Article

Hammill Institute on Disabilities

Journal of Positive Behavior Interventions
2020, Vol. 22(2) 116–126

DOI: 10.1177/1098300719876098
jpbi.sagepub.com
SAGE

Data-Based Decision Making for Social Behavior: Setting a Research Agenda

Allison L. Bruhn, PhD[1], Joseph H. Wehby, PhD[2], and Ted S. Hasselbring, EdD[2]

Abstract

With the advent of Precision Teaching more than 50 years ago, researchers and practitioners began to examine how to use K-12 students' academic data to guide instructional decisions. Although the field has advanced with the use of curriculum-based measurement and data-driven decision rules for reading and math, the same is not true in the area of social behavior. In this article, we provide a brief retrospective of academic decision making to inform an initial call for research related to behavioral decision making. Then, we highlight areas for exploration related to baseline data, measurement tools, and features of behavior that, if answered via rigorous scientific inquiry, can move the field forward in making data-based decisions to improve behavioral outcomes.

Keywords

assessment, challenging behavior, single-case designs, data analysis

With the recent passage of the Every Student Succeeds Act (ESSA; 2015), the federal government reaffirmed previous legislative language (e.g., No Child Left Behind [NCLB] Act, 2001) encouraging schools to enact systems-based improvements to school climate that will support all students, including those who are most at risk (McCurdy, Empson, Knoster, Fluke, & Grant, 2019). The logic is that a systems-based approach, which involves multiple elements or features functioning as a whole to achieve a common purpose (Betts, 1992), will improve academic and behavioral outcomes for K-12 students. Building on public health models of prevention and intervention, these systems involve the following core features: effective core instruction, universal screening, a continuum of interventions varying in intensity, progress monitoring, and data-driven decision making. Schools across the country have adopted preventive frameworks designed to provide a data-driven continuum of services for the varied needs of students with and without disabilities (Lloyd, Bruhn, Sutherland, & Bradshaw, 2019). These frameworks generally follow a tiered or leveled approach in which the intensity of supports provided at each tier reflects the intensity of students' needs within that tier. A well-known tiered academic framework is Response-to-Intervention (RTI) and a social-behavioral framework is School-Wide Positive Behavior Interventions and Supports (SWPBIS). Multi-Tiered Systems of Support (MTSS) is a more recent and comprehensive term encompassing academic, social, *and* behavioral supports. Although there has been debate on the best way to describe or name these systems, data-driven decision making focused on identification, implementation, and evaluation issues is a cornerstone of *all* preventive frameworks (Horner, Sugai, & Anderson, 2010). Moreover, data-based decision making (DBDM) is fundamental to effective implementation of a multitiered system of prevention and intervention.

Whereas DBDM related to academic skills has advanced in important ways that have led to improved student outcomes (Filderman, Toste, Didion, Peng, & Clemens, 2018), myriad questions remain about how to use data to monitor student progress in social-behavioral interventions (Maggin & Bruhn, 2018). For instance, some psychometric studies on the reliability, validity, sensitivity, and usability of behavioral progress monitoring tools have been conducted, yet limited information exists about how to use the data to inform intervention decisions (e.g., Briesch, Chafouleas, Riley-Tillman, & Contributors, 2016; Bruhn, Barron, Fernando, & Balint-Langel, 2018; Gage, Prykanowski, & Hirn, 2014; Riley-Tillman, Chafouleas, Sassu, Chanese, & Glazer, 2008). Furthermore, in research studies, generally DBDM is done by the researchers, not the practitioners who often lack knowledge, training, self-efficacy, and implementation experience in this area (Bruhn, Estrapala, Mahatmya,

[1]The University of Iowa, Iowa City, USA
[2]Vanderbilt University, Nashville, TN, USA

Corresponding Author:
Allison L. Bruhn, The University of Iowa, N252 Lindquist Center, Iowa City, IA 52242, USA.
Email: allison-bruhn@uiowa.edu

Action Editor: Grace Gengoux

Rila, & Vogelgesang, in press; Bruhn, Rila, Mahatmya, Estrapala, & Hendrix, 2018; Dunn, Airola, Lo, & Garrison, 2013; Stormont, Reinke, & Herman, 2011). This is not surprising given findings of a recent review of 57 undergraduate and graduate syllabi from special education courses in behavior or classroom management from universities in 31 different states (Majeika, Wilkinson, & Wehby, in preparation). Authors found that although college students received instruction on measuring, evaluating, and graphing data, courses lacked assignments requiring students to collect data, create decision rules, and use data to make decisions about intervention effects and next steps.

Thus, the purpose of this article is to raise an initial call for research related to behavioral DBDM. By first examining the history and current status of academic DBDM, and then highlighting the related yet underdeveloped practices in social behavior, we use this information to suggest potential areas of research for moving the field forward in making effective data-driven decisions to improve social-behavioral outcomes (e.g., increases in academic engagement and positive social interactions, decreases in off-task and disruptive behaviors).

A Brief History of Data-Based Decision Making

To better understand how we might conceptualize and apply DBDM to social behavior, we first examine related work in academic DBDM. This work began in the 1960s with Dr. Ogden Lindsley at the University of Kansas who pioneered "Precision Teaching" (Lindsley, 1990). In Precision Teaching, a performance goal or "aim" was set at a specified target date, then called an "aim star." Student performance data were collected and graphed for 3 days to determine the student's baseline level of performance. An "aim line" was then drawn from the midpoint of the baseline data to the "aim star." This line represented the performance trajectory needed for the student to meet the desired goal. Each day, the teacher would provide instruction, and student performance data were used to determine if the student was on track to meet the expected goal. This "teach and probe" procedure continued until either the student reached the expected aim, or the instructional program was deemed ineffective and the instruction was changed.

Lindsley's general rule of thumb for instructional effectiveness was that if a student was progressing at or above the aim line, then the program was appropriate for that student; otherwise, the instructional program was ineffective and needed to be changed in some way. In Lindsley's words, "The learner knows best." That is, the student's performance should determine the "right" teaching strategy. One limitation with Lindsley's rule was that teachers were left on their own to determine if and when an instructional program was ineffective, as there were no specific data-based decision rules to guide program changes. Without formal decision rules, teachers often were unsure about the successfulness of a program.

Building from the research on Precision Teaching, Haring, Liberty, and White (1980) took a progressive step toward addressing the lack of decision rules. They analyzed data over 2 years on student performance as well as the types of changes teachers made in their instructional programs when no decision rules were in place. When instructional changes were made without decision rules, student performance increased 33% and 41% of the time in Years 1 and 2 of the study, respectively. From these data, they developed formal decision rules to determine what type of programmatic change should occur (e.g., continuing, altering, discontinuing, or replacing instruction with something else). Common strategies for instructional changes included reteaching an earlier skill; providing more information in the form of instructions, cues, or feedback; delivering more powerful reinforcers or consequences; and moving ahead to a more advanced skill. Rules were applied as a series of yes/no questions in the form of a discriminate analysis, which guided teachers to a solution regarding the need for instructional changes. Haring et al. (1980) found that when teachers applied these rules to guide instructional changes, student performance improved 65% of the time, which was a large increase over changes made without decision rules.

Curriculum-Based Measurement (CBM)

Researchers extended early work on DBDM and developed CBM to measure students' academic performance through the ongoing assessment of student performance. Critical components of CBM, which is commonly used in schools today, are described in the following paragraphs.

Defining the general outcome measure. To begin, the teacher must identify the problem that needs to be addressed (e.g., the student reads too slowly or is unable to compute two-digit addition problems). Once the problem is specified, CBM involves probes used to measure the student's response to the instruction. Typically, CBM probes overlap closely with a school's curriculum, are quick to administer, can be given frequently, and are sensitive to short-term student gains (Stecker, Fuchs, & Fuchs, 2005). These measures are designed to provide the teacher with data on the effectiveness of the instructional program. In reading, the most common general outcome measure is proficient oral reading of a sample passage or a maze task, both of which are on a student's instructional grade level. In mathematics, the general outcome measure generally consists of a sample of critical grade-level computation skills (Stecker et al., 2005).

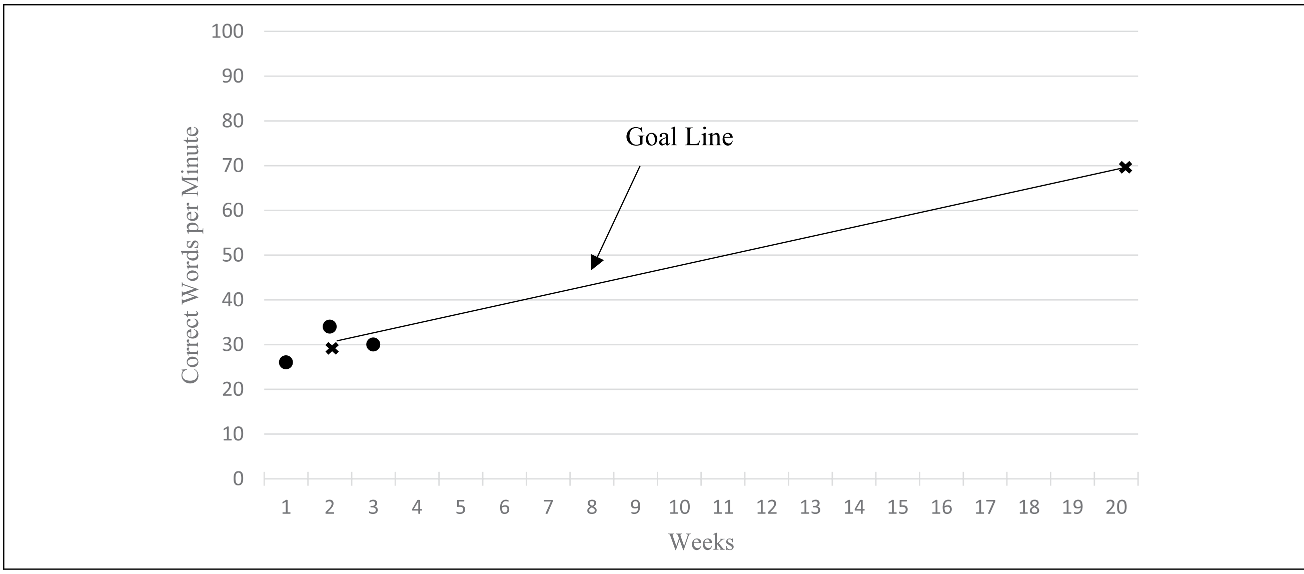

Figure 1. Establishing the goal line.

Setting the instructional goal. Once the problem has been identified and probes have been selected to measure student progress, an instructional goal must be established. This first requires baseline data on the student's present level of performance. Once a minimum of three baseline data points have been gathered, it is possible to establish a target performance goal for the student. By taking the median of the baseline data and multiplying the expected rate of improvement by the number of weeks of instruction, a performance goal (previously known as the aim star) can be established (Filderman & Toste, 2018). For example, if a student's median baseline reading rate is 30 correct words per min and an expected rate of improvement is two words per week and the monitoring will occur over 20 weeks, the goal would be 70 (i.e., $30 + [2 \times 20]$).

CBM requires probe data to be graphed for visual analysis. The skill being probed and the scale (e.g., correct words per min) are labeled on the vertical axis. The horizontal axis represents the time (number of weeks) over which the skill is to be monitored. The graphing process begins by plotting the baseline data points. Next, an outcome goal point is placed on the graph at the intersection of the expected goal and the week of expected goal achievement. A line is then drawn between the median of the baseline data points and the outcome goal. Using the previous example, a line is drawn connecting the median of the three baseline data points (30) and the goal (70) at week 20 on the graph. This line is called the *goal line* (previously known as the aim line in Precision Teaching). The goal line represents the growth rate students are expected to achieve to reach their instructional goal (Filderman & Toste, 2018; see Figure 1).

Frequency of progress monitoring. Unlike Precision Teaching in which the student is probed daily on each skill being taught, generally in CBM, the student is assessed only once or twice per week. Although probing provides the teacher with less data to evaluate the effectiveness of the instructional program, academic skills often do not change quickly (e.g., daily), and so less frequent monitoring does not present problems for evaluating student progress.

Implementing decision rules. Interpreting a student's RTI involves ongoing data collection and analysis—usually by comparing the student's actual performance data with the goal line and employing formal decision rules to determine when a program is effective or needs to be changed. Perhaps the simplest decision rule applied to graphed data involves data points above the goal line. When data are above the goal line, but not yet above the performance goal, no change in the instructional program is needed because it is presumed the student will reach the intended goal by the expected date if the data continue to stay above the goal line (see Figure 2).

When data fall below the goal line, the teacher must decide when it is appropriate to alter the instructional program. A common decision-making rule used in this case is the Points Below Method (Filderman & Toste, 2018). When this rule is applied, the teacher simply compares the student's last three data points with the goal line. If the last three data points are below the goal line, the teacher implements a program change as the student's progress will not meet the goal by the expected date (see Figure 3).

The Points Below Method is not the only one that can be used; more sophisticated rules or methods exist to guide the

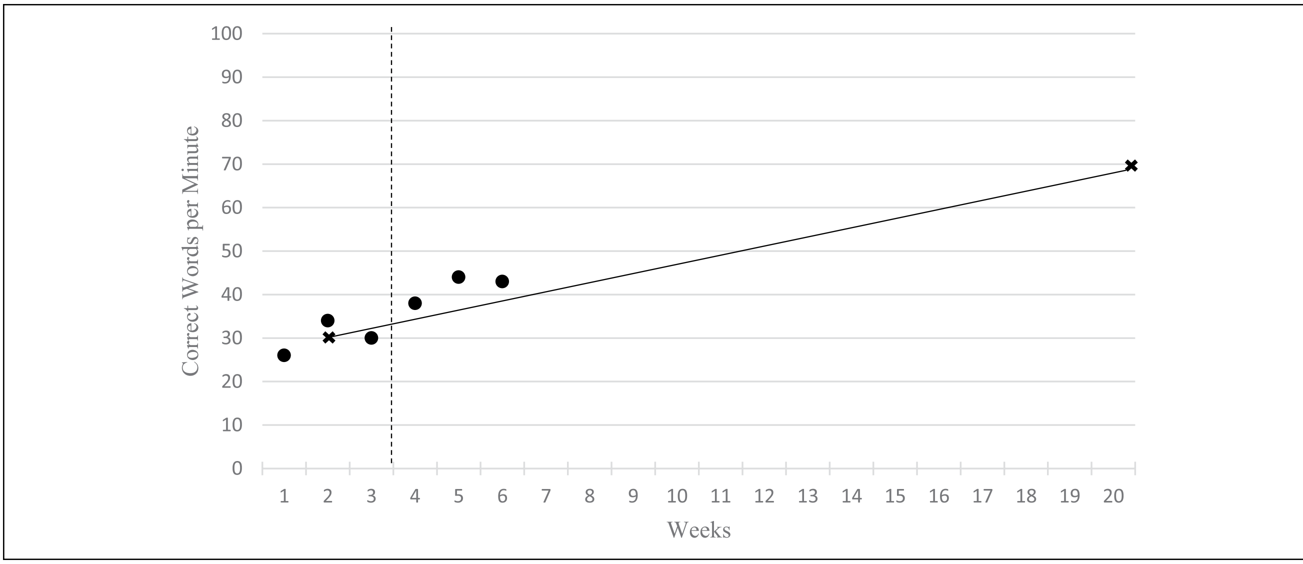

Figure 2. No instructional change needed.

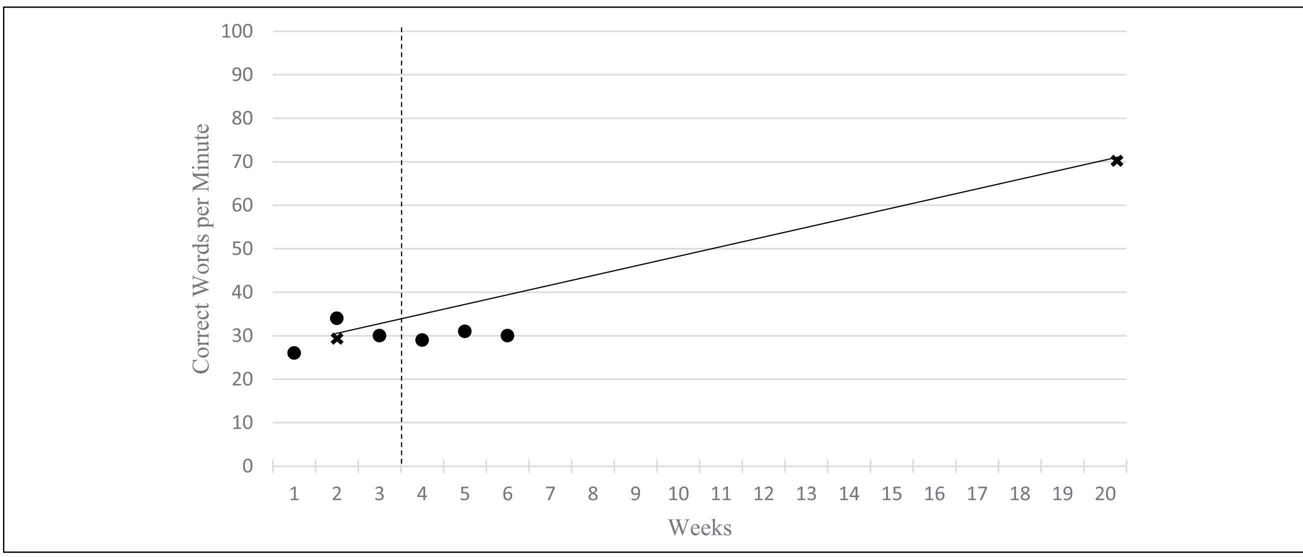

Figure 3. Instructional change needed.

teacher (e.g., slope method, criterion mastery; Filderman et al., 2018). For example, when at least eight data points have been collected and they closely follow the aim line but the last three are below, trend lines can be used to determine if a program change is needed. As shown in Figure 4, the last eight data points are used for making a decision. The data points are divided into as equal groups as possible (e.g., 3, 2, 3). Next, the median is found for the first and third triad and an X is drawn at each median. A *trend line* is then drawn between the two median marks. If this trend line is sloping in an upward or positive direction, the decision would be to continue with the same instructional program. If this trend line is flat or sloping in a negative direction, the decision would be to change the instructional program (Hutton, Dubes, & Muir, 1992).

Various decision rules have been used to help teachers make successful changes in their instructional programs for individual students, with each decision-making method having its own strengths and limitations related to feasibility and accuracy. For more information related to applying these methods, we refer the reader to Filderman and Toste (2018).

In a review of CBM research, Stecker et al. (2005) reported that student progress monitoring alone did not improve student achievement; teachers had to implement program changes when the data indicated a change was

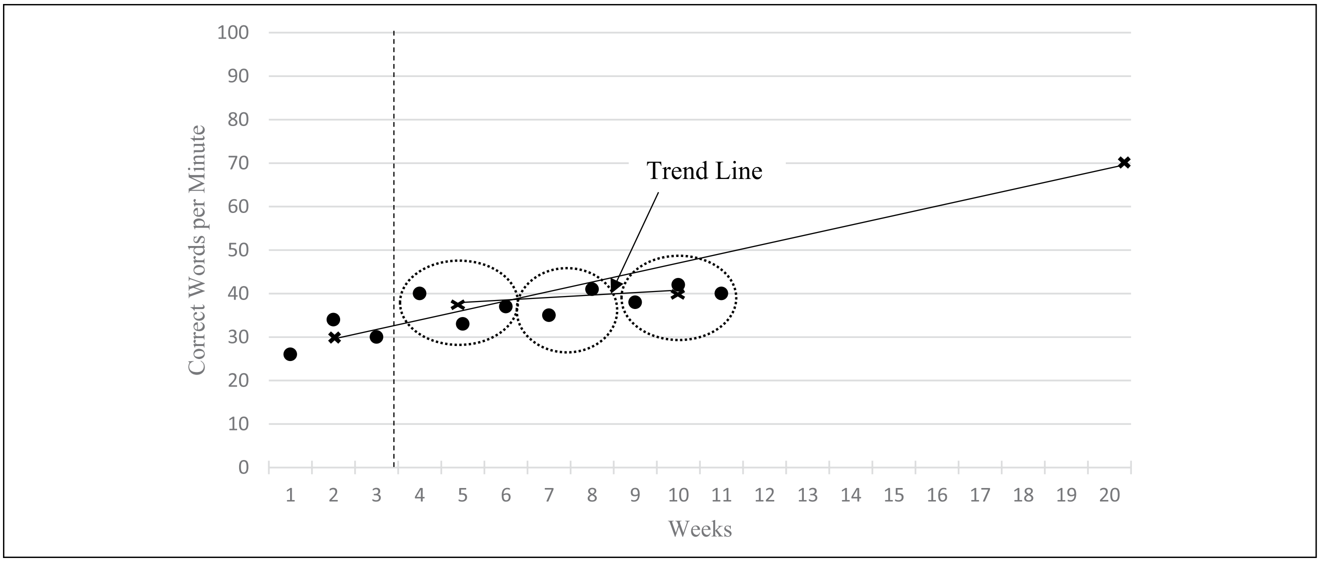

Figure 4. Comparing trend line with goal line.

needed to be successful. Furthermore, when data-based decision rules were used, teachers were more responsive to making the necessary changes in their instructional programs. Relatedly, in a synthesis and meta-analysis on the use of DBDM in reading interventions across 15 studies, Filderman et al. (2018) found overall positive effects for DBDM when compared with studies in which DBDM was not used. These positive effects were most pronounced in studies in which the DBDM process was clear, systematic, and aligned with recommendations from the National Center on Intensive Intervention (NCII; 2013; that is, deliver research-based intervention, monitor progress frequently, administer diagnostic assessments for students not making adequate progress, and adapt instruction as needed based on data). Meta-analysis indicates comparable findings also exist for math and writing (Jung, McMaster, Kunkel, Shin, & Stecker, 2018).

Data-Based Decision Making for Social-Behavioral Interventions

As noted previously, DBDM using a discrete set of rules has been established in the field of reading and math interventions. Similar work has not progressed at the same pace in the social-behavioral intervention field. Although DBDM is fundamental to tiered systems, most work in this area has focused only on Tier 1, or school-wide efforts, which are expected to be effective for 80% to 90% of the school population. In Tier 1, data are collected school-wide; that is, on every student in the building. These data are often required by districts and states to be collected and reported publicly, which may be why the bulk of research on DBDM for behavior is in Tier 1. Typically, school-wide behavioral data include attendance, suspensions, expulsions, and office discipline referrals (ODRs), though a small percentage of schools (<13%) use universal behavior screeners (Bruhn, Woods-Groves, & Huddle, 2014). These data are used to identify students who need Tier 2 or Tier 3 supports because Tier 1 alone is insufficient, with most schools relying on ODRs (Bruhn et al., 2014). Conventional wisdom has been that students who have two to five ODRs should qualify for Tier 2 supports (about 10%–15% of students), and six or more ODRs should qualify them for Tier 3 supports (about 5% of students; McIntosh, Frank, & Spaulding, 2010). Although these data indicate a change in behavioral support is necessary, ODR data do not indicate the type of intervention (e.g., small group social skills instruction, self-management strategies) a student should receive.

Using data to decide who needs Tier 2 or Tier 3 supports in addition to Tier 1 is a fairly straightforward process when using a criterion such as a certain number of ODRs or a cut score on a screening tool. Once students are in a Tier 2 or Tier 3 intervention, however, the DBDM process becomes much more complex. Unlike in Tier 1, a student's RTI ideally is measured on a more frequent, ongoing basis. The frequency of data collection alone makes monitoring progress in Tiers 2 and 3 a more arduous task. At Tier 2, efficiency is especially important because there are likely to be 3 times as many students requiring this level of support as compared with Tier 3 (Hirsch, Bruhn, & Kittleman, 2019). In Tier 2, the behaviors monitored may be more general and tied to the Tier 1 plan (e.g., Be Respectful), whereas in Tier 3, behaviors may be more specific and individualized to the student (e.g., keep hands, feet, and materials to self; Hirsch et al., 2019). The practical implementation of progress monitoring involves extensive coordination among school personnel to determine who, when, what, and how data will be collected (Bruhn, McDaniel, Rila, & Estrapala, 2018). Teams must determine if (a) the classroom teacher or another staff member (e.g., school counselor, paraprofessional) will collect the data, (b) data will be collected daily or weekly, or (c) paper or electronic forms will be used. These are just a few of the myriad issues, and perhaps the most difficult one is determining what type of data should be collected.

Data Collection

The type of data collected and the tool used to collect data depend on the behavior being monitored and the intervention being implemented (Bruhn, McDaniel, et al., 2018). For instance, in systematic direct observation (SDO), discrete behaviors (e.g., swearing) may be observed and recorded using basic frequency counts, whereas continuous behaviors (e.g., academic engagement) may be monitored using time sampling procedures. Although SDO is generally accepted as the "gold standard" in behavioral measurement, it is also time- and labor-intensive (Adamson & Wachsmuth, 2014). Relatedly, some interventions may have data collection tools (or intervention-based measures) built in (Bruhn, McDaniel, et al., 2018), such as the daily progress report in the Tier 2 intervention Check-in/Check-out (CICO; Crone, Hawken, & Horner, 2010). Not only should the tools or methods used to collect progress monitoring data be reliable and valid, but also they need to be feasible, efficient, and repeatable (Christ, Riley-Tillman, & Chafouleas, 2009).

Although ODR data may be readily available and accessible to all school personnel, making those data practical and feasible (Sugai, Sprague, Horner, & Walker, 2000), ODRs should not be used for progress monitoring for a number of reasons. First, the data are only as accurate as the system and practices are reliable, meaning if teachers' understanding of what constitutes an ODR differs across the building, then ODR data may not be a true reflection of student behavior (Pas, Bradshaw, & Mitchell, 2011). Additionally, data are generated only when an ODR is accrued, and ODR data tend to capture mostly externalizing behaviors (McIntosh, Campbell, Carter, & Zumbo, 2009).

Recently, scholars have begun to examine potential behavioral progress monitoring tools that can be tailored to individual behaviors and interventions at Tiers 2 and 3. These include SDO measures such as momentary time sampling of academic engagement, as well as direct behavior ratings (DBRs) of engagement and disruption, both of which have evidence of reliability (Chafouleas et al., 2010; Wood, Hojnoski, Laracy, & Olson, 2016). The National Center on Intensive Intervention (NCII; www.intensiveintervention.org) offers a critical analysis of these various tools, including information on the technical adequacy (e.g., reliability, validity, sensitivity to change) and social validity (e.g., acceptability, feasibility, contextual fit). Each of these methods has advantages and disadvantages, and selecting one often is dependent upon the contextual fit for the classroom or school (Burke et al., 2012).

Data Evaluation

Once a tool has been selected and data collection has commenced, the difficult practice of evaluating data to determine student responsiveness and making decisions is ongoing. Bruhn, McDaniel, et al. (2018) outlined a step-by-step process for monitoring behavioral progress using SDO, DBR, and intervention-based measures. This process includes selecting an appropriate method of measurement, logistical planning, analyzing data, incorporating treatment fidelity into the decision-making process, and adapting intervention based on student response. Although authors make recommendations for visually analyzing stability and trends in behavior, *no research-based rules or criteria are presented for helping practitioners determine if a student is making adequate progress*. For instance, it is unclear if there is a certain percentage a student needs to reach to be considered responsive. Also unclear is the duration a student needs to maintain that percentage to still be considered responsive, and at what point an adaptation to intervention should be made (e.g., fading intervention). In practice, these evaluative decisions generally are left to school site teams. Unfortunately, these teams rarely have any formal training on DBDM for behavior (Bruhn et al., in press). Further exacerbating the difficulty with this process is that no standard protocol exists for social-behavioral DBDM. This is in stark contrast to academic progress monitoring, which was founded on a set of basic decision rules nearly 50 years ago. Although other scholars have developed processes for teams to analyze data and problem-solve (e.g., Team-Initiated Problem Solving [TIPS], Data-Based Individualization [DBI]), the field has neglected to move beyond process to protocol for data evaluation.

Setting a Research Agenda

As evidenced above, there has been some movement on the development of measures to progress monitor social behavior; however, the critical issue of how to use these measures to make decisions about modifying/adapting behavior intervention remains unresolved. Although traditional methods of visual inspection (e.g., stability and trend analysis) are available, these approaches focus on describing data patterns, not informing decisions on programmatic changes to intervention procedures. To make programming changes, a set of decision rules that accounts for baseline levels of performance, the type of measurement tool used, and the type of behavior must be considered.

As we have seen in research on academic DBDM, having a set of rules or a protocol to follow has helped teachers make sound instructional decisions leading to improved student outcomes (Filderman et al., 2018; Jung et al., 2018; Stecker et al., 2005). These decision rules were based on years of research to determine age or grade-level goals, expected growth rates, and the type of tailored instruction necessary to meet students' needs. Comparable decision rules for behavior do not exist, and we hypothesize this is due to the vast complexity of behaviors that vary within and across individuals, as well as variations in the classroom environment (e.g., teachers' competence and attitude, classroom management, instructional activity) that shape behavior (Fabiano, Pyle, Kelty, & Parham, 2017). Despite this complexity, we are hopeful that by laying out these issues and the subsequent questions that need to be answered, this article will serve as a steppingstone to more dedicated work around DBDM for social behavior that follows similar logic to academic DBDM. These issues center on baseline data, measurement tools, and features of social behavior.

Baseline Data

Number of data points. Collecting baseline data allows researchers and practitioners to observe students' present level of performance without an intervention in place. In practice, these data should be used to set intervention goals and to determine if an intervention is working. From a research perspective, these data are imperative to establishing a functional relation between the dependent and independent variables, and thus the opportunity to demonstrate the necessary scientific rigor for establishing evidence-based practices. The What Works Clearinghouse standards for single-case design (Kratochwill et al., 2013) require a minimum of three baseline data points to meet the standards with reservation, whereas five are required to fully meet the standards. One of the first issues that must be addressed in creating decision rules is determining the number of baseline data points required prior to setting a behavioral goal. In academic DBDM, three baseline data points are collected. Applying this same logic to behavior, it is possible three data points are sufficient, so long as behavior is not highly variable. If data are variable, using only three data points to set an accurate and attainable initial behavioral goal may be problematic. Five data points, however, may allow behavior to stabilize. Or, at the very least, five may present a more accurate representation of behavior than three. It is possible, however, that even five data points are not sufficient for establishing a baseline level of performance, and as many as 10 may be needed (Chafouleas et al., 2013; Lewis, Scott, Wehby, & Wills, 2014). Although it may seem trivial to compare three versus five versus 10, the reality is when students have persistent challenging behavior in the classroom, teachers want immediate fixes. Furthermore, a core tenet of Tier 2 and Tier 3 interventions in which these decision rules would be applied is that they are readily available and implemented as soon as a student is identified as needing Tier 2 or Tier 3 support. For students with more severe challenging behaviors, withholding intervention for additional days while collecting baseline data may pose both ethical and social validity problems, particularly for the teachers charged with collecting the data and supporting the target student. Students exhibiting self-injury or physical aggression toward others, for example, are likely to need immediate intervention as their behavior could pose harm to themselves or others. Thus, withholding intervention until a certain amount of stable baseline data had been collected would be unethical. In short, additional research is needed to determine how much baseline data are needed to (a) accurately reflect typical student behavior absent intervention and (b) establish a reasonable initial behavioral goal.

Setting a goal. In academic progress monitoring, an initial performance goal is based on the median of baseline data, the expected growth rate, and the number of weeks of instruction. Unlike academic skills, there is no expected growth rate for behavior nor a standard duration for intervention. Whereas academic skills tend to change more slowly over time, behavioral change can happen rapidly, particularly when students are displaying performance deficits rather than acquisition deficits. With a performance deficit, the student may be refusing to perform appropriate behavior of which she or he is capable whereas an acquisition deficit indicates the student either has not acquired, or is not fluent in, the behavior. If the student has a performance deficit and is engaging in a reversible behavior (i.e., behavior that returns to baseline levels when intervention is withdrawn), the growth rate may be much faster than with an acquisition deficit because an intervention will likely focus on adjusting environmental conditions and reinforcement contingencies that result in immediate improvement (Gresham, Elliott, & Kettler, 2010). In this case, the goal may be set higher and goal attainment may be expected sooner than with a skill deficit. In contrast, acquisition deficits may require explicit teaching, modeling, practicing, and reinforcing of skills (Gresham et al., 2010). Like academic growth, this learning process may lead to slower behavioral change depending on the behavior targeted for change and the dosage of skill instruction. Thus, the goal may be set lower and raised incrementally over a longer period of time.

If we can determine an expected growth rate (e.g., five social interactions per day, 10 percentage points on-task behavior during class), which will likely vary by behavior, individual, and intervention, then we need to determine how long students will need intervention. Because there is no standard duration (e.g., 6 weeks) and practitioners have to decide this on an individual basis, research is needed to determine the optimal duration for intervention. How long does a student need to participate in a Tier 2 or Tier 3 intervention to reach mastery criteria and then maintain behavioral change? Research on growth rates by behavior (e.g., type of deficit), intensity, child characteristics (e.g., disability), and intervention, as well as intervention duration, will help inform researchers and practitioners on how to set attainable goals and then track progress with respect to the goal line.

Measurement Tools and Decision Rules

Assuming researchers can answer questions related to the number of baseline data points, expected growth rates, intervention duration, and attainable goals, decision rules about RTI may be created. However, another issue requiring further examination is the application of decision rules across various measurement tools. As noted above, there has been some progress related to the collection of social behavior data. Measures ranging from SDO, DBR, intervention-based measures, and ODR have been proposed as possible methods for charting an individual's response to an intervention (Bruhn, McDaniel, et al., 2018). These measurement methods produce data that range from counts of a specific topography of a discrete social behavior via SDO (e.g., number of talk-outs during an academic period) to general estimates of a child's adherence to school-wide expectations (e.g., total DBR points across a class period). Unfortunately, other than SDO and to some extent DBR (Briesch et al., 2016), the field of behavior progress monitoring lacks tools that are comparable with CBM in terms of technical adequacy and feasibility. Relatively little attention has been paid to the psychometric issues of behavioral progress monitoring tools, and these tools are critical to accurate decision making in problem-solving frameworks such as SWPBIS further research is needed to establish the reliability, validity, sensitivity, usability, and feasibility of behavioral progress monitoring tools (Maggin & Bruhn, 2018). Applying decision rules to behavioral data can be done only if the data are reliable and valid, as DBDM is predicated on psychometrically sound assessment tools (Chafouleas et al., 2010).

It is important to note that some of this work has begun and can help guide future research on measurement tools for monitoring social behavior. For example, within SDO alone, it has been reported that estimates of duration are more accurate when time sampling intervals are shorter and observation sessions are longer (Lane & Ledford, 2014; Sharp, Mudford, & Elliffe, 2015). In addition, some research has suggested SDO and DBR are significantly and positively correlated (Riley-Tillman, Chafouleas, Briesch, & Eckert, 2008; Smith, Eklund, & Kilgus, 2018) when using a momentary time sampling procedure. It is unknown whether this relationship stands across different recording systems (e.g., duration, event, or time sampling). While this type of work needs to continue to expand the range of behavioral progress monitoring tools as well as the usability and sensitivity of existing tools, concurrently, research needs to occur on the degree to which specific measurement tools influence the types of decision rules that are being used.

A major question in the development of decision rules involves the degree to which different progress monitoring tools (i.e., SDO, DBR, or other) require a set of different decision rules. It could be that a general outcome measure such as DBR would involve a different algorithm than an SDO monitoring approach using momentary time sampling to estimate off-task behavior. As work on improving the technical adequacy of these measures continues, questions on how the features of these monitoring systems impact DBDM also should be explored. It should be noted that the type of behavior being monitored (e.g., prosocial or antisocial) also may influence decision rules and should be included in any research in this area.

To this end, we envision a series of studies that would involve researchers collecting multiple measures (e.g., DBR and SDO) on behavior (both prosocial and antisocial) concurrently, graphing the data, applying a hypothetical set of decision rules, and determining if the same decisions are made at the same time point for each measure or tool. This work would be enhanced by creating standards for using visual analysis to make decisions, especially given research indicating there is often low to moderate agreement when different reviewers analyze the same graphed data (Barton, Meadan-Kaplansky, & Fettig, 2019; Ninci, Vannest, Willson, & Zhang, 2015).

Features of Behavior

A third area of exploration involves understanding how features of behavior interact with decision rules. Like the influence on expected growth rates, a skill deficit or a performance deficit may influence how decision rules are applied. One might presume that with a powerful intervention in place with fidelity, a student with a performance deficit may improve faster because the skill is already in their behavioral repertoire, whereas an acquisition deficit requires the student to acquire the skill and then develop fluency. Thus, it is possible different rules should be created to account for behavior that is expected to quickly change versus that in which slower growth over time is expected. Again, understanding these issues requires thoughtful, rigorous research. Quantitative syntheses of intervention research could begin to answer these questions. As an example, researchers could conduct a systematic search for single-case design studies of a specific intervention (e.g., small group social skills instruction).

Graphed data could be extracted using software that essentially recreates the raw data set (e.g., Plot Digitizer). These data can then be statistically analyzed for growth rates, while accounting for intervention dosage, duration, and measurement variation. Furthermore, within-child characteristics such as behavior type and severity, as well as demographics, could be included as moderators, thus allowing for a more in-depth analysis.

Another feature of behavior that may influence how decision rules function is whether the behavior being measured is positive (e.g., on-task) or negative (e.g., disruption). Although one might presume growth occurs in a positive direction, when negative behaviors such as inappropriate language, disruption, and aggression are measured, "growth" is in a negative direction as evidenced by decreases in behavior. It is not clear if a simple transformation of rules is appropriate. For instance, a decision rule for positive behavior might indicate that, after the student has met the initial goal, the goal should be raised by 10 percentage points. If the behavior is negative, we cannot presume the corresponding rule should be to decrease the goal by 10 percentage points. Research on this issue will involve determining acceptable levels of performance for both increasing and decreasing behaviors. Additionally, for positive behaviors, research is necessary to determine decision rules for adapting interventions as students move from skill acquisition to fluency to maintenance.

Conclusion

Given the complexity of social behavior, developing a set of data-based decision rules for determining students' behavioral response to intervention is a formidable task. This challenge is exacerbated by the dearth of teacher knowledge and self-efficacy in DBDM related, in part, to inadequate preservice preparation (Majeika et al., in preparation). However, research in academic DBDM has demonstrated that when provided clear and discriminate decision rules, teachers can make impactful instructional changes resulting in improved academic outcomes (Filderman et al., 2018; Haring et al., 1980; Jung et al., 2018; Stecker et al., 2005). In light of the success of academic DBDM, we are hopeful about the promise of behavioral DBDM. We have outlined three overarching themes for extending the research in this area specifically related to baseline data, measurement tools, and behavioral features. Understanding how we measure behavior, how often we measure it, how we set goals, and how dimensions of the behavior itself (e.g., topography, type of deficit, definition) alter the functioning of decision rules will be important in this line of DBDM research. With a retrospective look at academic decision making followed by suggestions for research in behavioral DBDM, we recommend researchers heed this initial call and, in turn, conduct DBDM research to improve practice and student outcomes.

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

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Funding for this article was provided by the Institute of Education Sciences (510-14-2540-00000-13607400-6200-000-00000-20-0000).

References

Adamson, R. M., & Wachsmuth, S. T. (2014). A review of direct observation research within the past decade in the field of emotional and behavioral disorders. *Behavioral Disorders*, *39*, 181–189.

Barton, E. E., Meadan-Kaplansky, H., & Fettig, A. (2019). Comparison of visual analysis, non-overlap methods, and effect sizes in the evaluation of parent implemented functional assessment based interventions. *Research in Developmental Disabilities*, *85*, 31–41.

Betts, F. (1992). How systems thinking applies to education. *Educational Leadership*, *50*, 38–41.

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

Bruhn, A. L., Barron, S., Fernando, J. A., & Balint-Langel, K. (2018). Extending the direct behavior rating: An examination of schoolwide behavior ratings and academic engagement. *Journal of Positive Behavior Interventions*, *20*, 31–42. doi:10.1177/1098300717713441

Bruhn, A. L., Estrapala, S., Mahatmya, D., Rila, A., & Vogelgesang, K. (in press). Training teachers on data-based individualization: A mixed methods study. *Behavioral Disorders*.

Bruhn, A. L., McDaniel, S., Rila, A., & Estrapala, S. (2018). A step-by-step guide to tier 2 behavioral progress monitoring. *Beyond Behavior*, *27*, 15–27. doi:10.1177/1074295618756984

Bruhn, A. L., Rila, A., Mahatmya, D., Estrapala, S., & Hendrix, N. (2018). Analyzing the effects of data-based, individualized interventions using multilevel modeling and visual analysis. *Journal of Emotional and Behavioral Disorders*. Advance online publication. doi:10.1177/1063426618806279

Bruhn, A. L., Woods-Groves, S., & Huddle, S. (2014). A preliminary investigation of emotional and behavioral screening practices in K-12 schools. *Education and Treatment of Children*, *37*, 611–634.

Burke, M. D., Davis, J. L., Lee, Y. H., Hagan-Burke, S., Kwok, O., & Sugai, G. (2012). Universal screening for behavioral risk in elementary schools using SWPBS expectations. *Journal of Emotional and Behavioral Disorders*, *20*, 38–54. doi:10.1177/1063426610377328

Chafouleas, S. M., Briesch, A. M., Riley-Tillman, T. C., Christ, T. J., Black, A. C., & Kilgus, S. P. (2010). An investigation of the generalizability and dependability of Direct Behavior Rating Single Item Scales (DBR-SIS) to measure academic engagement and disruptive behavior of middle school students. *Journal of School Psychology*, *48*, 219–246.

Chafouleas, S. M., Kilgus, S. P., Jaffery, R., Riley-Tillman, T. C., Welsh, M., & Christ, T. J. (2013). Direct Behavior Rating as a school-based behavior screener for elementary and middle grades. *Journal of School Psychology*, *51*, 367–385. doi:10.1016/j.jsp.2013.04.002

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

Crone, D. A., Hawken, L. S., & Horner, R. H. (2010). *Responding to problem behavior in schools: The behavior education program* (2nd ed.). New York, NY: Guilford Press.

Dunn, K. E., Airola, D. T., Lo, W. J., & Garrison, M. (2013). Becoming data driven: The influence of teachers' sense of efficacy on concerns related to data-driven decision making. *The Journal of Experimental Education*, *81*, 222–241.

Every Student Succeeds Act of 2015, Pub. L. No. 114-95.

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.

Filderman, M. J., & Toste, J. R. (2018). Decisions, decisions, decisions: Using data to make instructional decisions for struggling readers. *TEACHING Exceptional Children*, *50*, 130–140.

Filderman, M. J., Toste, J. R., Didion, L. A., Peng, P., & Clemens, N. H. (2018). Data-based decision making in reading interventions: a synthesis and meta-analysis of the effects for struggling readers. *The Journal of Special Education*, *52*, 174–187.

Gage, N., Prykanowski, D., & Hirn, R. (2014). Increasing reliability of direct observation measurement approaches in emotional and/or behavioral disorders research using generalizability theory. *Behavioral Disorders*, *39*, 228–244.

Gresham, F. M., Elliott, S. N., & Kettler, R. J. (2010). Base rates of social skills acquisition/performance deficits, strengths, and problem behaviors: An analysis of the Social Skills Improvement System–Rating Scales. *Psychological Assessment*, *22*, 809–815.

Haring, N. G., Liberty, K. A., & White, O. R. (1980). Rules for data-based strategy decisions in instructional programs: Current research and instructional implications. In W. Sailor, B. Wilcox, & L. Brown (Eds.), *Methods of instruction for severely handicapped students* (pp. 159–192). Baltimore, MD: Brookes.

Hirsch, S. E., Bruhn, A. L., & Kittleman, A. (February, 2019). *Technology across the tiers*. A Presentation at the International Association for Positive Behavior Support Conference, Washington, DC.

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

Hutton, J. B., Dubes, R., & Muir, S. (1992). Estimating trend in progress monitoring data: A comparison of simple line-fitting methods. *School Psychology Review*, *21*, 300–312.

Jung, P. G., McMaster, K. L., Kunkel, A. K., Shin, P., & Stecker, P. M. (2018). Effects of data-based individualization for students with intensive learning needs: A meta-analysis. *Learning Disabilities Research & Practice*, *33*, 144–155.

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

Lane, J. D., & Ledford, J. R. (2014). Using interval-based systems to measure behavior in early childhood special education and early intervention. *Topics in Early Childhood Special Education*, *34*, 83–93.

Lewis, T. J., Scott, T. M., Wehby, J. H., & Wills, H. P. (2014). Direct observation of teacher and student behavior in school settings: Trends, issues, and future directions. *Behavioral Disorders*, *39*, 190–200.

Lindsley, O. R. (1990). Precision Teaching: By teachers for children. *Teaching Exceptional Children*, *22*(3), 10–15.

Lloyd, B., Bruhn, A. L., Sutherland, K., & Bradshaw, C. (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. doi:10.1177/0198742918808485

Maggin, D., & Bruhn, A. L. (2018). Evidence-based assessment and single-subject research: A commentary for the special issue on direct behavior ratings. *Assessment for Effective Intervention*, *43*, 71–78. doi:10.1177/1534508417738722

Majeika, C. E., Wilkinson, S., & Wehby, J. H. (in preparation). *How are preservice special education teachers trained to use data to monitor the progress of behavioral interventions?*

McCurdy, B. L., Empson, D. N., Knoster, T., Fluke, S. M., & Grant, C. A. (2019). School resource officers and schoolwide PBIS: Considerations for training. *Preventing School Failure: Alternative Education for Children and Youth*, *63*, 1–8.

McIntosh, K., Campbell, A. L., Carter, D. R., & Zumbo, B. D. (2009). Concurrent validity of office discipline referral cut points used in schoolwide positive behavior support. *Behavioral Disorders*, *34*, 100–113.

McIntosh, K., Frank, J. L., & Spaulding, S. A. (2010). Establishing research-based trajectories of office discipline referrals for individual students. *School Psychology Review*, *39*, 380–394.

National Center on Intensive Intervention. (2013). *Data-based individualization: A framework for intensive intervention*. Washington, DC: Office of Special Education, U.S. Department of Education.

Ninci, J., Vannest, K. J., Willson, V., & Zhang, N. (2015). Interrater agreement between visual analysts of single-case data: A meta-analysis. *Behavior Modification*, *39*, 510–541. doi:10.1177/0145445515581327

No Child Left Behind Act of 2001. Pub. L. No. 107-110, y 115, Stat. 1425. (2002).

Pas, E., Bradshaw, C. P., & Mitchell, M. M. (2011). Examining the validity of office discipline referrals as an indicator of student behavior problems. *Psychology in the Schools*, *48*, 541–555.

Riley-Tillman, T. C., Chafouleas, S. M., Briesch, A. M., & Eckert, T. L. (2008). Daily behavior report cards and systematic direct observation: An investigation of the acceptability, reported training and use, and decision reliability among school psychologists. *Journal of Behavioral Education*, *17*, 313–327. doi:10.1007/s10864-008-9070-5

Riley-Tillman, T. C., Chafouleas, S. M., Sassu, K. A., Chanese, J. A. M., & Glazer, A. D. (2008). Examining the agreement of Direct Behavior Ratings and systematic direct observation for on-task and disruptive behavior. *Journal of Positive Behavior Interventions*, *10*, 136–143. doi:10.1177/1098300707312542

Sharp, R. A., Mudford, O. C., & Elliffe, D. (2015). A data-based method for selecting parameters of momentary time sampling to provide representative data. *European Journal of Behavior Analysis*, *16*, 279–294. doi:10.1080/15021149.2015.1083294

Smith, R. L., Eklund, K., & Kilgus, S. P. (2018). Concurrent validity and sensitivity to change of Direct Behavior Rating Single–Item Scales (DBR-SIS) within an elementary sample. *School Psychology Quarterly*, *33*, 83–93. doi:10.1037/spq0000209

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in Schools*, *42*, 795–819.

Stormont, M., Reinke, W., & Herman, K. (2011). Teachers' knowledge of evidence-based interventions and available school resources for children with emotional and behavioral problems. *Journal of Behavioral Education*, *20*, 138–147.

Sugai, G., Sprague, J. R., Horner, R. H., & Walker, H. M. (2000). Preventing school violence: The use of office discipline referrals to assess and monitor school-wide discipline interventions. *Journal of Emotional and Behavioral Disorders*, *8*, 94–101.

Wood, B. K., Hojnoski, R. L., Laracy, S. D., & Olson, C. L. (2016). Comparison of observational methods and their relation to ratings of engagement in young children. *Topics in Early Childhood Special Education*, *35*, 211–222.