the potential impact of low or variable integrity on intervention decisions made by MTSS teams, we also assessed integrity of the DRC intervention. We evaluated integrity with regard to percent observed adherence, average observed implementation quality, and percent of days with data tracked by the teacher. These indicators of integrity were examined for trends and in relation to emerging evidence regarding acceptable levels of intervention integrity (Owens, et al., in press).

Treatment sensitivity. To assess the treatment sensitivity, we compared the daily DBR-MIS ratings with the daily data from the DRC. Namely, we calculated Tau-U or Tau_{novlap} ES for the DBR-MIS ratings at baseline, Month 1, and Month 2, using the same procedures as described above. The ESs as quantified by the DRC data and the DBR-MIS were compared (with regard to magnitude categories; small, medium, large) for each case and target behavior.

Results

Intervention Effectiveness and Integrity

Quantitative analyses. Across the six students, there were 16 DRC target behaviors, 15 of which demonstrated a stable baseline. One target behavior demonstrated improvement prior to the intervention initiation (Child B—English language arts [ELA] target). This downward trend was corrected for when calculating the ESs. Table 1 provides descriptive data for all DRC target behaviors at baseline, Month 1, and Month 2 of intervention, as well as ESs representing the magnitude of change at Months 1 and 2. Of the 16 DRC target behaviors, all demonstrated a positive response to the intervention during Month 1, with eight of 16 demonstrating a response that was moderate to large in magnitude (Tau_{novlap} range = .62–.91; see Table 1). Similarly, all target behaviors demonstrated a positive response to intervention during Month 2, with nine of 16 demonstrating a response that was moderate to large in magnitude (Tau_{novlap} range = .62–1.00; see Table 1). These data represent a continuation of the treatment effect.

DRC integrity, defined as the percent of days for which the teacher collected DRC data, ranged from 79.41% to 95.30% (M = 87.14%, SD = 5.93; see Table 1). Despite these relatively high rates, there was wide variability in observed adherence and quality (see Table 1). Thus, teachers were grouped based on observed adherence. There were three teachers who demonstrated higher observed adherence (defined as at or above 65%: Child A, Child C, and Child E special education teacher) and three teachers who demonstrated lower observed adherence (below 65%: Child B, Child D, and Child E general education teacher). The average ES for DRC targets associated with teachers whose adherence was 65% or higher was .72 (SD = 0.21) compared with an average ES of .46 (SD = .30) for targets associated with teachers falling below 65% adherence. This level (65% adherence) was selected because it provided an even split across cases and it is in alignment with emerging evidence of minimum benchmarks of classroom intervention integrity associated with change in child behavior (Owens et al., in press). DBR-MIS sensitivity was interpreted while considering these levels of integrity.

Visual analysis. There was one teacher in each of the three randomized start cohorts with higher integrity and one teacher in each cohort with lower integrity. Thus, when depicting the daily data for the DRC target behavior in alignment with the multiple baseline A-B design, the three cases with higher integrity are depicted in one panel (see Figure 1) and the three cases with lower integrity are depicted in a separate panel (see Figure 2).

Visual analysis of Figure 1 reveals that for Child A and Child C (Child E did not have a comparable positive target behavior), the levels of all positive behaviors (*% attention check questions* and *% returns to class on time* for Child A; *% of morning routine complete* for Child C) were higher during the intervention phase than during baseline. Similarly, the levels of all negative behaviors (*leaves seat* for Child A; *interruptions* and *leaves seat* for both Child C and Child E) were lower during intervention as compared with baseline. With regard to variability, reduced variability during intervention as compared with baseline was observed for one of three targets for Child A (*% returns to class*), all three targets for Child C, and both targets for Child E in the special education classroom. Taken together, these data

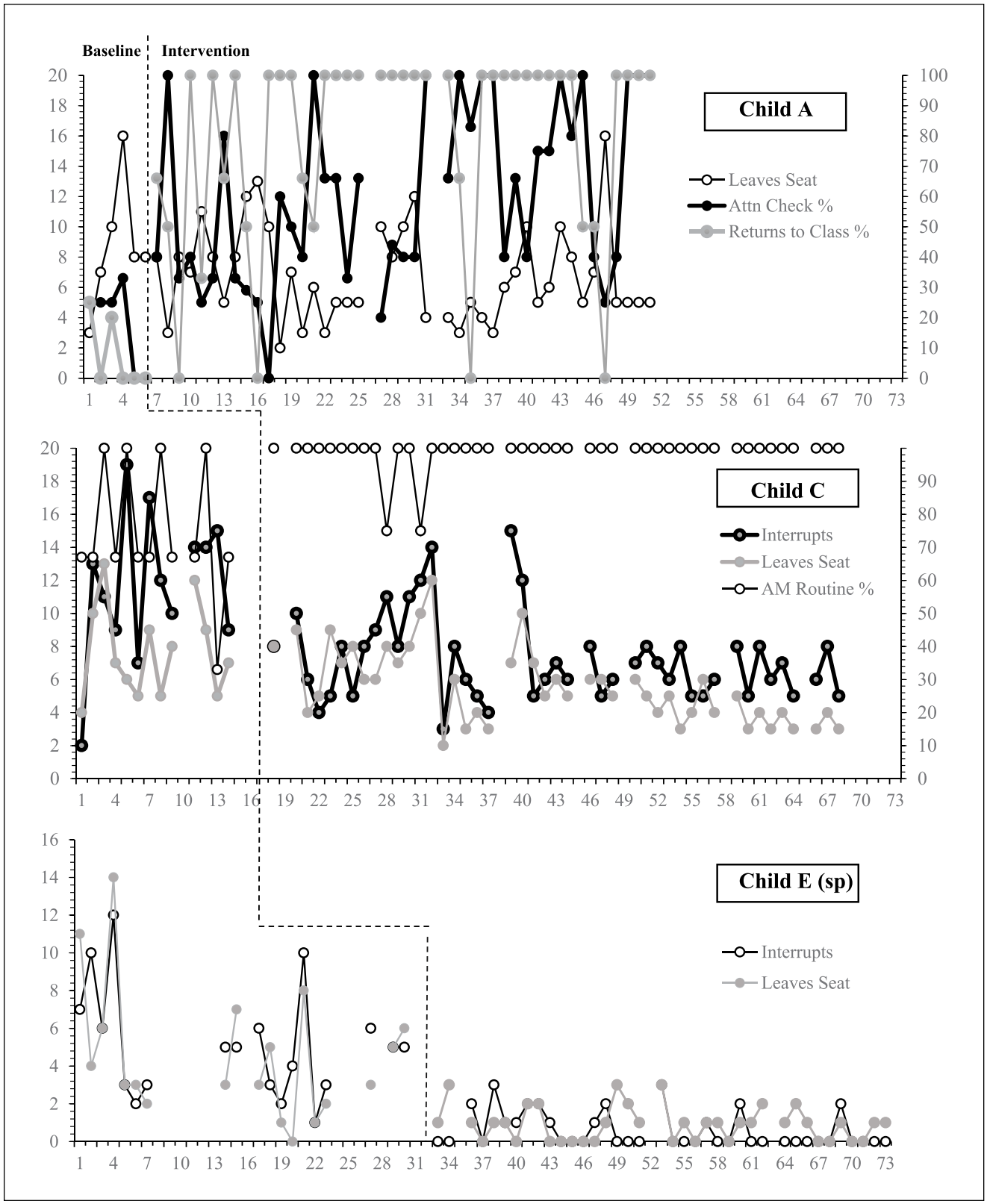

Figure 1. Daily DRC behaviors as a function of A-B design among cases with higher integrity.
Note. AM routine = morning routine; sp = special education class; DRC = Daily Report Card.

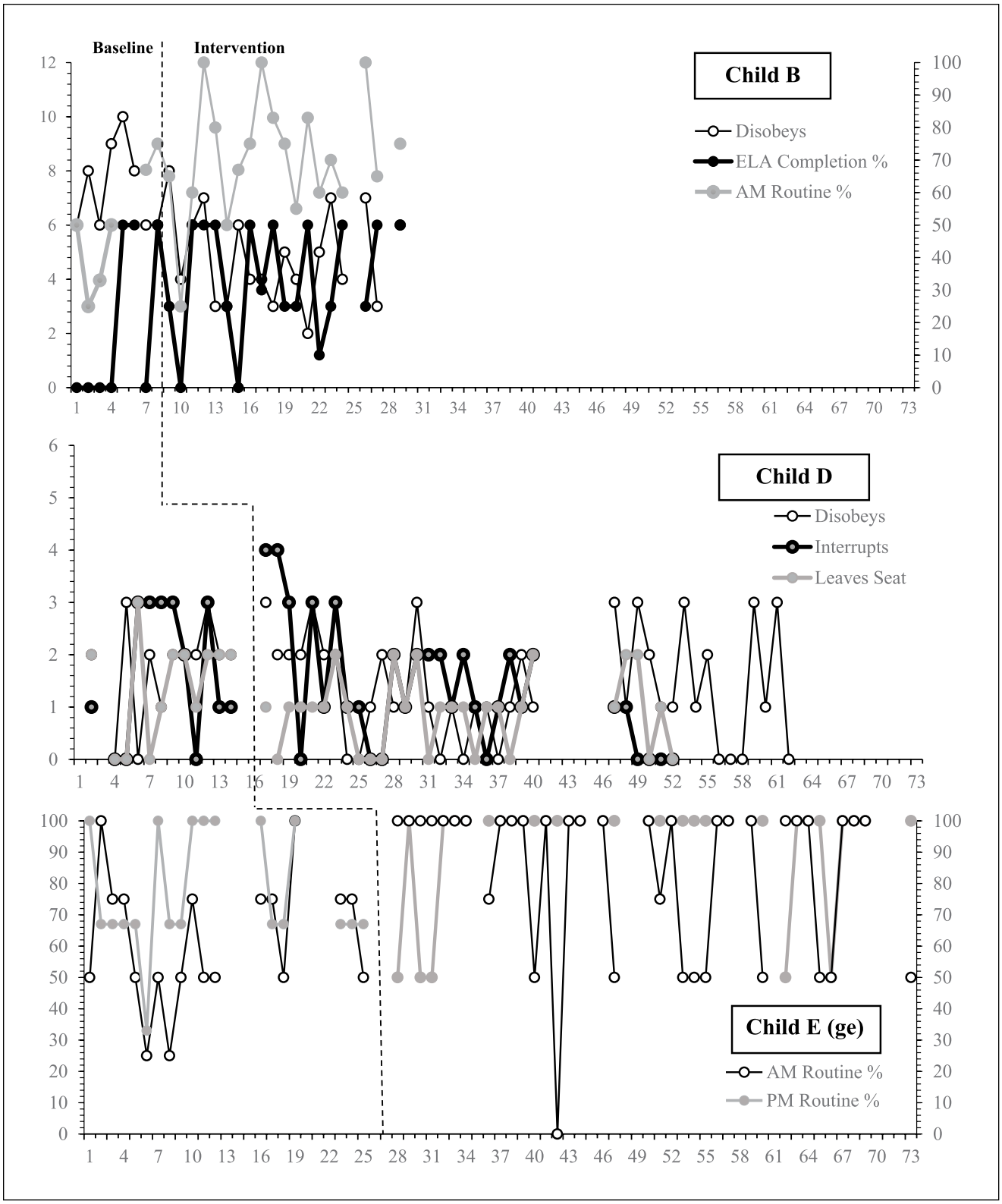

Figure 2. Daily DRC behaviors as a function of A-B design among cases with lower integrity.
Note. ELA = English language arts; AM routine = morning routine; ge = general education class; PM routine = end-of-day routine; DRC = Daily Report Card.

support the effectiveness of the DRC in reducing negative behaviors, increasing positive behaviors, and producing more consistency in both types of behaviors as a function of the DRC interventions.

Visual analyses of the data in Figure 2 reveal a generally similar picture with regard to changes in the level of target behaviors as a function of intervention onset for Child B and Child E (not Child D). However, variability was reduced for one of three behaviors for Child B (*% work complete*), one of two for Child E in the general education classroom, and zero of three for Child D. These muted intervention effects may be related to lower intervention adherence.

Treatment Sensitivity of DBR-MIS

Table 1 provides descriptive data for all DBR-MIS across baseline, Month 1, and Month 2 of intervention. On average, teachers completed DBR-MIS ratings on 79% of days (*SD* = 23.39; range = 50 to 100) during baseline and on 74% of days (*SD* = 32.38; range = 31 to 100) during intervention (see Table 1). Across the six cases, there were 12 DBR-MIS; nine of the 12 demonstrated a stable baseline (baseline trends were accounted for when calculating the ES for the other three).

DBR-MIS disruptive. Three students had DRC targets that aligned with the DBR-MIS Disruptive scale (Child C, Child D, and Child E with special education teacher). In all three cases, the pattern of the ESs for the DBR-MIS mirrored that observed for at least one disruptive target on the DRC. For example, for Child C, the ES for the DBR-MIS (.56 at Month 1 and .78 at Month 2) mirrored the pattern of the ES for the *interruptions* target (.49 at Month 1 and .80 at Month 2). For Child E (with the special education teacher), the ES for the DBR-MIS was .62 and .58 at Months 1 and 2, respectively; and the ES for the *leaves seat* target was .75 and .70 at Months 1 and 2, respectively. For Child D, the changes in the two disruptive DRC target behaviors (i.e., *leaves seat* and *interruptions*) were limited and mixed (both ES below .35 at Month 1; and .23 and .62, respectively at Month 2). In alignment with this, the DBR-MIS disruptive ratings were also modest (.46 at Month 1 and .22 at Month 2).

DBR-MIS academic engagement. Four students had DRC targets that were associated with the DBR-MIS Academic Engagement scale (Child A, Child B, Child C, and Child E with special education teacher). At Month 1, treatment sensitivity was demonstrated for three of four cases (Child A, Child B, and Child C). For example, for Child A and Child C, the ESs for the DBR-MIS (.88 and .63) mirrored the pattern of the ES for the *attention check questions* and *returns to class on time* targets for Child A (.75 and .84) and the *morning routine completion* target for Child C (.66). Similarly, Child B showed limited gains on *work completion* during Month 1 (.30) and the DBR-MIS mirrored this at Month 1 (.09). Treatment sensitivity was demonstrated for all four cases by Month 2. Namely, for Child A, Child C, and Child E, ESs for relevant DRC targets were at or above .69 and ESs for DBR-MIS ratings were all at or above .66. Child B continued to show limited gains in *work completion* and the ES for the DBR-MIS reflected this. Overall, these data provide evidence of the treatment sensitivity of the Academic Engagement scale.

DBR-MIS organizational skills. Three students had DRC targets that aligned with the DBR-MIS Organizational Skills scale (Child A, Child B, and Child E with general education teacher). Although out of seat behavior is typically conceptualized as a disruptive behavior, in Child A, the teacher had indicated that the child was most frequently leaving his seat to ask for help. The teacher and consultant had conceptualized this student as needing to develop the organizational skills of being prepared for lesson, following instructions for assignments, and asking for help appropriately. Thus, this DRC target was associated with the Organization DBR-MIS. Child A had modest changes in the *leaves seat* target (ESs were .24 and .31 at Months 1 and 2, respectively), yet the DBR-MIS ratings showed strong changes (ESs were .71 and 1.0, respectively). This may have occurred as a function of Child A's strong response on his other two DRC target behaviors (i.e., *% attention check questions correct, % returns to class on time*). Improvement in these key domains may have been associated with overall improvements in engagement and organization. For Child B, the ESs suggest large improvements in the percent of the *morning routine complete* (.88 and 1.00 at Months 1 and 2, respectively), albeit with continued variability. The ESs for the DBR-MIS were of lower magnitude (i.e., .24 and .42). However, it is also important to keep in mind that implementation integrity for this DRC was low, which may have affected change in student behavior and tracking of data by the teacher. Finally, Child E (in general education classroom) showed modest improvement in his *a.m. and p.m. routine* DRC targets (ESs ranged from .38 to .62), whereas the DBR-MIS showed small improvements (i.e., ESs were below .30). These data provide restricted evidence for the treatment sensitivity of the DBR-MIS Organizational Skills.

DBR-MIS oppositional. Two students had target behaviors that aligned with the Oppositional DBR-MIS (Child D and Child E with general education teacher). The data in Table 1 reveal limited correspondence between the DRC targets and the DBR-MIS Oppositional outcomes for Child E. Despite significant, moderate improvements in the DRC targets related to *completion of routines* (ESs range = .38 to .62), the Oppositional DBR-MIS demonstrated only small change by Month 2 (ES was .28). Given that this child spent a considerable portion of his day in the special education room, the DRC

targets that were selected for use in the general education classroom were limited (i.e., only targeting *morning and end-of-day routine*) and theoretically less associated with the oppositional DBR-MIS. Per teacher report, this child was motivated by the DRC rewards to improve his routines, yet he continued to show persistent oppositional behavior in other interactions, which likely affected the DBR-MIS ratings. Child D showed minimal change in the level of his behaviors (i.e., disobeys; ESs were .36 and .15 at Months 1 and 2, respectively); the DBR-MIS ratings showed small to moderate improvement (ESs were .41 and .69 at Months 1 and 2, respectively). The DBR-MIS ratings may have improved in association with the student's improvement in *interruptions* (ES at Month 2 was .62). However, the limited evidence of correspondence in these cases may also be associated with low implementation integrity. With only two cases and the variability in implementation integrity, the conclusions that can be drawn about the DBR-MIS Oppositional scale are limited.

Discussion

This study provides important information about the treatment sensitivity of four newly developed DBR-MIS that assess academic engagement, disruptive behavior, organization skills, and oppositional behavior, and does so in the context of an evidence-based classroom intervention applied over a 2-month time span. We found strong evidence for the treatment sensitivity of the DBR-MIS assessing academic engagement and disruptive behavior, whereas results for the DBR-MIS assessing organizational skills were tempered and those for the oppositional scale were inconclusive. This study also replicates previous findings that the DRC is effective in improving students' academic and behavioral functioning (Owens et al., 2012) and highlights the role of integrity and benchmarks when determining a student's response to intervention.

Replication of DRC Effectiveness and Integrity

The magnitude of change observed in DRC target behaviors was similar to previously established monthly benchmarks (Owens et al., 2012), even when using a more conservative ES indicator (Tau-U and Tau_{novlap}). This replication enhances our confidence in recommending benchmarks for intervention decision making. However, the additional information about the relationship between intervention integrity and student outcomes offers important implications for actions to be taken when making data-driven decisions. For example, the average ES for DRC targets associated with teachers whose adherence was 65% or higher was .72 (SD = 0.21) compared with an average ES of .46 (SD = 0.30) for DRC targets associated with teachers falling below 65% adherence. This pattern, and findings from the Owens et al. (2012) study, indicate that, when using a DRC intervention, educators can expect to see change in target behaviors by the end of 1 month that is equivalent to an ES of .50 or higher. However, if the ES is lower, educators should consider (and address) the level of intervention integrity before selecting an alternative or more intensive intervention. Because the DBR-MIS produced similar conclusions (particularly for the Disruptive and Engagement scales), this raises confidence in the use of the DBR-MIS as progress monitoring tools.

It also appears that integrity indicators measured via observation (i.e., adherence and quality) were more variable across cases than integrity indicators derived from the permanent products of the DRC. That is, just because a teacher is regularly completing the DRC does not necessarily mean he or she is implementing the intervention with adequate levels of adherence and quality. For example, despite the fact that the teacher of Child B provided DRC data on more than 90% of days, this teacher repeatedly failed to provide the student with feedback on target behaviors or provide praise for positive behaviors, resulting in very low adherence scores. Consistent with other studies (e.g., Fabiano et al., 2010; Owens et al., 2008), such variable integrity suggests that some teachers need ongoing support to attain high-quality implementation.

Finally, these data suggest that teachers may need to achieve a minimum benchmark (e.g., 65% integrity) to produce a moderate change in student behavior. This is consistent with recent studies revealing minimum benchmarks in teacher classroom management strategies needed to produce change in child behavior (Owens et al., in press) and offers guidelines to consultants who are supporting teachers in making decisions about a student's response to intervention.

Treatment Sensitivity of the DBR-MIS

First, we found strong evidence for the treatment sensitivity of the DBR-MIS Disruptive scale. There were three cases for whom this scale was used (Child C, Child D, and Child E special education). As can be seen from Table 1, when changes in the disruptive behavior targets on the DRC were large (e.g., Child C Month 2, Child E special education both months), change in the DBR-MIS Disruptive scale was moderate to large. When changes in the disruptive behavior targets on the DRC were small or variable (e.g., Child C Month 1, Child D across months), change in the Disruptive scale was also small to modest. These data suggest that the expanded Disruptive scale, which includes multiple disruptive behaviors associated with hyperactivity/impulsivity, is capable of detecting early changes in target behaviors. Furthermore, the magnitude of change in the DBR-MIS is well aligned with the magnitude of change in DRC behaviors.

We also found strong evidence for the treatment sensitivity of the expanded DBR-MIS Academic Engagement scale that includes items assessing on-task, task initiation and completion, and participation. Namely, Child A and Child C both had engagement-related DRC targets that improved significantly (ESs >.65) and concomitantly, the Engagement scale scores also improved significantly (ESs >.63). Child B showed modest changes in percent of work complete with continued variability (perhaps as a function of variable implementation) and the DBR-MIS accurately reflected this (i.e., modest, nonsignificant ESs with variability). Furthermore, for Child E, both DRC targets were more related to disruptive behavior than engagement (leaves seat and interruptions), thus, the Engagement scale changed less in Month 1 (ES = .20) despite improvement in these target behaviors. However, over time with continued improvement in the DRC target behavior, the Engagement scale also showed moderate improvement (ES = .66). This may suggest that changing disruptive behavior may not immediately influence academic engagement, but may affect engagement over time as reductions in disruptive behavior are sustained (see Rapport, Scanlan, & Denney, 1999; Volpe et al., 2006). This also may suggest that teachers need to use two DBR-MISs to adequately capture a more complete picture of the student's overall response to an intervention as each DBR-MIS may highlight areas of improvement and areas for continued intervention. Both of these interpretations warrant further investigation.

In contrast, the evidence for the DBR-MIS Organizational Skills and Oppositional scales are preliminary. Three cases had DRC targets aligned with the Organizational Skills scale, with some evidence of treatment sensitivity across all cases. For Child A, the nature of the target behaviors was overlapping. For example, the ESs for the *returns to class on time* target aligned with the items of the Engagement DBR-MIS scale, however, it also aligned with items on the Organizational Skills scale. Similarly, although the *leaves seat* target is typically considered to align with the Disruptive scale, in this case, the teacher viewed the child's behavior as being related to the Organizational Skills scale (as previously described). It appears the DBR-MIS Organizational Skills scale aligned more with his strong improvement on returning to class on time, rather than his out of seat behavior. This case highlights the importance of further study on how to best match DBR-MISs to target behaviors. Although change in the DBR-MIS Organizational Skills scale score may not directly reflect change in the *leaves seat* target, the DBR-MIS Organizational Skills scale appeared to be sensitive to changes in other related behaviors targeted for intervention and warrants further examination (as this domain has not been studied in DBR-SIS studies)

For Child B, the ESs showed large changes in *percent of morning routine complete*, yet some variability in performance remained. The Organizational Skills DBR-MIS revealed only small to moderate change, with continued variability. The lower implementation integrity may have been related to the child's continued variable performance over time, and thus the mixed results for the DBR-MIS ratings. Namely, as a broader tool for progress monitoring, the results indicate a need for higher integrity and continued intervention in this case. Finally, Child E with the general education teacher showed modest improvement in his *a.m. and p.m. routines* (ESs ranged from .38 to .62), and the DBR-MIS accordingly showed small improvements (ES ranged from .04 to .26).

We only had two cases for which the Oppositional DBR-MIS was used (Child D and Child E with the general education teacher). These two cases also had low DRC integrity data. Consistent with this pattern, the DBR-MIS Oppositional scale data did not show strong ESs. Thus, this pattern offers preliminary evidence of specificity to change (i.e., stability of the measure in the absence of an effect), in that when oppositional behaviors are not changing, the DBR-MIS is not changing. However, additional evaluation of this scale in the context of positive outcomes is warranted. Furthermore, additional study of the sensitivity of this scale relative to the Disruptive DBR-MIS is warranted to examine the nuances between hyperactive/impulsive disruptions and oppositional/defiant disruptions.

Limitations

The results of this study must be interpreted in the context of limitations. First, generalizations must be made cautiously due to the small, heterogeneous nature of the sample in the current study. Although a small sample can never be representative of the population, the sample characteristics represent the majority of elementary teachers (81.2% of primary-grade teachers in U.S. public schools identify as Non-Hispanic and Caucasian; Goldring, Gray, & Bitterman, 2013). Second, under typical school procedures, teachers may select only one DBR-MIS scale for progress monitoring. However, for research purposes (efficiency in obtaining treatment sensitivity data), we had teachers complete two DBR-MIS. It is therefore unknown whether ratings may have been different if teachers were asked to complete only one scale. Third, the teachers in the study implemented the DRC and completed the DBR-MIS with assistance and reminders from a project consultant. The presence of the consultant may have affected both the integrity of the intervention and the likelihood of the DBR-MIS completion on a daily basis. As such, the documented effects may be stronger than those that would be expected in the absence of ongoing implementation support.

Finally, the realities of the actual school setting resulted in lack of control over some aspects of the project (e.g., Child B changed classrooms after 5 weeks of implementation; two teachers intervened with Child E). Similarly, target behaviors were selected collaboratively with the teachers. For this reason, we ended up with the ability to evaluate some of the DBR-MIS scales (i.e., Disruptive and Engagement) with more cases or repetitions, whereas others (i.e., Organizational Skills, Oppositional) will need further evaluation. Finally, interobserver ratings were not obtained for observed adherence and quality indicators.

Conclusion and Implications

In conclusion, results of the current study contribute to the literature in two ways. First, this study replicates that the DRC intervention is effective for changing target behaviors exhibited by students in general education classrooms, and further highlights the importance of integrity. Second, the findings provide support for the use of DBR-MIS methods for monitoring a student's response to a classroom intervention. Although previous research supported the dependability of data obtained from the DBR-MIS Engagement and Disruptive scales (Volpe & Briesch, 2012, 2015), the current results provide strong evidence for the treatment sensitivity of the expanded versions of these scales, as well as promising evidence for the Oppositional and Organizational Skills constructs. This continued strengthening of evidence in support of the DBR-MIS is important when considering feasibility, particularly in comparison with the use of the DBR-SIS. As noted above, recent studies have demonstrated that fewer repetitions of a DBR are needed to obtain a dependable and valid picture of student behavior when there are more items completed at each time point (Daniels et al., 2017; Volpe et al., 2011). Given persistent concerns related to teachers' lack of time and resources, there may be greater buy-in and acceptability for a progress monitoring tool that is both brief and able to be completed less frequently. Overall, this evidence of treatment sensitivity of the newly developed DBR-MIS provides support for this progress monitoring tool.

Declaration of Conflicting Interests

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

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant (R324A150071). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

ORCID iD

Chelsea L. Hustus https://orcid.org/0000-0003-1306-9165

References

Benner, G. J., Kutash, K., Nelson, J. R., & Fisher, M. B. (2013). Closing the achievement gap of youth with emotional and behavioral disorders through multi-tiered systems of support. *Education and Treatment of Children*, *36*, 15–29.

Bramlett, R. K., Murphy, J. J., Johnson, J., Wallingsford, L., & Hall, J. D. (2002). Contemporary practices in school psychology: A national survey of roles and referral problems. *Psychology in the Schools*, *39*, 327–335.

Briesch, A. M., Chafouleas, S. M., & Riley-Tillman, T. C. (2016). *Direct Behavior Rating: Linking assessment, communication, and intervention*. New York, NY: Guilford Press.

Briesch, A. M., & Volpe, R. J. (2007). Important considerations in the selection of progress monitoring measures for classroom behaviors. *School Psychology Forum*, *1*, 59–74.

Briesch, A. M., Volpe, R. J., & Floyd, R. G. (2018). *A practical guide to school-based observation*. New York, NY: Guilford Press.

Chafouleas, S. M., Christ, T. J., & Riley-Tillman, T. C. (2009). Generalizability of scaling gradients on Direct Behavior Ratings. *Educational & Psychological Measurement*, *69*, 157–173.

Chafouleas, S. M., Riley-Tillman, T. C., & Christ, T. J. (2009). Direct Behavior Rating (DBR): An emerging method for assessing social behavior within a tiered intervention system. *Assessment for Effective Intervention*, *34*, 157–173.

Chafouleas, S. M., Sanetti, L. M. H., Kilgus, S. P., & Maggin, D. (2012). Evaluating sensitivity to behavioral change across consultation cases using Direct Behavior Rating–Single-Item Scales (DBR-SIS). *Exceptional Children*, *34*, 575–591.

Chafouleas, S. M., Volpe, R. J., Gresham, F. M., & Cook, C. R. (2010). School-based behavioral assessment within problem-solving models: Current status and future directions. *School Psychology Review*, *39*, 343–349.

Christ, T. J., Riley-Tillman, T. C., & Chafouleas, S. M. (2009). Direct Behavior Ratings (DBRs): Foundation for the development and use of DBR to assess and evaluate of behavior. *Assessment for Effective Intervention*, *34*, 201–213.

Daniels, B., Volpe, R. J., Briesch, A. M., & Fabiano, G. A. (2014). Development of a problem-focused behavioral screener linked to evidence-based intervention. *School Psychology Quarterly*, *29*, 438–451.

Daniels, B., Volpe, R. J., Briesch, A. M., & Gadow, K. D. (2017). Dependability and treatment sensitivity of Multi-Item Direct Behavior Rating Scales for interpersonal peer conflict. *Assessment for Effective Intervention*, *43*, 48–59.

Daniels, B., Volpe, R. J., Fabiano, G. A., & Briesch, A. M. (2017). Classification accuracy and acceptability of the Integrated Screening and Intervention System Teacher Rating Form. *School Psychology Quarterly, 32*, 212–225.

Daniels, B., Volpe, R. J., Briesch, A. M., & Owens, J. S. (under review). Content validation of direct behavior rating multi-item scales using exploratory factor analysis.

Fabiano, G. A., Pyle, K., Kelty, M. B., & Parham, B. R. (2017). Progress monitoring using Direct Behavior Rating Single Item

scales in a multiple-baseline design study of the daily report card intervention. *Assessment for Effective Intervention*, *43*, 21–33.

Fabiano, G. A., Vujnovic, R. K., Pelham, W. E., Waschbusch, D. A., Massetti, G. M., Pariseau, M. E., . . . Greiner, A. R. (2010). Enhancing the effectiveness of special education programming for children with attention deficit hyperactivity disorder using a daily report card. *School Psychology Review, 39*, 219–239.

Girio, E. L., & Owens, J. S. (2009). Teacher acceptability of evidence-based and promising treatments for children with attention-deficit/hyperactivity disorder. *School Mental Health*, *1*, 16–25.

Goldring, R., Gray, L., & Bitterman, A. (2013). *Characteristics of public and private elementary and secondary school teachers in the United States: Results from the 2011–2012 schools and staffing survey* (NCES 2013-314). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch

Gresham, F. M. (2005). Response to intervention: An alternative means of identifying students as emotionally disturbed. *Education and Treatment of Children*, *28*, 328–344.

Hintze, J. M. (2005). Psychometrics of direct observation. *School Psychology Review*, *34*, 507–519.

Miller, F. G., Crovello, N. J., & Chafouleas, S. M. (2017). Progress monitoring the effects of daily report cards across elementary and secondary settings using Direct Behavior Rating: Single Item Scales. *Assessment for Effective Intervention*, *43*, 34–47.

National Center on Intensive Intervention. (2014). *[Behavioral Progress Monitoring Tool Chart, 2008]*. National Center on Intensive Intervention at American Institutes for Research. Retrieved from https://charts.intensiveintervention.org/chart/behavioral-progress-monitoring-tools

Owens, J. S., Evans, S. W., Coles, E. K., Himawan, L. K., Holdaway, A. S., Mixon, C., . . . Egan, T. (in press). Consultation for classroom management and targeted interventions: Examining benchmarks for teacher practices that produce desired change in student Behavior. *Journal of Emotional and Behavioral Disorders*.

Owens, J. S., & Evans, S. W. (2018). Progress monitoring change in children's social, emotional, and behavioral functioning: Advancing the state of the science. *Assessment for Effective Intervention, 43*, 67–70.

Owens, J. S., Holdaway, A. S., Zoromski, A. K., Evans, S. W., Himawan, L. K., Girio-Herrera, E., . . . Murphy, C. E. (2012). Incremental benefits of a daily report card intervention over time for youth with disruptive behavior. *Behavior Therapy, 43*, 848–861.

Owens, J. S., Murphy, C. E., Richerson, L., Girio, E. L., & Himawan, L. K. (2008). Science to practice in underserved communities: The effectiveness of school mental health programming. *Journal of Clinical Child & Adolescent Psychology*, *37*, 434–447.

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, *42*, 284–299.

Rapport, M. D., Scanlan, S. W., & Denney, C. B. (1999). Attention-deficit/hyperactivity disorder and scholastic achievement: A model of dual developmental pathways. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *40*, 1169–1183.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children-2 (BASC-2)*. Circle Pines, MN: American Guidance Service.

Sims, W. A., Riley-Tillman, C., & Cohen, D. R. (2017). Formative assessment using Direct Behavior Ratings: Evaluating intervention effects of daily behavior report cards. *Assessment for Effective Intervention*, *43*, 6–20.

Spiel, C. F., Evans, S. W., & Langberg, J. M. (2014). Evaluating the content of individualized education programs and 504 plans of young adolescents with attention deficit/hyperactivity disorder. *School Psychology Quarterly*, *29*, 452–468.

Suhr, J. A. (2015). *Psychological assessment: A problem-solving approach*. New York, NY: Guilford Press.

Vannest, K. J., Davis, J. L., Davis, C. R., Mason, B. A., & Burke, M. D. (2010). Effective intervention for behavior with a daily behavior report card: A meta-analysis. *School Psychology Review*, *39*, 654–672.

Vannest, K. J., & Ninci, J. (2015). Evaluating intervention effects in single-case research designs. *Journal of Counseling & Development*, *93*, 403–411.

Volpe, R. J., Briesch, A., & Gadow, K. D. (2011). The efficiency of behavior rating scales to assess disruptive classroom behavior: Applying generalizability theory to streamline assessment. *Journal of School Psychology*, *49*, 131–155.

Volpe, R. J., DuPaul, G. J., DiPerna, J. C., Jitendra, A. K., Lutz, J. G., Tresco, K., . . . Junod, R. V. (2006). Attention-deficit/hyperactivity disorder and scholastic achievement: Models of mediation via academic enablers. *School Psychology Review, 35*, 47–61.

Volpe, R. J., & Briesch, A. M. (2012). Generalizability and dependability of Single-Item and Multiple-Item Direct Behavior Rating Scales for engagement and disruptive behavior. *School Psychology Review*, *41*, 246–261.

Volpe, R. J., & Briesch, A. M. (2015). Multi-item direct behavior ratings: Dependability of two levels of assessment specificity. *School Psychology Quarterly, 30*, 431–442.

Volpe, R. J., & Fabiano, G. A. (2013). *Daily behavior report cards: An evidence-based system of assessment and intervention*. New York, NY: Guilford Press.

Volpe, R. J., & Gadow, K. D. (2010). Creating abbreviated rating scales to monitor classroom inattention-overactivity, aggression, and peer conflict: Reliability, validity, and treatment sensitivity. *School Psychology Review*, *39*, 350–363.

Volpe, R. J., McConaughy, S. H., & Hintze, J. M. (2009). Generalizability of classroom behavior problem and on-task scores from the Direct Observation Form. *School Psychology Review*, *38*, 382–401.

Vujnovic, R. K., Holdaway, A. S., Owens, J. S., & Fabiano, G. A. (2014). Response to Intervention for youth with attention-deficit/hyperactivity disorder: Incorporating an evidence-based intervention within a multi-tiered framework. In M. S. Weist, N. A. Lever, C. P. Bradshaw, & J. S. Owens (Eds.), *Handbook of school mental health: Research, training, practice, and policy* (2nd ed., pp. 399–412). New York, NY: Springer.

Whitcomb, S., & Merrell, K. W. (2013). *Behavioral, social, and emotional assessment of children and adolescents*. New York, NY: Routledge.