

---

# Measuring Outcomes in Schools

Laura C. Chezan, Thomas R. Kratochwill,  
Mark D. Terjesen, and Kim Van H. Nguyen

---

## Measuring Outcomes in the Schools

In light of their training in psychometrics, evaluation, and statistical analysis, school psychologists may have the most knowledge, of any professional in the school setting, of evaluating the impact of educational and psychotherapeutic interventions (Fagan & Wise, 2000; National Association of School Psychologists; NSAP, 2010). School settings provide a natural opportunity for important, real-world, and meaningful data that will enhance the delivery of interventions and decision-making as to who benefits from what intervention under what conditions. In this chapter, we provide detailed guidelines for strategies for the working school psychologist to measure system-level change, classroom-wide change, and individual change. We first present various methodologies for collecting data to evaluate change at all levels of intervention implementation. Next, we describe in detail straightforward approaches to calculate the

effectiveness of an intervention. Throughout the chapter, we also review the practical barriers that may exist in the Australian educational system that challenge the collection of outcome data, and we present potential strategies to help manage these barriers.

---

## Overview of Data Collection and Measuring Outcomes

School psychologists provide a wide range of services to children and youth, parents, and educators. Services provided by school psychologists include but are not limited to academic, behavioral, and psychological assessment and interventions for children and youth; collaboration and consultation with educators, parents, and other professionals; or development of crisis management procedures (Australian Psychological Society; APS, 2013). Therefore, school psychologists may intervene at multiple levels with different targets for change. For example, they may consult with the school administrative staff to develop a school-wide policy to reduce the frequency of bullying within the school and implement a system-wide approach to promote appropriate behavior. School psychologists may also work with classroom teachers on developing class-wide strategies that target specific academic or behavioral concerns for groups of students or perhaps the whole class. For example, the school psychologist and educator may collaborate

---

L.C. Chezan (✉)  
Old Dominion University, 110 Lions Child  
Study Center, Norfolk, VA 23529, USA  
e-mail: [lchezan@odu.edu](mailto:lchezan@odu.edu)

T.R. Kratochwill  
University of Wisconsin, Madison, WI, USA

M.D. Terjesen • K.V.H. Nguyen  
St. John's University, New York, NY, USA

to improve early literacy skills for students in an entire class and develop teachers' instructional and managerial skills to support this intervention. Thus, it is important that school psychologists would continuously gather data to monitor and evaluate the effectiveness of the intervention. Finally, school psychologists may also work on the individual student level with parents and/or teachers to collect data to understand a student's abilities and challenges and then use these data to develop an intervention plan and to evaluate the outcome of the intervention. An example of this option would be developing an intervention for a student who is disruptive in class. The school psychologist may conduct a functional analysis to understand the purpose (e.g., attention) of the problem behavior and then work with the parent and teacher on strategies to change that behavior while continuing to collect data throughout the intervention.

There is much more to providing services to children and youth than just selection of an intervention and its implementation. Providing services in school settings also involves program evaluation (NASP, 2010). Program evaluation refers to the systematic analysis and interpretation of data to evaluate the effectiveness of a specific intervention. Two aspects of program evaluation merit further discussion. One aspect refers to data collection. Data collection is the process of gathering information to document and measure outcomes to evaluate a student's progress or lack of progress on a specific behavior, thus allowing the school psychologist to monitor a student's acquisition during intervention (Riley-Tillman & Burns, 2009). Data collection has three purposes. First, it establishes a basis for developing individual, classroom, or school-wide goals at the onset of the intervention program. Second, it provides information on a student, class, or school's level of performance on the target behavior before and after the intervention has been implemented. Third, it allows the school psychologist to determine whether an intervention was successful, thus answering questions such as "How well does an intervention work?" or "Does the student learn the skill or behavior being taught?" (NASP, 2010).

## The Importance of Measuring Outcomes and the Role of the School Psychologist

When delivering services in school settings, school psychologists must practice within the scope of their profession and meet the required standards of professional competence. According to the *NASP Model for Comprehensive and Integrated School Psychological Services* (NASP, 2010) domain of data-based decision and accountability, school psychologists should practice within the context of a problem-solving framework, systematically collect data to understand students' abilities and challenges, and use such data on an ongoing basis to evaluate and monitor the effectiveness of school-based interventions as well as the school psychologist's services. Moreover, school psychologists must acquire theoretical knowledge and demonstrate technical skills that enable them to address complex student needs while providing comprehensive and effective services that have direct and measurable impact (Gibbons & Brown, 2014).

When assessing the impact of services provided, school psychologists should follow several evaluation guidelines: (a) the use of multiple measures, including at least one measure of impact on student outcomes; (b) the use of valid and reliable measures of student outcomes; (c) the use of measures that are sensitive to documenting different levels of proficiency; and (d) systems that are linked to professional development and improvement (Waldron & Prus, 2006). In addition, they need to gather information from a multitude of sources including informant ratings from each member of the evaluation team and other individuals who can provide relevant information on a student's current level of functioning on the behavior of interest. School psychologists will also collect direct observation data in the student's natural environment to verify the information gathered through informant ratings and to obtain a better understanding of the student's behavior targeted for intervention as well as other environmental variables that may influence his or her behavior (Gibbons & Brown, 2014). Throughout

the evaluation process, school psychologists must select assessment tools and implement evaluation practices that have supporting empirical evidence documenting their effectiveness at the individual, group, or systems levels (NASP, 2010).

All Australian schools have a responsibility, under the Disability Discrimination Act (1992) and Disability Standards for Education (2005), to ensure that students with disabilities are given equal rights to access and participate in education and to make *reasonable* adjustments to allow that participation. However, this is in contrast to the United States, for example, where there is specific legislation that makes school districts legally accountable for not only the *success* of the whole school but especially children who are at risk, which is encapsulated in the No Child Left Behind Act (NCLB, 2001). Aligned with the reauthorization of the Individuals with Disabilities Education Act (IDEA, 2004), there is an emphasis in schools on implementing preventive practices and linking assessment to intervention to promote academic and behavioral success for all students. A preventive intervention designed to address students' academic needs is Response to Intervention (RTI). RTI is a multi-tiered model that is used to evaluate psychoeducational services within core instruction (primary prevention; Tier 1), supplemental instruction (secondary prevention; Tier 2), and intensive instruction (tertiary prevention; Tier 3). RTI is noticeably different than the "wait-to-fail" model often seen in the traditional special education system (Walker & Shinn, 2010).

RTI is frequently discussed as a multi-tiered system of supports (MTSS) for matching instruction and intervention to student needs (Batsche et al., 2005). This conceptualization shifts more focus to using data to inform prevention and early intervention activities, rather than simply identifying students who may be eligible for special education (Castillo, 2014). The application of data-based decision criteria to school-wide screening and regular monitoring of at-risk students is needed as part of the RTI process to help ensure that those in need are provided the appropriate services to increase the likelihood of their

academic success (Glover & DiPerna, 2007). School psychologists' knowledge and skills in research and program evaluation make them ideal candidates to both advocate for and support efforts to engage in program evaluations (Castillo, 2014).

The expertise that school psychologists have in assessment, data collection and interpretation, and evidence-based practices is a critical factor in ensuring that all aspects of measuring student performance and evaluating outcomes are executed appropriately by educators and other professionals involved in a student's education. The main purpose of measuring student outcomes is to promote student success. Thus, implementing screening measures and continuously evaluating progress and outcome data are critical aspects in the early identification, prevention, and intervention to address academic and behavioral problems displayed by students in school settings.

---

### Practical Considerations for Evaluating Outcomes Assessment

Measuring outcomes, evaluating the effectiveness of interventions, and making data-based programmatic decisions to address students' academic and social behavior in school settings are high priority in many countries worldwide including Australia. Specifically, the importance of gathering and analyzing data within the school is reflected in the Australian Education Act 2013: "Support will be provided to schools to find ways to improve continuously by: (a) analysing and applying data on the educational outcomes of school students (including outcomes relating to the academic performance, attendance, behaviour and well-being of school students); and (b) making schools more accountable to the community in relation to their performance and the performance of their school students" (Part 1, Section 3). One way of doing this is through the introduction of the National Assessment Program—Literacy and Numeracy (NAPLAN), which was implemented in Australia in 2008(<http://www.nap.edu.au>). It is an annual national assessment for all students in years 3, 5, 7, and 9.

NAPLAN's data have been acclaimed to providing the basis for standardized national monitoring of student achievement in Australia. All students in these year levels are expected to participate in tests in literacy and numeracy that are developed over time through the school curriculum. KidsMatter is an initiative of mental health promotion, prevention, and early intervention initiative set in Australian primary schools and in early childhood education and care services (Slee et al., 2009). KidsMatter is supported by a strong evidence base and is a comprehensive model for improving mental health in schools that involves the entire school community. By incorporating components of the Collaborative for Academic, Social, and Emotional Learning program (CASEL, [www.casel.org](http://www.casel.org)), it targets the mental health and well-being of all students in primary schools through promoting a positive school environment and providing education on social and emotional skills for life.

In this section of the chapter, we first discuss practical challenges encountered by school psychologists when designing and implementing evaluations within the school setting. Next, we present several guidelines to be considered by school psychologists when selecting assessment tools to evaluate student outcomes and the effectiveness of specific interventions. We end by describing how school psychologists should select effective interventions to address academic and social behaviors and by discussing several aspects related to intervention implementation in school settings.

### Timing of Outcomes Assessment

Although school settings and the role of school psychologists involve a significant amount of data collection to assist in educational placement and recommendations, the school culture, from our experience, appears more reticent to allow for gathering of data to evaluate the effectiveness of interventions on the individual, classroom, or school-wide level. We have had educators question *why* we need data and whether we can just “see if things improve.” This issue may be one of

the challenges that the school psychologist will encounter—to establish a culture of expectation for data collection to evaluate the effectiveness of interventions. This concern may require them to communicate the importance of data collection and also communicate that data are not being collected to evaluate *their* effectiveness in the classroom or as a school administrator. Developing a respectful and collaborative relationship with school educators, support staff, and administrators can increase the likelihood of “buy in” when it comes to gathering data to evaluate the effectiveness of an intervention. One way that we may work toward overcoming that school cultural barrier is helping educators and administrators recognize the benefits for not just the student but for the educator and for the school as a whole. By gathering data and having the school-based professionals be part of this process, it may help educators believe that they are valued not just as sources of information about the students but also as experts on the student and what may work within the classroom or school context.

When a collaborative relationship with school staff has been established, it is important to develop an assessment strategy to clarify goals and measure intervention outcomes. This process may vary depending upon the method or source of assessment as well as the characteristics of the reporters. Cappella, Massetti, and Yampolsky (2009) identify three strategies that would be beneficial to consider with regard to outcomes assessment: (a) timing of assessment, (b) method of source of assessment, and (c) level of outcomes assessed. We briefly discuss these strategies as they relate to practical issues in gathering data within the context of developing and measuring outcomes within the school. It is important to consider *when* assessment is going to occur and *how frequently*. Depending upon the identified goals, assessment could be evaluated via a checklist or brief rating scale or perhaps a more intensive observational approach is warranted. Further, our experience in schools has led us to conclude that it is important that we try to balance the desire for multiple data assessments with an understanding that an educator or parent who is asked to complete a rating scale too

frequently may resist doing so. That is, it may be better to choose less frequent ratings of student academic or student behavior if it is going to increase the likelihood of teachers continuing to be a source of measurement. It is important for the school psychologist to discuss the practicality of data collection with the educator and if they believe that it would be an acceptable approach that would provide meaningful data to evaluate the success of the intervention. This strategy may further engage educators in the process and reinforce that they are an active part of the decision-making and as such will be more likely to continue to be a part of data collection.

With regard to the method or source of the assessment, it really depends upon what the specific objectives of the intervention are as well as the source of the data. For example, if a school-wide program is being developed to increase school attendance, data could be gathered from the administrative offices. If a new reading curriculum is introduced in the classroom, data could be gathered from standardized or end-of-semester reading examinations. If the intervention is geared toward promoting the social skills development of an individual student, data could be gathered from behavioral observations, student report, and/or teacher or parent ratings. Ultimately, the decision for what method of assessment will be utilized will need to tie both the objectives of the evaluation along with the practical nature of data collection together.

Finally, it is important to consider the desired level of outcomes assessed and the goals of the intervention. This consideration would be important for establishment at the beginning as to what level of desired change or outcome is expected and at what point and to discuss the likelihood of the intervention leading to the desired outcome within a specific time frame. The school psychologist, educators, and school administrators may wish to discuss if they are looking at change at the individual level toward more adaptive functioning in the individual or change at the class or school level with specific goals set (e.g., 70% of the students will be reading at or above grade level). Data collection prior to implementation of the intervention is key at this stage, as it is impor-

tant to ascertain if the goals are realistic and, even if met, would they have a noticeable improvement in classroom behavior. Watson and Watson (2009) offer an example of this consideration: if a student was referred for intervention because he/she was out of her seat 63% of the time, but comparison data shows that her peers are out of their seats 57% of the time, the focus of the intervention may shift from the individual to a more large-scale intervention. Similarly, it is important to consider current and desired level of performance of student academic functioning and behavior at the classroom and school-wide level in comparison with local norms and standardized comparisons and what the desired and realistic effect of the intervention would be. As such, it is important for collection of baseline data for the referred and comparison students at the individual, classroom, or school-wide level and continuous monitoring throughout the intervention. As the number of assessments increases, so does the reliability and validity about the conclusion drawn from the data (Watson & Watson, 2009).

### Choosing Outcomes Assessments

In addition to considering some of the practical issues described above, there are a number of important characteristics that warrant consideration in choosing measures to evaluate the effectiveness of an intervention. Rating scales completed by the educator, the parent, and the student continue to be used both within traditional models of assessment and for evaluating the effectiveness of these scales in response to the implementation of an intervention.

These behavior rating scales are standardized measures consisting of a list of behavioral descriptions. The rater is asked to indicate the degree to which the behavior is present. These rating scales may represent a wide range of student behavioral constructs with standard scores on each construct for comparison purposes. The information collected through rating scales may assist school psychologists in educational and diagnostic decision-making (Merrell, 2008; Shapiro & Heick, 2004) as well as in monitoring



the effectiveness of an intervention. Behavior rating scales have improved in their psychometric characteristics and, therefore, may provide a significant amount of data about student behavior in a fairly efficient manner. Space prohibits a detailed discussion of the best practices in use of the selection of third-party behavior rating scale approaches, and the reader is referred to Campbell and Hammond (2014) and McConaughy and Ritter (2008) for a more complete review.

Nevertheless, we think that it is important to highlight some specific aspects that may warrant consideration for the school psychologist when selecting rating scales for use in evaluation of an individual, classroom, or school-wide intervention. First, rating scales are considered to be an indirect method of data collection that are highly inferential in nature. That is, the information provided by a rater on a rating scale may be limited by their exposure to the student performing the behavior and possibly their personal beliefs about the behavior and the student. As such, these rating scales may be somewhat subjective in nature and more easily influenced by rater variables and motivation. Second, rating scales are varied in terms of their focus and their breadth. That is, some scales are more broad based and include items that cover a range of emotional, behavioral, or social constructs, while others are more narrowly focused on specific areas. The choice of which rating scale to use will be based on the purpose of the evaluation. For implementation of a school-wide mental health program, it might be advantageous to have more of a broadband measure that assesses a number of clinical constructs. Alternatively, for a targeted intervention, such as a program to reduce anxiety among students, it may be important to have more of a narrow band measure that focuses on that specific aspect of state and provides more in-depth detail.

A more direct and less inferential approach to data collection, but admittedly one that requires more effort, is that of behavioral observation. Conducting classroom observations can be a particularly effective way to determine whether interventions implemented at the individual, classroom, and school-wide levels have been effective. The challenges here may come back to

deciding *what* should be measured and *when* it should be observed. When considering using observations to evaluate the effectiveness of the intervention, it is important to make sure that a relevant sampling of behavior has been conducted. This issue may be particularly important at the systems-wide level, as observations of specific groups of students may not be reflective of the impact of the system-wide intervention on the student body as a whole. As such, we recommend observations at the individual level or the classroom level rather than the system level.

When conducting observations as part of an evaluative process to measure the effectiveness of an intervention, it is important for the school psychologist to recognize that there are practical challenges associated with this approach. To begin, observations may warrant more time and effort on behalf of the school psychologist and/or educational staff to develop a reliable, valid, and easy to record methodology of an objectively defined student behavior. Staff resistance to observation recording in the classroom would be an important variable to consider and address. Second, it is important to make sure that what is being observed is related to the goal of the intervention and is meaningful in nature. As an example, while collecting data on out-of-seat behavior among the students might be easily defined and gathered, it does not provide data as to the social context of the classroom. That is, perhaps there are environmental variables (e.g., peer behavior, academic content, or teacher classroom management) that may be impacting upon this behavior. Rather, it may be more important to gather data about the interaction between the student, their peers, and the teacher regarding the behavior in a way that is manageable for data collection purposes to guide intervention selection. For example, Hamre and Pianta (2007) proposed a model to guide classroom observations, the *Classroom Assessment Scoring System* (CLASS), that may help guide school psychologists in choosing what aspects of the classroom interaction relate to student behavior that they may wish to focus upon. Finally, when conducting classroom observations, it may be important to decide whether to focus on the frequency or the quality of the

observed behaviors (Hamre, Pianta, & Chomat-Mooney, 2009). For an in-depth review of behavioral observations, the reader is referred to Volpe, DiPerna, Hintze, and Shapiro (2005).

---

## Intervention Implementation

When selecting an intervention that is to be implemented in the school, there are a number of variables to consider. First, and probably foremost, is that the school psychologist considers the evidence for a specific intervention as it relates to the goals and objectives for the student, classroom, and school. Historically, one of the challenges for this concern is due to the fact that much of the early school psychology research was not developed and conducted by school psychologists in the school setting. As such it is important for the school psychologist to consider the ability to translate the evidence-based interventions (EBIs) that may not have been developed in schools into school-based practice. More specifically, they may wish to consider to what degree would this intervention work within their school and how would it work toward meeting the goals of improving the student, classroom, and school. A number of resources for EBIS in the school exist (see [ebi.missouri.edu](http://ebi.missouri.edu); <http://effectivechild-therapy.com>; [http://faculty.uca.edu/ronkb/bramlett/empirical\\_interventions.htm](http://faculty.uca.edu/ronkb/bramlett/empirical_interventions.htm); <http://ies.ed.gov/ncee/wwc>; <http://www.interventioncentral.org>). School psychologists may wish to consider whether to use one of many optional interventions included on these sites and modify an existing intervention that has been in place in the school setting currently, as well as the practicality of implementing this intervention. When selecting an intervention, the school psychologist should also consider how acceptable the intervention would be to those who most likely are in the position to implement it: educators, paraprofessionals, and school administrators. Consideration of the degree of training involved prior to implementing the intervention as well as the fact that EBIs not be equally effective for all cultural groups are both additional variables for the school psychologist to consider the population with which the intervention will be delivered.

## Intervention Integrity and Outcomes Assessment

One of the most important and challenging variables that relates to evaluating the effectiveness of an intervention is to determine if it was implemented as planned. This concept is often called treatment integrity (Berryhill & Prinz, 2003; Durlak & Dupre, 2008; Martens & McIntyre, 2009). As the interventions adopted require school staff to perform certain behaviors at specific times with individual students, small groups or classes, and entire schools, it is important that the nature and timing of educator behavior is examined as it relates to the impact on student behavior (Martens & McIntyre, 2009). Failure to do so limits the degree of confidence as to whether it was the intervention that led to the desired student outcome or some other variable. Cordray and Pion (2006) posit that treatment effects should only be described relative to the manner in which the treatment was delivered and received. Further, treatment integrity has been closely linked to treatment outcome (Sanetti & Kratochwill, 2014; Sheridan, Swanger-Gagne, Welch, Kwon, & Garbacz, 2009; Vollmer, Roane, Ringdahl, & Marcus, 1999; Wilder, Atwell, & Wine, 2006). Higher levels of treatment integrity are associated with better client outcomes and suggest that subsequent exposure to lower levels of treatment integrity can compromise intervention effects (Vollmer et al., 1999). In addition, assessing treatment integrity can help identify those aspects of a plan that were difficult to implement, focus efforts to revise the plan, and ultimately lead to greater acceptance and use of the plan over time (Erchul & Martens, 2010).

A special series of *School Psychology Review* was devoted to the science of treatment integrity and efforts to promote greater attention to the development of it in school-based practice (Hagermoser-Sanetti & Kratochwill, 2009). Concerns about the degree to which practitioners and research assess integrity are noted with specific recognition of the importance of developing assessment methodologies that are not as intervention specific. For a more complete in-depth review of treatment integrity conceptual and measurement strategies, the reader is referred to

Sanetti and Kratochwill (2014). A brief review of varied strategies for measuring integrity is discussed below.

To begin, regardless of what integrity assessment methodology is used, it is important that a predetermined criterion for what constitutes treatment integrity be determined. That is, school psychologists want to have a numerical value established to draw conclusions as to whether the intervention was delivered in the manner it was intended to be. As an example, with the collection of interobserver agreement data in single-case research studies, Kratochwill et al. (2013) recommend that data should be collected at each phase in the study and for at least 20% of each phase. Neely, Davis, Davis, and Rispoli (2015) describe minimal percentage agreement for observations should range between 80 and 90%. The opportunity to have multiple observational raters for practice in each stage of the implementation may be a challenge for schools with deleted resources, but a level of integrity should be established a priori.

Intervention integrity may be assessed via reports by the implementer, the participant, or by experts in the intervention. Ratings by the implementer or participants in the intervention may be accomplished through their completion of a self-report survey or checklists in which they indicate to what degree they perceived that specific aspects of the intervention were delivered and how effectively they were administered. A percentage of steps completed is typically recorded and used as a measure of integrity. These data may then be reviewed with the implementer, and possible modifications to the procedures may be offered to promote greater fidelity to the treatment. These assessments are fairly easy to do, but there is certainly a risk for bias in completion of these checklists. However, Sanetti and Kratochwill (2009) found a high level of agreement between self-report and more objective recordings of fidelity.

Another approach that may be more labor intensive involves direct observations of a partial or entire delivery of the intervention by the implementer. These observations may be live or via recordings of the intervention, and generally

the observer is an expert in the intervention and has received training in delivery implementation and recording. Typically, they will view the session with a record of the expected steps of the intervention that are to be delivered and will record whether it was delivered as specified in the intervention protocol. Finally, a psychologist may use a permanent product recording to evaluate treatment integrity. A permanent product is an indirect measure of behavior and consists of evaluating the product of a specific behavior after the fact. For example, a psychologist may record an intervention session implemented by an educator and then watch the video at the end of the day to score how many steps of the intervention the educator implemented correctly. Neely et al. (2015) provide a number of examples of permanent products that could involve home-school notes, charts, or tokens.

---

### **Measuring Outcomes Assessment on the System and Classroom Level**

Examining outcomes on the school-wide and classroom level allows the school psychologist and school officials to make data-based decisions on their school's classroom curriculum and interventions (Gibbons & Brown, 2014; Howell, Hosp, & Kurns, 2008). The school psychologist is expected to analyze data on both academic and social behaviors and use progress-monitoring data, to assist in making and evaluating instructional and behavioral recommendations (Howell et al., 2008). In this section, we provide a brief overview on the importance of measuring outcomes, as well as the different methods for evaluating interventions on the systems and classroom level.

It is crucial for educators and school administrative personnel to regularly engage with each other and discuss expectations for their students. For instance, there should be knowledge of state standards and determination of curriculum that lead to the goals for students. There needs to be consistent communication on students' progress, such as through reviews of academic grades and instructional support team (IST) meetings.



Formative and summative evaluations should occur to determine if goals have been met. That is, evaluation should occur during instruction to allow an opportunity for modifications and, after instruction, such as at the end of the academic year, to assess whether goals have been met. Gibbons and Brown (2014) outline a number of best practices in evaluating psychoeducational services using outcome data that include (a) developing clearly defined goals, (b) identifying performance indicators, (c) determining criteria for success, (d) describing the relationship between psychoeducational services and goals, (e) focusing on collaboration and teamwork, and (f) evaluating progress toward goals. Goals are typically district based and should be concise and measurable. Once goals are clearly defined, the next step would be to determine how student performance as related to those goals would be measured. The performance measure should be reliable, valid, and sensitive, meaning it should capture student achievement or behavior and his or her improvement over time (Deno, 1986). For example, schools may use curriculum-based measures (CBMs) to monitor students' basic skill areas or the amount of office referrals for behavioral indicators. Once performance indicators are identified, a criterion for success should be set that describes the desired outcome. Schools may administer state proficiency exams every term to assess their students' reading and math level. In the United States, schools commonly use CBM assessment materials like AIMSweb (<http://www.aimsweb.com>) or the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), which are a set of short procedures and measures for assessing the acquisition of early literacy skills from kindergarten through sixth grade (University of Oregon, 2009; <https://dibels.uoregon.edu>). School psychologists and other educational professionals use DIBELS to identify students who are in need of early intervention (Goffreda & DiPerna, 2010). These screening assessments allow for comparative data (e.g., national, state, or district norms) and benchmark goals (e.g., assessing three times per year) and help determine whether a student is below their target proficiency. They are also useful in comparing and predicting student achieve-

ment and growth over time via research-based normative and relative growth information. The NAPLAN tests in Australia support schools to undertake continuous curriculum-wide progress monitoring in the areas of reading, writing, language conventions (spelling, grammar, and punctuation), and numeracy. NAPLAN testing gives schools the ability to map student progress, identify strengths and weaknesses in teaching programs, and set goals. More information on the reliability, validity, generalizability, and efficiency of universal academic measures can be found at <http://www.rti4success.org/resources/tools-charts/screening-tools-chart>.

Outcome data on a school-wide level involves collecting and organizing existing data, such as statewide assessment scores and office discipline referrals. For example, if fewer than 60% of students perform at the proficient level each of the last five years and rates of office referrals are above the district average, it should lead to data analysis and development of systematic change (Castillo, 2014). Universal screening procedures allow collection of data within educational settings ranging from the individual to district level. They are designed to (a) be administered to all students; (b) identify students who are at risk of future academic, behavioral, or emotional difficulties, thereby be considered for prevention services or more intensive interventions; (c) provide data regarding the degree to which school-based academic instruction, behavioral assistance, and social-emotional programs are meeting the needs of students at the classroom, grade, school, and district levels; and (d) provide information to school psychologists and other educators about individual students' and systems' academic, behavioral, and social-emotional needs (Kettler, Glover, Albers, & Feeney-Kettler, 2014).

Universal screening data should directly link to intervention services. The school psychologist should be aware of measurement error when considering the universal screening data. Errors can often be overlooked in universal screening since the stakes tend to be lower compared to more stringent classification procedures on the individual level (Albers & Kettler, 2014). There are broadband, narrowband, and multiple-gate

approaches to universal screening. A broadband approach assesses several domains concurrently, and a narrowband approach assesses a specific domain, such as early literacy skills or disruptive behavior (Albers & Kettler, 2014).

Multi-gate approaches are conducted on fewer students and provide better accuracy in identifying at-risk students. Albers and Kettler (2014) offer an example of a multi-gate approach toward screening wherein a teacher completes an initial screen of the whole classroom at Gate 1, which consists of ranking the students according to frequency of disruptive behavior. In Gate 2, the teacher would complete a standardized behavior rating scale, for the five students who were ranked in Gate 1 with high frequency of disruptive behaviors. Gate 3 would then consist of the school psychologist performing a standardized observation of students who scored within the at risk or clinically significant range on the behavior rating scales. Parental ratings could also be supplemented for additional information. Although the initial process of the multi-gate approach can produce false positives (identified by the screening system but not truly warranting an intervention), it is in the later stages/gates that lead to more accurate identification.

The use of student performance data to evaluate core instructional programs, to arrive at data-based decisions about certain groups of students and at-risk students, and to deliver intervention to such students effectively and efficiently falls under the framework of RTI (VanDerHeyden & Harvey, 2013). Within the RTI model, there may exist a class-wide problem wherein too many students experience difficulty for a Tier 2 intervention to be effective, suggesting that the problem lies within the classroom rather than with an individual student (VanDerHeyden & Burns, 2005). Thus, it is important to consider class-wide EBIs to provide effective instruction to the students. If there is a difference between the student's performance and the performance standard, the school psychologist should determine whether there are other students with similar differences. If there are other students with similar difficulties, it is possible that a group intervention will be the most efficient approach as long as the problem analysis determines similar needs (Upah, 2008).

Monitoring student progress is critical to ensuring that students receive effective interventions and educational services. Collecting and analyzing progress-monitoring data are part of the problem-solving process. The data are critical to developing and evaluating the quality and effectiveness of interventions. Progress-monitoring data are most often in the form of a curriculum-based measure (Shinn, 2007).

---

## School-Wide Academic Screening and Intervention

Curriculum-based evaluation (CBE) is a systematic approach to problem solving that can help school psychologists make data-based decisions about intervention planning that focuses on improving student outcomes (Howell & Hosp, 2014). Curriculum-based assessment (CBA) includes the knowledge and use of a variety of measurement and assessment tools (Howell & Hosp, 2014). Curriculum-based measures are a type of CBA and refer to a standardized set of procedures used to measure student performance in the areas of reading, math, and written expression (Howell, Hosp, & Howell, 2007). CBM usually includes a set of standardized and short duration tests (i.e., 1–5 min) used to evaluate the effects of instructional interventions in the basic skills of reading, mathematics, spelling, and written expression (Shinn, 1998). The use of CBM is suggested as performance indicators to assess student progress toward long-term goals (Shinn, 2010). With CBM, alternate forms of short tests are developed that sample performance toward the long-term goal or general outcome (Fuchs & Deno, 1991). Another feature of CBM is frequent monitoring and graphical depiction of student scores for decision-making, such as once or twice weekly assessments plotted on a time-series, equal-interval graph (Stecker, Fuchs, & Fuchs, 2005). Thus, CBM is used as a predictive approach to estimate whether students are on target toward meeting long-term goals and to determine whether current instruction is contributing to student growth. This assessment will help school psychologists and teachers modify instructional plans to meet individual student's needs.

## School-Wide Behavioral Screening and Intervention

Schools are encouraged (e.g., US Department of Education, 2012; IDEA, 2004) to adopt evidence-based positive strategies for dealing with problem behaviors rather than discipline tactics that include punishment procedures (e.g., suspensions) (McKevitt & Braaksma Fynaardt, 2014). Implementation of whole-school mental health promotion, such as Australia's KidsMatter framework for supporting social and emotional well-being, has shown to improve children's ability to learn (Dix, Slee, Lawson, & Keeves, 2011). School-wide positive behavior interventions and supports (SWPBIS), also called positive behavior support and positive behavior interventions and supports (PBIS), are a broad set of research-validated strategies based on a problem-solving model that aims to prevent inappropriate behavior through teaching and reinforcing appropriate behaviors (OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports, 2010; Sugai & Horner, 2006). Such strategies align with the core principles of the RTI model and are grounded in multi-tiered systems of support. Common behavioral expectations are identified, taught, and applied to all students. When students demonstrate the desired behaviors, they are to be acknowledged by verbal praise and tangible rewards (e.g., tickets to be redeemed for prizes, preferred activities). Proper implementation of SWPBIS strategies at the whole-school level has been shown to significantly reduce office discipline referrals (Bradshaw, Mitchell, & Leaf, 2010). SWPBIS supports the domain of Preventive and Responsive Services under the NASP Model for Comprehensive and Integrated School Psychological Services (NASP, 2010).

Data should be collected on SWPBIS' effects on student outcomes. Instruments such as the School-Wide Evaluation Tool (SET; Sugai, Lewis-Palmer, Todd, & Horner, 2001; <http://www.pbis.org>) and the Team Implementation Checklist (TIC; Sugai, Horner, & Lewis-Palmer, 2001; <http://www.pbis.org>) measure the implementation integrity of the universal level of a behavioral support model. These tools can pro-

vide scores that display whether a goal is obtained (e.g., at least 80 % on expectations rewarded or at least 80 % of steps achieved). To also evaluate the impact of a model such as SWPBIS, data of behavior incidents, such as discipline referrals, should also be collected. By regularly reviewing data, school personnel can analyze the frequencies and locations of certain problem behaviors and develop action plans to rectify any identified behavioral issues. The overall data on the school level can then be used to identify those students who need more support beyond the universal instruction (McKevitt & Braaksma Fynaardt, 2014). Targeted support should be implemented on a group or individual level to specifically address problem behaviors. Such interventions can range from social skills or problem-solving skills groups to individualized behavior intervention plans.

A classroom positive behavior support system can be established that is congruent to its school's SWPBIS (Tier 1) system. Teachers may incorporate a reinforcement system for their students. The SWPBIS data can help determine whether it is a system's issue or an issue with a select number of teachers. Teachers may also consult with the school psychologist if there is additional classroom support needed. For instance, if a classroom has a number of students with disruptive behaviors, the teacher and school psychologist may collaborate with each other to identify and analyze the problem. Baseline data should be collected, patterns of behavior should be analyzed, and a hypothesis of the function of the behavior should be developed. Then, an intervention plan should be developed that is linked to the hypothesis. The next step would be to implement the intervention with data recording to monitor effectiveness. During intervention, office discipline referrals can provide supplemental information. There are various instruments that assess a specific range of behaviors and serve as a tool to monitor progress and evaluate the intervention plan. The Behavioral Observation of Students in Schools (BOSS; Shapiro, 2003) is a useful observation code for assessing child academic behavior (i.e., on-task or off-task behavior). Other evidence-based measuring tools of social, emotional, and

behavioral functioning include the Behavior Intervention Monitoring Assessment System (BIMAS; McDougal, Bardos, & Meier, 2010), Behavior Assessment System for Children, Second Edition, Progress Monitor (BASC-2 Progress Monitor; Reynolds & Kamphaus, 2009), Direct Behavior Rating—Single Item Scales (DBR-SIS; Chafouleas, Riley-Tillman, & Christ, 2009), and Social Skills Improvement System (SSIS) Rating Scales (Gresham & Elliott, 2008).

---

## Measuring Outcomes on the Individual Level

### Overview of Measuring Individual Outcomes

Measuring individual outcomes is one of the main responsibilities of school psychologists providing services to address students' academic or social behavior in school settings. For example, a school psychologist may, in consultation with a teacher, design a behavioral intervention to increase the on-task behavior during instructional time for a student diagnosed with attention deficit hyperactivity disorder (ADHD) while implementing an ongoing data collection to measure the change in student's behavior before and after intervention. A school psychologist may also implement a cognitive behavioral intervention to teach a youth to manage his or her own behavior through cognitive self-regulation while collecting data to make programmatic decisions and evaluate the effectiveness of the intervention.

Numerous data collection systems exist that allow a school psychologist to measure individual outcomes to determine a student's progress or lack of progress. In this section of the chapter, we discuss the steps involved in designing interventions using some elements of single-case design (SCD) methodology that allow school psychologists to collect objective data and to assist with program evaluation and data-based decision making. Next, we present several guidelines for evaluating the effectiveness of

such interventions at the individual level in school settings.

---

## Designing Interventions and Measuring Individual Outcomes Using Single-Case Designs

*Step 1: Defining the target behavior.* A prerequisite for selecting a specific data collection system from a wide range of available options is the operational definition of the target behavior to be measured. An operational definition is a clear, accurate, and complete description of the target behavior in observable and measurable terms (Kazdin, 2011). For example, a school psychologist may define the behavior of signing "Please" as tapping one's chest with an open palm within 2–3 s of being presented with a preferred item or activity when working with a student who has limited or no speech. An academic behavior such as identifying sight words may be defined as reading words orally within 2 s of when presented with a flash card. The clear identification and operational definition of the target behavior is extremely significant because it guides the selection of the appropriate data collection system depending on the characteristics of the behavior and the purpose of the intervention (Brown, Steege, & Bickford, 2014).

*Step 2: Writing an intervention goal.* As we mentioned previously, one of the purposes of data collection is to inform the process of writing individual goals prior to intervention. Specifically, when writing individual goals, a school psychologist or a teacher must use data to not only justify the need for a specific intervention but also to formulate goals that are realistic and ambitious at the same time and build on a student's strengths with the ultimate goal of increasing his or her academic success or personal independence and self-sufficiency. For example, if data reveal that John completes a one-digit addition correctly within 30 min, an intervention goal for John may be to increase his fluency with this skill, so that he completes a one-digit addition within 2 or 3 min, an amount of time comparable with the

time needed by his peers to complete the same task. In this case, selecting an acceptable performance criterion for John is based on data collected on the amount of time necessary to complete the task.

Individual goals usually describe in measurable terms what the student needs to accomplish by the end of the intervention or the outcome of the intervention and consist of four components: the student, the target behavior, the conditions or instructional circumstances under which the student has to display the behavior, and the criterion for mastery (Mager, 1997; Wolery, Bailey, & Sugai, 1988). An example of an academic individual goal is “Tim will read the number orally when presented with ten flash cards displaying one-digit numerals and the instruction ‘Read the number,’ within 3 seconds for each number with 85 % accuracy for five consecutive intervention sessions.” In this case, Tim is the student, reading numbers orally is the target behavior, when presented with ten flash cards displaying one-digit numerals and the instruction “Read the number” is the condition, and within 3 s for each number with 85 % accuracy for five consecutive intervention sessions is the criterion for mastery.

*Step 3: Selecting a data collection system.* Once the school psychologist operationally defines the target behavior, the next step is to select the data collection system from many available options (see Cone, 2001). In general, a data collection system has three components: a recording method, a recording instrument, and a recording schedule (Miltenberger, 2012). The first component of a data collection system is the *recording method*. The recording method refers to the procedure used to measure a student’s performance of the target behavior. Multiple recording methods exist that allow a school psychologist to measure different characteristics or dimensions of a behavior. The most often used methods include frequency, duration, interval recording, latency, magnitude, and topography (Brown et al., 2014).

*Frequency* refers to the number of times a behavior occurs in a predetermined period of time. When recording frequency, a school psychologist counts and records each occurrence of the target behavior during a specific observation

period (e.g., 30-min. instructional time). Examples of behaviors that can be measured using frequency recording include hand-raising, requests for assistance, words spelled correctly, math problems solved correctly, out-of-seat behavior, off-task behavior, or episodes of tantrums. Frequency recording is appropriate when the length of the observation period is constant across sessions or days. However, one of the advantages of frequency recording is that it allows a school psychologist to compare data collected across observation periods of various lengths by converting frequency to rate. Converting frequency to rate consists of dividing the number of times the target behavior occurred by the length of the observation period. For example, if Amy raised her hand 10 times during the 40-min science class on Monday, her rate is 0.4 per minute (10 occurrences divided by 40 min.). If she raised her hand five times during a 30-min math class on Tuesday, her rate is 0.6 per minute (5 occurrences divided by 30 min.).

*Duration* refers to the amount of time a behavior is performed or how long a behavior lasts. When recording duration, a school psychologist simply records the time when the behavior begins and ends. For example, the school psychologist starts the timer when Steve begins completing his math assignment and stops the timer when he discontinues working on the assignment. Duration data indicate that Steve worked on his math assignment for 20 min. Duration is appropriate for behaviors that are continuous or high frequency, and the purpose of the intervention is to increase or decrease the length of time a child performs a target behavior. Examples of behaviors for which duration recording is appropriate include on-task behavior, off-task behavior, social interaction, cooperative learning, cooperative play, rocking, crying, or tantrums.

*Interval recording* consists of recording the occurrence or nonoccurrence of the target behavior within specified time intervals. When using interval recording, a school psychologist divides the observation period in equal intervals and records whether the behavior occurs within those intervals. Depending on the type of interval recording, the psychologist may select to record



whether the behavior occurs during any part of the interval (i.e., partial interval recording), throughout the interval (i.e., whole interval recording), or at the end of the interval (i.e., time sampling). The length of the intervals usually ranges from 5 to 30 s depending on the characteristics of the target behavior. Examples of behaviors that can be measured using interval recording include on-task behavior, off-task behavior, completing a task, physical aggression, self-injurious behavior, or social interactions.

*Latency* refers to the amount of time that elapses between the presentation of an instruction and initiation of the target behavior. In other words, latency refers to how long it takes a child to engage in a behavior. When recording latency, a school psychologist documents the length of time between the end of an instruction and when the child began following the direction. For example, after being asked to start completing addition problems, Joe needed 20 min to start working. Latency is appropriate when the purpose of the intervention is to increase the amount of time an individual complies with a specific instruction or to decrease latencies that are too short. For example, a child may raise his or her hand during instructional time when a teacher asks a question but before she ends the question, and thus engages in incorrect responding.

*Magnitude* refers to the intensity or force with which a behavior is emitted. For example, a school psychologist may record how loud a child's vocalizations are or how intense is a self-injurious behavior such as banging one's head against a desk. Examples of behaviors for which magnitude may be appropriate include speech or other vocalizations, physical aggression, self-injurious behavior, using a pen to write, or using a keyboard to type. It is important to note that measuring the magnitude of a target behavior may sometimes result in subjective data that are difficult to quantify. For example, when asking a child to rate the intensity of pain inflicted by a peer during physical aggression using a scale, a school psychologist may obtain a subjective measure based on the child's sensitivity and tolerance to pain.

*Topography* refers to the shape or form of a target behavior or what the behavior looks like. For example, a school psychologist may use topography recording to collect data on the form of various behaviors displayed by a child during a tantrum or the letter formation during cursive handwriting. Topography recording is appropriate when collecting data on behaviors that must meet specific topographical criteria (Cooper, Heron, & Heward, 2007). Examples of behaviors that can be measured using topography recording include cursive handwriting or correct posture.

Because many behaviors can be measured using multiple recording methods, the selection of the appropriate method depends on the characteristics of the target behavior and the purpose of the evaluation (Cone, 2001; Kazdin, 2011). For example, a school psychologist may use frequency to record the number of correct words per minute read by a child within 10 min, if he or she is interested in measuring the child's progress on the number of words acquired within a specified period of time (e.g., 2 months). However, the school psychologist may use duration to record the amount of time necessary for a child to read a paragraph from a story book, if he or she is interested in measuring the reading fluency after the acquisition of newly taught words.

The second component of a data collection system is the *recording instrument*. The recording instrument is a data sheet or tool that allows a school psychologist to document one or multiple dimensions of a target behavior depending on the purpose of the evaluation. There are multiple formats of data sheets that can be adapted or created to meet the requirements of a specific situation. Most data sheets have several common components including the name of the student for which data are collected, the name of the recorder, the date of recording, a legend listing the behaviors recorded and their operational definitions, and directions on how to record data on the target behavior. Regardless of its format, each data sheet should be simple and easy to use and yield accurate and objective data to assist school psychologists and other professionals in making data-based decisions. Several resources are

available to assist school personnel with development of data collection tools. Examples of resources include the Intervention Central (<http://www.interventioncentral.org>), National Center on Student Progress Monitoring (<http://www.studentprogress.org/families.asp>), and PBIS (<https://www.pbis.org>).

The third component of a data collection system is the *recording schedule*. The recording schedule refers to the frequency with which data are collected on a target behavior. In general, this decision is made based on the type of target behavior on which data are collected. The guidelines published in the professional literature suggest that the most effective recording schedule consists of frequent data collection such as every two or three days for most social, academic, or other classroom related behaviors (Fuchs & Fuchs, 1986). In some cases, especially when a new intervention is first introduced to teach a skill or when the target behavior is problematic or dangerous, a school psychologist may decide to collect data every day or every intervention session until a change in behavior is documented (Brown et al., 2014; Farlow & Snell, 1994).

One aspect of data collection that merits further discussion refers to challenges associated with the data collection process in clinical and applied settings including schools. One challenge of data collection relates to the accuracy and reliability of data. Variables such as unintended changes in the way data are collected, recorder's expectations of how data should look like before and after the intervention, or student's awareness the he or she is observed may result in data that do not reflect an accurate level of student's performance on the target behavior (Kazdin, 2011). A second challenge of data collection relates to lack of knowledge and skills required to collect data including the recording schedule, analysis of data, and use of data in decision-making (Sandall, Schwartz, & Lacroix, 2004). A third challenge refers to the amount of time necessary to collect data on individual goals. Specifically, practitioners report that data collection can be too time consuming and sometimes interferes with other job responsibilities (Cooke, Heward, Test, Spooner, & Courson, 1991). Thus, it is important for school psychologists to become

familiar and fluent with the most effective and feasible data collection systems to overcome various challenges encountered by professionals when implementing these procedures in applied settings and to provide training, assistance, and feedback to teachers who may be directly involved in data collection on student outcomes.

*Step 4: Selecting a design or evaluation strategy.* Continuous data collection is essential to program evaluation and data-based decision-making. Yet, as we mentioned previously, data collection is only one aspect of program evaluation. The second aspect of program evaluation refers to establishing a functional relation between a specific intervention and a child's performance on a target behavior. Ideally when evaluating the effectiveness of an intervention, a school psychologist makes statements about the effects of an intervention on a child's level of performance on a target behavior as compared to the effects of no intervention or alternative interventions, thus determining the presence or the absence of a functional or causal relation between an intervention and a behavior (American Psychological Association; APA, 2002). A functional relation exists when a child's level of performance on a specific behavior has changed when, and only when, the intervention was introduced. In other words, the intervention is responsible for the change in the child's behavior.

It is important to note that continuous data collection to measure intervention outcomes allows school psychologists to evaluate the direction and the magnitude of a child's progress after an intervention was introduced. However, data collection alone does not allow school psychologists to determine whether a functional relation exists between the intervention and target behavior (Alberto & Troutman, 2013). To document a functional relation, school psychologists must use experimental methods within certain design structures. The decision on whether to use a non-experimental design or an experimental design depends on the purpose of program evaluation, the skills and knowledge of the person implementing the program, and the complexity of variables characteristic to the applied setting in which an intervention is implemented. In most

applied settings, it is extremely difficult to use an experimental SCD to evaluate a treatment (Kratochwill & Piersel, 1983). However, various causal inference procedures can be used to increase the credibility of the conclusions that are drawn from nonexperimental SCDs (Kazdin, 1981; 2011).

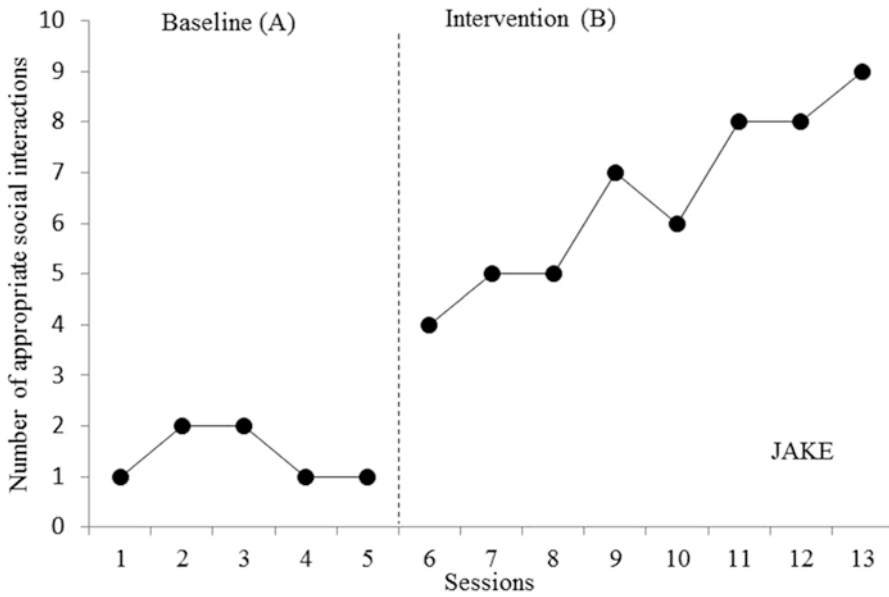
Consider the following complexities of practice for evaluation of an intervention: if the goal is to increase the number of times a child raises his or her hand to ask questions during instructional time, the school psychologist may use a data collection system to measure the child's performance before and after the intervention within the context of a pre-post nonexperimental design. However, if the goal is to demonstrate that a child's on-task behavior across three different settings (i.e., reading class, math class, and science class) has increased when, and only when, the teacher used positive reinforcement in the form of specific verbal praise and the change in the on-task behavior is not the result of other variables (e.g., peer attention), the school psychologist may decide to measure individual outcomes repeatedly across the three classroom settings. The repeated measurement itself can increase the inference that praise was responsible for the increase in on-task behavior. However, an even stronger procedure would be to stagger in time the point at which teacher used the positive reinforcement in each of these settings (a SCD that is called a multiple-baseline procedure; see below).

Numerous types of SCDs exist that allow a school psychologist to measure individual outcomes and have demonstrated effectiveness for measuring and monitoring individual progress on a target behavior (Riley-Tillman & Burns, 2009). SCDs are experimental methods consisting of various designs involving repeated measures of a specific behavior under different conditions to evaluate the effectiveness of a treatment for an individual or a small group of individuals that serve as their own control (Kazdin, 2011). Multiple types of SCDs ranging from simple to complex can be used in applied settings to evaluate interventions and document functional relations between a behavior and an intervention. These designs will require replica-

tion to establish functional or experimental control. Again, in most school and applied settings, it can be very challenging to use any type of experimental SCD that involves formal replication and would meet the rigorous standards advanced for research. We mention these designs because they might be used under some rare and unusual circumstances in school settings for a clinical purpose rather than in the context of formal research.

*Basic or case study designs.* One of the most basic SCDs is the *AB design*. The AB design consists of two phases: the A phase (i.e., baseline) and the B phase (i.e., intervention). Baseline represents the phase of the design in which the school psychologist collects data on the target behavior performed by a child before the intervention is introduced, and it represents the child's current or existing level of performance. The B phase represents the phase of the design in which the school psychologist collects data on the target behavior after the intervention has been introduced. The AB design allows a school psychologist to make some inferences about the effectiveness of a specific intervention by comparing the child's performance before and after the intervention. Based on data collected, the school psychologist can also make decisions about the continuation, revision, or interruption of the intervention. The AB design is one of the most often used designs in applied settings because it is easy to implement and does not require extensive trainings, and it can provide school psychologists and other professionals an accurate and objective visual representation of a student's performance on a target behavior when these data are presented in a graph. Although it does not meet the rigor of experimental methods, the AB design is a feasible alternative that allows the evaluation of practices in applied settings.

Consider an example of an AB design (see Fig. 1) in which a school psychologist may collect data on the number of times a student initiates appropriate interactions with an adult during break. In this case, the school psychologist records baseline data for 3–5 days to document the number of appropriate interactions initiated by the student during break (the A phase). Next,



**Fig. 1** Number of appropriate social interactions with an adult during break

the school psychologist introduces an intervention consisting of specific verbal praise (e.g., “Jake, I like the way you asked permission to go outside”) contingent on the student’s appropriate interaction with an adult and continues to record the number of appropriate interactions (the B phase). Finally, the school psychologist can analyze data collected and make revisions to instruction as necessary. The AB design could allow the psychologist to make some assumptions about the effectiveness of the intervention when an increase in the number of appropriate interactions is documented during intervention compared to baseline.

The AB design is a nonexperimental procedure, and thus, it cannot be used to demonstrate a functional relation between a target behavior and a specific intervention. Although data may indicate that a student’s performance has increased after the intervention has been introduced thereby suggesting that the intervention is responsible for the change in behavior, a school psychologist cannot be confident that the increase in the target behavior is indeed the result of the intervention and not the result of other variables. For example, a school psychologist may implement an intervention to increase the in-seat behavior for a

student with ADHD. She collected baseline data and determined that the student stayed in his seat an average of 3 min during the 15-min. independent work. The psychologist implemented an intervention consisting of a token economy system in which the student could earn one token for each 5-min in-seat behavior for a total of up to three tokens. Data indicated that the student was able to sit in his chair an average of 10 min. during the 15-min., independent work as soon as the intervention was introduced, and thus concluded that the intervention was effective. However, the same day when the intervention was implemented, the student’s parents reported that he had a cold and the doctor prescribed some medication that made the student sleepy. In this situation, the medication may be responsible for the increase in the in-seat behavior, and therefore, the school psychologist cannot be confident that the token economy system produced the change in the student’s behavior.

Despite the fact that the AB design does not demonstrate a functional relation between a behavior and an intervention, practitioners working in applied settings, including school psychologists, can draw valid inferences from AB designs by following several procedures (Kazdin, 2011). First, it

is important to *collect systematic data* on the target behavior. Collecting systematic data allows a school psychologist to document whether a change in behavior occurred after the intervention was introduced, but it does not provide information on why or how the change has occurred. Consider the case when a school psychologist collects systematic data on academic task completion after the implementation of a self-monitoring system for a middle-school student during math instruction. In this case, the school psychologist may document an increase in the percentage of tasks completed by the student shortly after self-monitoring was introduced, but he or she cannot draw any conclusions of why the increase in task completion occurred or explain how the change in task completion happened.

Second, *continuous measurement* of the target behavior has the potential to improve the quality of inferences drawn from an AB design as opposed to measurement at two points in time (i.e., pre- and posttest). Continuous measurement of the target behavior provides a school psychologist the opportunity to identify data patterns and to analyze whether the change in pattern coincided with the introduction of an intervention. On the other hand, collecting data on one or two occasions makes the process of ruling out alternative explanations for changes in target behavior very difficult. For example, a school psychologist assessing a student's reading fluency by administering a pre- and a posttest may not be able to rule out the possibility that changes in the assessment procedures lead to an increase in the student's performance on the posttest compared to the pretest.

Third, a school psychologist should consider an individual's *past and future projections of performance* when drawing inferences from AB designs. For example, a student experiencing chronic academic failure as documented by his or her performance on tests and assignments has a history of a low-level performance in academic areas, thus suggesting that the student's low-level performance is likely to continue unless an intervention is implemented. In this case, if a change in academic performance is documented after the onset of an intervention, the inference that the

intervention is responsible for change is much stronger compared with a situation in which the student's previous or past performance is not documented and no comparison can be made.

Fourth, *immediacy and magnitude of change* are two important variables to consider when drawing inferences from AB designs. In general, the more immediate a change in the level of performance after the introduction of an intervention, the stronger an inference can be drawn that the intervention rather than other variables was responsible for change in behavior. The same logic applies to the magnitude of change. Specifically, a larger change in the level of performance after intervention allows for a more valid inference about the effectiveness of intervention compared to a smaller change. Please see below for a more detailed description of immediacy and magnitude of change.

Finally, *the number and the type of individuals* for whom the same intervention is implemented may influence the validity of inferences about the effectiveness of an intervention. Specifically, when an intervention is implemented with more than one individual or with individuals with different demographic characteristics, the inference that the intervention was responsible for change is much stronger compared to situations in which the intervention is implemented with only one individual. For example, a school psychologist implementing a token economy system to increase the on-task behavior for three students during science instruction is much more confident that the intervention produced an increase in the on-task behavior compared to a situation in which the token economy was implemented only with one student.

*Classroom designs.* It may be useful to evaluate the effectiveness of universal interventions by considering classroom-level outcome data, and evaluation of academic and behavioral concerns on the classroom level can be practical and efficient. The following description on the "B design," or Tier 1 practice, is discussed in detail by Riley-Tillman and Burns (2009). Within a B design, it is recognized that some intervention is currently in place, where some formal instructional practices are instituted in every regular



classroom. To assess the effectiveness of Tier 1 practices, whole class outcome data are collected from unit tests and standardized assessments. Riley-Tillman and Burns (2009) offer three critical decisions to consider when developing a B design. First, the selection of the target should preferably be a whole class or group. From gathering data of a whole group, all of the students' progress and their patterns can be monitored, in addition to providing important information on individual students who deem to be outliers or display problematic behavior. After the target is determined, some assessment method that feasibly and repeatedly measures that behavior is selected (e.g., office discipline referrals of a whole class). Finally, the target group must be compared to some standard, such as some growth benchmark like a standardized norm group or to other students in same educational setting. Although a B design can show what the target behavior looks like in a standard (Tier 1) environment, its limitation is that it cannot distinguish whether a particular intervention is responsible for any observed change in the target behavior or whether the change in the behavior would occur in the absence of the intervention. More formal options for evaluation of outcomes are discussed in the next section of the chapter.

*Experimental single-case designs.* In addition to case studies or AB designs, several experimental designs may be implemented in applied settings in certain circumstances. Table 1 displays the most common SCDs used in applied settings and their defining characteristics. The SCDs are represented in Fig. 2 (reversal design), Fig. 3 (multiple-baseline design across participants), and Fig. 4 (alternating treatments design). These designs each involve a replication component and can be used to establish a functional or experimental relation between the independent and dependent variable. They are used extensively in SCD experimental research in psychology and education.

*Step 5: Evaluating data.* Once the school psychologist has selected the most appropriate and feasible SCD to determine the effectiveness of a specific intervention, the next step is to evaluate data within the context of the selected design.

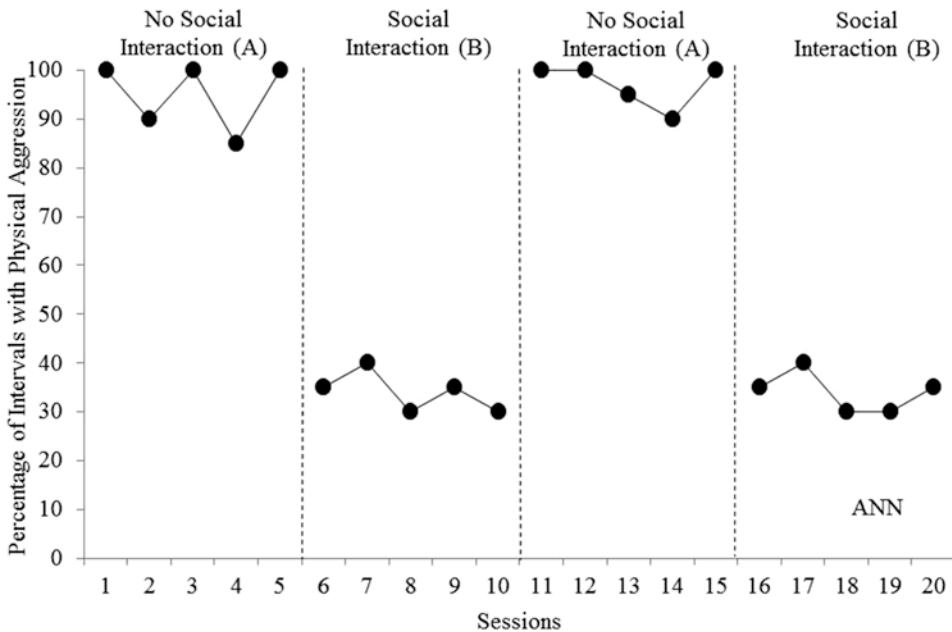
The traditional method to evaluate data in SCDs consists of visual analysis. When using visual analysis to determine the effectiveness of an intervention, a school psychologist should evaluate several characteristics of data within and across phases including level, trend, variability, overlap, immediacy of effect, and consistency across similar phases (Kazdin, 2011; Kennedy, 2005). In recent years, more formal procedures have been advanced to assist with the visual analysis of data in designing studies and conducting literature reviews in which SCDs are used (see Kratochwill et al., 2010). *Level* refers to the magnitude and direction of change in a student's performance from the end of one phase to the beginning of the next phase. A large and immediate change in level is an indicator of an effective intervention. For example, if the percentage of correct single-digit additions increased from 30% at the end of the baseline phase to 70% at the beginning of the intervention phase, this would be considered an immediate and large change in level.

*Trend or slope* is the consistent decrease or increase in performance and requires at least three data points. A consistent performance in the desired direction after an intervention has been introduced suggests that the intervention was successful. For example, if data indicate that John engaged in 20, 18, and 17 episodes of verbal aggression during small group instruction, the school psychologist would argue that data suggest a decreasing trend in the desired direction. On the other hand, if data indicate that John engaged in 20, 30, and 35 episodes of verbal aggression during small group instruction, the school psychologist would say that data suggest an increasing trend in the opposite direction than the desired one. It is also possible that data may indicate no trend. For example, the number of episodes of aggressive behavior could remain relatively constant (e.g., 20, 21, 19) which would suggest no trend.

Numerous methods exist to calculate a trend to determine whether a student makes progress toward his or her intervention goal. One simple and straightforward method to calculate a trend is to use Microsoft Excel, a software program that

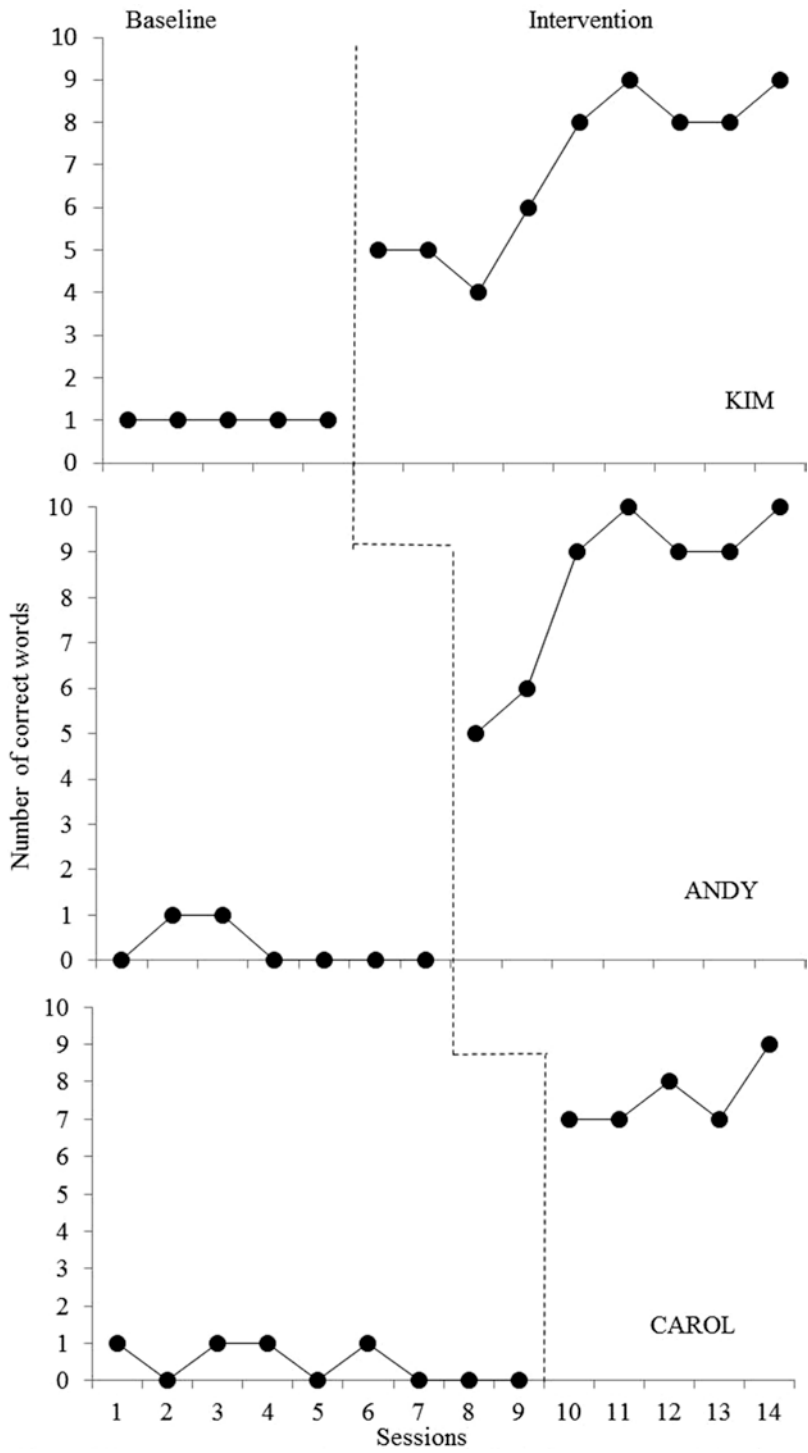
**Table 1** Types of single-case designs

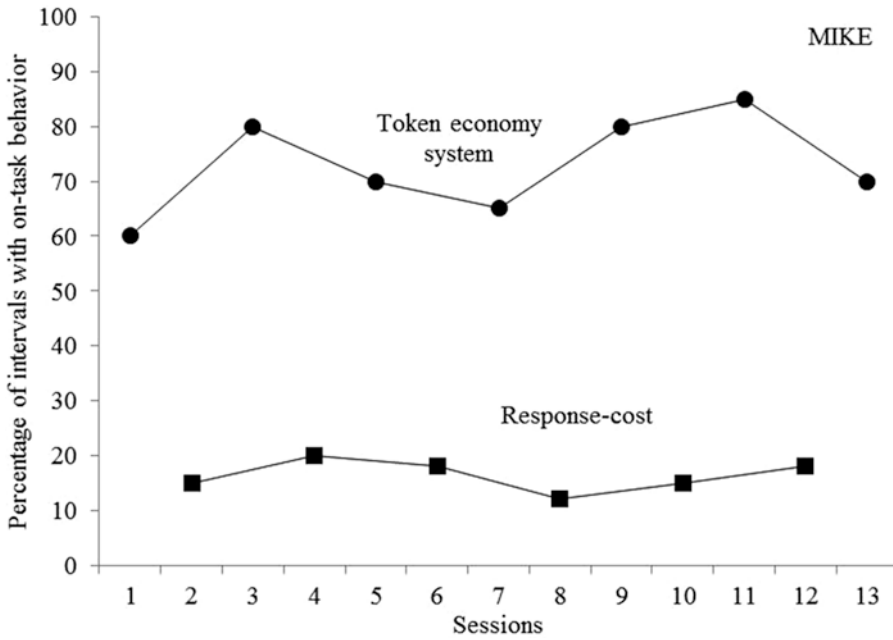
Design	Definition	When to use the design
Reversal (e.g., ABAB)	Examines the effectiveness of an intervention on a single behavior by replicating a baseline phase (i.e., A) and an intervention phase (i.e., B) that are repeated at least twice	When the intervention can be discontinued and the behavior can be reversed to baseline conditions
Multiple baseline across participants, settings, or behavior	Examines the effectiveness of an intervention across several baselines by implementing it at different points in time for each baseline	When the purpose of the intervention is to experimentally test the effectiveness of an intervention across participants, behaviors, and settings and when the behavior cannot be reversed
Alternating treatment	Consists of two or three interventions that are alternated in a counterbalanced or random order	When comparing the effectiveness of two or more interventions that can be alternated relatively rapidly



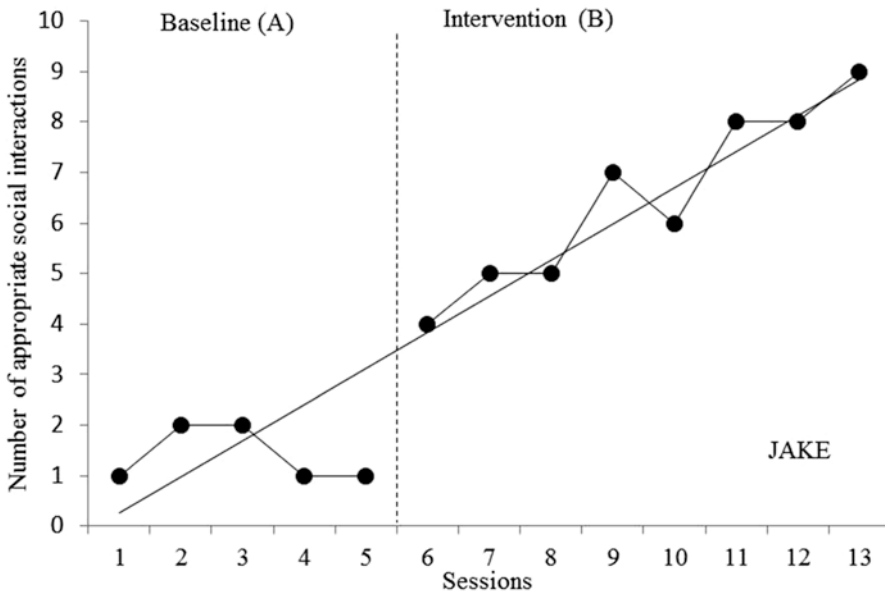
**Fig. 2** Percentage of intervals with physical aggression across two conditions: no social interaction (a) and social interaction (b)

**Fig. 3** Number of correct words read independently during one-on-one instruction





**Fig. 4** Percentage of intervals with on-task behavior across two interventions: token economy system and response cost



**Fig. 5** Number of appropriate social interactions with an adult during break

can be used to record data. Figure 5 shows a trend generated in Excel that allows a school psychologist to make decisions about Jake’s progress on initiating appropriate social interactions. To calculate a trend, the school psychologist (a) entered

data in a spreadsheet and generated a line chart, (b) selected the chart displaying Jake’s data, (c) chose the “Layout” option available in “Chart Tools”, and (d) added a linear trendline (see Cummings & Martinez, 2012 for additional information).

*Variability* refers to the fluctuation of data from one intervention session to the next one or the range of data from the best-fit straight line (Horner & Spaulding, 2010). Data characterized by high variability are difficult to interpret and to evaluate the effectiveness of an intervention compared to data that indicate low or no variability. For example, reading fluency scores of 10, 35, 2, 50, and 20 indicate high variability, whereas scores of 10, 12, 15, 11, and 13 indicate low variability and have the potential to demonstrate a clear functional relation between reading fluency and a specific intervention.

*Overlap* refers to the percentage of data from the intervention phase that overlaps with the data from the baseline phase. An intervention is considered to be effective if data indicate a small percentage of overlapping data as opposed to larger percentages of overlapping data. For example, baseline data indicate that the percentage of on-task behavior ranges from 20 to 35. In the first intervention phase (i.e., verbal praise), seven of ten data points (70% overlap) fall within the same range as baseline, whereas in the second intervention phase (i.e., token economy), one of ten data points (10% overlap) falls within the same range as baseline. The overlap percentages suggest that the token economy may be more efficient to increasing the on-task behavior compared to verbal praise.

*Immediacy of effect* refers to the length of time between the onset or termination of one phase and the change in behavior. When determining the immediacy of effect, a school psychologist evaluates the change between the last three data points in one phase and the first three data points in the next phase (Horner & Spaulding, 2010). A more immediate change is an indicator of a more effective intervention compared to a delayed change. *Consistency across similar phases* refers to a data pattern across all baseline and all intervention sessions within a design. Greater consistency across similar phases indicates a clearer functional relation between a behavior and a specific intervention.

It is important to consider all the abovementioned elements when using visual analysis to evaluate the effectiveness of an intervention. In

addition to the visual analysis, a school psychologist should also assess the clinical or practical significance of the intervention. Specifically, it is important to determine whether the change in a student's behavior as a result of an intervention is large enough to make a difference in the student's daily functioning and improve his or her level of functioning and quality of life. Methods for assessing the clinical significance of an intervention may include surveys, questionnaire, and interviews with the students, educators, or parents and professional judgment (see Kazdin, 2011, for more information on social validity of intervention outcomes).

*Data-based decision-making.* An effective intervention requires ongoing data collection and analysis to make data-based informed decisions leading to successful individual outcomes. Data-based decision-making consists of revisions or modifications of an intervention to increase the likelihood of goal attainment. For example, if data indicate that a student's level of performance has increased after the intervention was implemented, then the school psychologist may decide to continue the intervention in its current form without any revisions. However, the school psychologist may decide to modify or revise the intervention when data indicate that the student's performance has not changed after the intervention was introduced. Revisions may also be necessary when data indicate that the student's level of performance after the intervention decreased compared to his or her performance before the intervention was introduced.

Data-based decision-making using SCD data is sometimes a very complex process due to the lack of concrete decision rules for determining the strength of a functional effect (Kazdin, 1998). When data-based decision-making using SCD data is used in applied settings including schools, it is recommended that school psychologists and other professionals use dual criteria to evaluate the effectiveness of an intervention, namely, the overall change in the target behavior and the social validity of the outcome and/or the social consequence of the target behavior (Alberto & Troutman, 2013; Stoiber & Kratochwill, 2002).



## Conclusion

Measuring individual, classroom, and school-based outcomes using some basic elements of SCD data and larger classroom or systemic approaches may assist school psychologists and other professionals with program evaluation and data-based decision-making and are practical and important methodologies that may be used in the schools. The core elements of program evaluation and data-based decision-making are: defining the target behavior, selecting a data collection system, selecting an appropriate design or evaluation strategy, and evaluating data. It is important for the school psychologist to become familiar and fluent in using the methods described in this chapter to be able to collect objective data and make informed decisions and to train other professionals working with children and youth in applied settings.

## Test Yourself Quiz

1. What are the five steps involved in designing interventions using single-case design methodology to address students' academic or social behavior in school settings?
2. Describe the six elements of visual analysis used to determine the effectiveness of an intervention within the context of a single-case design.
3. What are some of the variables to consider when implementing an outcomes' assessment within a school setting?
4. Consider what the differences are between individual, classroom, and system-wide outcomes assessment, and what would guide your decision-making process in choosing outcomes assessment.

## References

- Albers, C. A., & Kettler, R. J. (2014). Best practices in universal screening. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Data-based and collaborative decision making* (pp. 121–131). Washington, DC: National Association of School Psychologists.
- Alberto, P., & Troutman, A. C. (2013). *Applied behaviour analysis for teachers* (9th ed.). Upper Saddle River, NJ: Pearson Education, Inc.
- American Psychological Association. (2002). Criteria for evaluating treatment guidelines. *American Psychologist*, *57*, 1052–1059. doi:10.1037/0003-066X.57.12.1052.
- Australian Psychological Society. (2013). *Framework for the effective delivery of school psychological services*. Flinders Lane, VIC: The Australian Psychological Society Limited.
- Batsche, G. M., Elliot, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., et al. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria, VA: National Association of State Directors of Special Education.
- Berryhill, J. C., & Prinz, R. J. (2003). Environmental interventions to enhance student adjustment: Implications for prevention. *Prevention Science*, *4*, 65–87.
- Bradshaw, C., Mitchell, M., & Leaf, P. (2010). Examining the effects of school-wide positive behavioural interventions and supports on student outcomes. *Journal of Positive Behavior Interventions*, *12*, 133–148.
- Brown, R., Steege, M. W., & Bickford, R. (2014). Best practices in evaluating the effectiveness of interventions using single-case methods. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Foundations* (pp. 371–380). Washington, DC: National Association of School Psychologists.
- Campbell, J. M., & Hammond, R. K. (2014). Best practices in rating scale assessment of children's behaviour. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Data-based and collaborative decision making* (pp. 287–304). Washington, DC: National Association of School Psychologists.
- Cappella, E., Massetti, G. M., & Yampolsky, S. (2009). Rigorous, responsive, and responsible: Experimental designs in school intervention research. In L. M. Dinella (Ed.), *Conducting science based psychology research in schools* (pp. 51–78). Washington, DC: American Psychological Association. doi:10.1037/11881-003.
- Castillo, J. M. (2014). Best practices in program evaluation in a model of response to intervention/multitiered system of supports. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Foundations* (pp. 329–342). Washington, DC: National Association of School Psychologists.
- Chafouleas, S. M., Riley-Tillman, T. C., & Christ, T. J. (2009). Direct Behavior Rating (DBR): An emerging method for assessing social behaviour within a tiered intervention system. *Assessment for Effective Intervention*, *34*, 195–200. doi:10.1177/1534508409340391.
- Commonwealth of Australia. Australian Education Act 2013, No. 67, 2013. Retrieved from [https://www.comlaw.gov.au/Details/C2013A00067/Html/Text#\\_Toc361054816](https://www.comlaw.gov.au/Details/C2013A00067/Html/Text#_Toc361054816)
- Cone, J. D. (2001). *Evaluating outcomes: Empirical tools for effective practice*. Washington, DC: American Psychological Association. doi:10.1037/10384-000.

- Cooke, N., Heward, W. L., Test, D. W., Spooner, F., & Courson, F. H. (1991). Student performance data in the classroom: Measurement and evaluation of student progress. *Teacher Education and Special Education, 14*, 155–161. doi:10.1177/088840649101400301.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behaviour analysis* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Cordray, D. S., & Pion, G. M. (2006). Treatment strength and integrity: Models and methods. In R. R. Bootzin & P. E. McKnight (Eds.), *Strengthening research methodology: Psychological measurement and evaluation* (pp. 103–124). Washington, DC: American Psychological Association. doi:10.1037/11384-006.
- Cummings, J. A., & Martinez, R. S. (2012). Visual representation of progress monitoring and academic achievement data. In R. Brown-Chidsey & K. J. Andren (Eds.), *Assessment for intervention: A problem-solving approach* (2nd ed., pp. 321–343). New York, NY: Guilford Press.
- Deno, S. L. (1986). Formative evaluation of individual student programs. A new role for school psychologists. *School Psychology Review, 15*, 358–374.
- Dix, K. L., Slep, P. T., Lawson, M. J., & Keeves, J. P. (2011). Implementation quality of whole-school mental health promotion and students' academic performance. *Child and Adolescent Mental Health, 17*(1), 45–51.
- Durlak, J. A., & Dupre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting the implementation. *American Journal of Community Psychology, 41*, 327–350. doi:10.1007/s10464-008-9165-0.
- Erchul, W. P., & Martens, B. K. (2010). *School consultation: Conceptual and empirical bases of practice* (3rd ed.). New York, NY: Springer Science + Business Media.
- Fagan, T. K., & Wise, P. S. (2000). *Chapter 4: Roles and functions of school psychologists* (School psychology: Past, present, and future). Bethesda, MD: NASP Publications.
- Farlow, L., & Snell, M. (1994). *Making the most of student performance data*. Washington, DC: American Association on Mental Retardation.
- Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*(6), 488–500.
- Fuchs, L., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53*, 199–208.
- Gibbons, K., & Brown, S. (2014). Best practices in evaluating psychoeducational services based on student outcome data. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Foundations* (pp. 355–370). Washington, DC: National Association of School Psychologists.
- Glover, T. A., & DiPerna, J. C. (2007). Service delivery for response to intervention: Core components and directions for future research. *School Psychology Review, 36*(4), 526–540.
- Goffreda, C. T., & DiPerna, J. C. (2010). An empirical review of psychometric evidence for the Dynamic Indicators of Basic Early Literacy Skills. *School Psychology Review, 39*(3), 463–483.
- Gresham, F. M., & Elliott, S. N. (2008). *Social skills improvement system: Rating scales*. Bloomington, MN: Pearson Assessments.
- Hamre, B. K., & Pianta, R. C. (2007). Learning opportunities in preschool and early elementary classrooms. In R. Pianta, M. Cox, & K. Snow (Eds.), *School readiness & the transition to school* (pp. 49–84). Baltimore: Brookes.
- Hamre, B. K., Pianta, R. C., & Chomat-Mooney, L. (2009). Conducting classroom observations in school-based research. In L. M. Dinella (Ed.), *Conducting science based psychology research in schools* (pp. 79–105). Washington, DC: American Psychological Association. doi:10.1037/11881-004.
- Horner, R. H., & Spaulding, S. A. (2010). Single-case research designs. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 1386–1394). Thousand Oaks, CA: SAGE.
- Howell, K. W., & Hosp, J. L. (2014). Best practices in curriculum-based evaluation. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Data-based and collaborative decision making* (pp. 159–171). Washington, DC: National Association of School Psychologists.
- Howell, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York, NY: Guilford Press.
- Howell, K. W., Hosp, J. L., & Kurns, S. (2008). Best practices in curriculum-based evaluation. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vth ed., pp. 349–362). Bethesda, MD: National Association of School Psychologists.
- Individuals With Disabilities Education Act, 20 U.S.C. § 1400 (2004).
- Kazdin, A. E. (1981). Drawing valid inferences from case studies. *Journal of Consulting and Clinical Psychology, 49*, 183–192. doi:10.1037/0022-006X.49.2.183.
- Kazdin, A. E. (1998). *Research design in clinical psychology* (3rd ed.). Boston: Allyn & Bacon.
- Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York, NY: Oxford University Press.
- Kennedy, C. (2005). *Single-case designs for educational research*. Boston: Allyn & Bacon.
- Kettler, R. J., Glover, T. A., Albers, C. A., & Feeney-Kettler, K. A. (2014). An introduction to universal screening in educational settings. In R. J. Kettler, T. A. Glover, C. A. Albers, & K. A. Feeney-Kettler (Eds.), *Universal screening in educational settings: Evidence-based decision making for schools* (pp. 3–16). Washington, DC: American Psychological Association. doi:10.1037/14316-001.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M. et al. (2010). Single-case designs technical documentation. In *What Works Clearinghouse: Procedures and standards handbook (version 2.0)*. Retrieved from What Works Clearinghouse website: [http://ies.ed.gov/ncee/wwc/pdf/wwc\\_procedures\\_v2\\_standards\\_handbook.pdf](http://ies.ed.gov/ncee/wwc/pdf/wwc_procedures_v2_standards_handbook.pdf).

- Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., et al. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34*, 26–38. doi:10.1177/0741932512452794.
- Kratochwill, T. R., & Piersel, W. C. (1983). Time-series research: Contributions to empirical clinical practice. *Behavioral Assessment, 5*, 165–176.
- Mager, R. (1997). *Preparing instructional objectives* (3rd ed.). Atlanta, GA: The Center for Effective Performance, Inc.
- Martens, B. K., & McIntyre, L. L. (2009). The importance of treatment integrity in school-based behavioural interventions. In A. Akin-Little, S. Little, M. Bray, & T. Kehle (Eds.), *Behavioral interventions in schools: Evidence-based positive strategies* (pp. 59–71). Washington, DC: American Psychological Association. doi:10.1037/11886-004.
- McConaughy, S. H., & Ritter, D. R. (2008). Best practices in multimethod assessment of emotional and behavioural disorders. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vth ed., pp. 697–715). Bethesda, MD: National Association of School Psychologists.
- McDougal, J. L., Bardos, A. N., & Meier, S. T. (2010). *Behavior intervention monitoring assessment system (BIMAS)*. Toronto: Multi-Health Systems.
- McKeivitt, B. C., & Braaksma Fynaardt, A. (2014). Best practices in developing a positive behaviour support system at the school level. In P. L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Systems-level services* (pp. 165–180). Washington, DC: National Association of School Psychologists.
- Merrell, K. W. (2008). *Behavioral, social, and emotional assessment of children and adolescents* (3rd ed.). New York, NY: Lawrence Erlbaum Associates.
- Miltenberger, R. G. (2012). *Behavior modifications: Principles and procedures* (5th ed.). Belmont, CA: Thomson/Wadsworth.
- National Assessment Program – Literacy and Numeracy (NAPLAN). (2013). Retrieved from <http://www.nap.edu.au/>.
- National Association of School Psychologists. (2010). *Model for comprehensive and integrated school psychological services*. Bethesda, MD: National Association of School Psychologists. Retrieved from: [http://www.nasponline.org/standards/2010standards/2\\_PracticeModel.pdf](http://www.nasponline.org/standards/2010standards/2_PracticeModel.pdf).
- Neely, L., Davis, H., Davis, J., & Rispoli, M. (2015). Review of reliability and treatment integrity trends in autism-focused research. *Research in Autism Spectrum Disorders, 9*(1), 1–12. doi:10.1016/j.rasd.2014.09.011.
- No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).
- Reynolds, C. R., & Kamphaus, R. W. (2009). *BASC-2 progress monitor manual*. Bloomington, MN: Pearson.
- Riley-Tillman, C., & Burns, M. K. (2009). *Evaluating educational interventions: Single-case designs for measuring response to intervention*. New York, NY: Guilford Press.
- Sandall, S. R., Schwartz, I. S., & Lacroix, B. (2004). Interventionists' perspectives about data collection in integrated early childhood classrooms. *Journal of Early Intervention, 26*, 161–174. doi:10.1177/105381510402600301.
- Sanetti, L. M. H., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction To the special series. *School Psychology Review, 38*, 445–459. Retrieved from: <http://www.nasponline.org/>.
- Sanetti, L. M., & Kratochwill, T. R. (Eds.). (2014). *Treatment integrity: A foundation for evidence-based practice in applied psychology*. Washington, DC: American Psychological Association.
- Shapiro, E. S. (2003). *Behavioral Observation of Students in Schools (BOSS)*. Computer Software. San Antonio, TX: Psychological Corporation.
- Shapiro, E. S., & Heick, P. F. (2004). School psychologist assessment practices in the evaluation of students referred for social/behavioural/emotional problems. *Psychology in the Schools, 41*, 551–561. doi:10.1002/pits.10176.
- Sheridan, S. M., Swanger-Gagné, M., Welch, G. W., Kwon, K., & Garbacz, S. A. (2009). Fidelity measurement in consultation: Psychometric issues and preliminary examination. *School Psychology Review, 38*(4), 476–495.
- Shinn, M. R. (Ed.). (1998). *Advanced applications of curriculum-based measurement*. New York: Guilford.
- Shinn, M. R. (2007). Identifying students at risk, monitoring performance, and determining eligibility within response to intervention: Research on educational need and benefit from academic intervention. *School Psychology Review, 36*, 601–617.
- Shinn, M. R. (2010). Building a scientifically based data system for progress monitoring and universal screening across three tiers, including RTI using curriculum-based measurement. In M. R. Shinn & H. M. Walker (Eds.), *Interventions for achievement and behaviour problems in a three-tier model including RTI* (pp. 259–292). Bethesda, MD: National Association of School Psychologists.
- Slee, P. T., Lawson, M. J., Russell, A., Askill-Williams, H., Dix, K. L., Owens, L., et al. (2009). *KidsMatter Primary Evaluation Final Report*. Centre for Analysis of Educational Futures, Flinders University of South Australia.
- Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*(8), 795–819. doi:10.1002/pits.20113.
- Stoiber, K. C., & Kratochwill, T. R. (2002). *Outcomes: Planning, monitoring, evaluating*. San Antonio, TX: The Psychological Corporation.
- Sugai, G., & Horner, R. H. (2006). A promising approach for expanding and sustaining school-wide positive behaviour support. *School Psychology Review, 35*, 245–259.
- Sugai, G., Horner, R. H., & Lewis-Palmer, T. (2001). *Team implementation checklist*. Eugene, OR: Educational and Community Supports, University of Oregon.

- Sugai, G., Lewis-Palmer, T. L., Todd, A. W., & Horner, R. H. (2001). *School-wide evaluation tool (SET)*. Eugene, OR: Center for Positive Behavioral Supports, University of Oregon.
- Technical Assistance Center on Positive Behavioral Interventions and Supports. (2010). *Implementation blueprint and self-assessment: Positive behavioural interventions and supports*. Washington, DC: Technical Assistance Center on Positive Behavioral Interventions and Supports. Retrieved from [http://www.pbis.org/common/pbisresources/publications/SWPBS\\_ImplementationBlueprint\\_vSep\\_23\\_2010.pdf](http://www.pbis.org/common/pbisresources/publications/SWPBS_ImplementationBlueprint_vSep_23_2010.pdf).
- University of Oregon Center on Teaching and Learning. (2009). *Dynamic indicators of basic early literacy skills*. Retrieved January 10, 2009, from <http://dibels.uoregon.edu/>.
- Upah, K. R. F. (2008). Best practices in designing, implementing, and evaluating quality interventions. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vth ed., pp. 209–223). Bethesda, MD: National Association of School Psychologist.
- VanDerHeyden, A. M., & Burns, M. K. (2005). Using curriculum-based assessment and curriculum-based measurement to guide elementary mathematics instruction: Effect on individual and group accountability scores. *Assessment for Effective Intervention, 30*(3), 15–31. doi:10.1177/073724770503000302.
- VanDerHeyden, A., & Harvey, M. (2013). Using data to advance learning outcomes in schools. *Journal of Positive Behavior Interventions, 15*(4), 205–213. doi:10.1177/1098300712442387.
- Vollmer, T. R., Roane, H. S., Ringdahl, J. E., & Marcus, B. A. (1999). Evaluating treatment challenges with differential reinforcement of alternative behavior. *Journal of Applied Behavior Analysis, 32*(1), 9–23. doi:10.1901/jaba.1999.32-9.
- Volpe, R. J., DiPerna, J. C., Hintze, J. M., and Shapiro, E. S. (2005). Observing students in classroom settings: A review of seven coding schemes. *School Psychology Review, 34*, 454–474.
- Waldron, N., & Prus, J. (2006). *A guide for performance-based assessment, accountability, and program development in school psychology training programs* (2nd ed.). Bethesda, MD: National Association of School Psychologists.
- Walker, H. M., & Shinn, M. R. (2010). Systemic, evidence-based approaches for promoting positive student outcomes within a multitier framework: Moving from efficacy to effectiveness. In M. R. Shinn & H. M. Walker (Eds.), *Interventions for achievement and behaviour problems in a three-tier model including RTI* (pp. 1–26). Bethesda, MD: National Association of School Psychologists.
- Watson, T. S., & Watson, T. S. (2009). Behavioral assessment in the schools. In A. Akin-Little, S. G. Little, M. A. Bray, & T. J. Kehle (Eds.), *Behavioral interventions in schools: Evidence-based positive strategies* (pp. 27–41). Washington, DC: American Psychological Association. doi:10.1037/11886-002.
- Wilder, D. A., Atwell, J., & Wine, B. (2006). The effects of varying levels of treatment integrity on child compliance during treatment with a three-step prompting procedure. *Journal of Applied Behavior Analysis, 39*, 369–373. doi:10.1901/jaba.2006.144-05.
- Wolery, M., Bailey, D. B., & Sugai, G. M. (1988). *Effective teaching: Principles and procedures of applied behaviour analysis with exceptional children*. Boston: Allyn & Bacon.