Chapter XX: Making Sense of Multiple Data Sources: Using Single Case Design Research for Behavioral Decision Making

Elizabeth Talbott
https://orcid.org/0000-0002-9163-3414

Andres De Los Reyes
https://orcid.org/0000-0003-0214-983X

Abstract
Data from multiple school sources, including informant ratings, systematic direct observations (SDOs), and school wide data (e.g., office disciplinary referrals [ODRs]) are routinely used to guide decision making in the delivery of evidence based practices for students with externalizing and internalizing behavior problems. Over 50 years of research show that these data sources are likely to diverge. Yet, there is little to no guidance from research about how to interpret and use these divergent data to adapt and tailor evidence based interventions for individual students. We describe how researchers can employ Kraemer et al.'s (2003) framework to guide the selection of data sources and single case design (SCD) research to analyze data from these sources. We then apply the Operations Triad Model (OTM) to interpret teacher ratings and SDOs in SCD and recommend future directions for research in behavioral decision making using multiple sources of data within SCD.

Keywords: Single Case Research Design, Operations Triad Model, Converging Operations, Diverging Operations, Data Based Decision Making

Evidence based practice consists of three core features: (a) evidence based interventions, (b) evidence based assessment, and (c) evidence based analysis (Riley-Tillman & Johnson, 2017). In the multi-tiered systems of support (MTSS) framework, assessment is at the heart of evidence based practice. Behavioral assessment methods are aligned with the three tiers of intervention and are focused on (a) screening procedures for all students to determine who is at greatest risk (Lane et al., 2014); (b) progress monitoring for individuals, to determine whether they are making adequate progress in response to interventions (Chafouleas et al., 2021); and (c) diagnostic procedures, such as functional behavioral assessment (FBA), to develop hypotheses about contexts and contingencies that prompt and maintain problem behavior (Dunlap & Kern, 2018; Hanley et al., 2003). Throughout the MTSS framework, the data based decision making process relies on informant ratings, systematic direct observations (SDOs), and ongoing review of extant data (e.g., office disciplinary referrals [ODRs]) conducted by a range of school professionals, including special and general education teachers, school psychologists, social workers, counselors, and paraprofessionals (see Christ et al., 2009; Dunlap & Kern, 2018; Lane et al., 2014).

Yet, we know from more than 50 years of educational and psychological research that diverse sources of data (e.g., survey reports from parent, teacher, and youth informants) are likely to provide estimates of child and adolescent behavior that disagree with one another (Achenbach et al., 1987; De Los Reyes et al., 2015; Talbott et al., 2018). Informants disagree even when the measures used are psychometrically sound, suggesting that measurement confounds do not fully explain the disagreements (Achenbach, 2011; Dirks et al., 2012). Without clear guidance about how to select informants and analyze and interpret their data for decision making about behavior, practitioners may discard those data that do not fit their expectations,

and make intervention decisions based on the informant who, based on intuition, is deemed the most "insightful" (Marsh et al., 2020).

Likewise, informant ratings may or may not disagree with data from SDOs (Chafouleas et al., 2005; Riley-Tillman et al. 2008). Historically, the discrepancies commonly observed across diverse data sources have created considerable uncertainties, and as such they have posed barriers to the development and implementation of precise, tailored, evidence based interventions for students with behavioral disabilities (De Los Reyes et al., 2019a). Yet, these discrepancies (or lack thereof) among data sources can reveal meaning in children's behavior across contexts as they interact with different individuals (De Los Reyes et al., 2009). By *context*, we mean setting events (i.e., structure of a given setting, such as physical arrangements, adult to student ratio, schedules, and behavioral expectations); ongoing interactions among adults and children in a given setting; and intervention effects, which may be positive, negative, small, large, or negligible (Sims et al., 2017).

Thus, rather than viewing them as barriers, we view these discrepancies as opportunities to analyze and interpret behavior in context (De Los Reyes et al., 2019b). Indeed, as others have noted (Sims et al., 2017), the implementation of evidence based interventions essentially consists of a series of experiments. Within these experiments, the practitioners who deliver them typically use multiple sources of data (e.g., teacher ratings and SDOs). When these data are not aligned, disagreements are less likely to be seen as the measurement confounds that they are often interpreted to reflect in the larger literature (Achenbach, 2011). Rather, they are likely to serve as *prompts* to continue probing the concerns experienced by the student receiving services; to observe and record contingencies for behavior in *context* as part of a diagnostic procedure, such as an FBA.

This aspect of school-based service delivery—the potential for disagreements among data sources to improve our understanding of students' needs—serves as important "proof of concept" for the utility of multi-informant approaches to assessment. In fact, these disagreements reflect a key reality of this approach: informants provide non-redundant information about students' needs. Unfortunately, there has been little guidance from research about how to capitalize on the knowledge gained from this non-redundant information. That is, if disagreements reflect information relevant to understanding students' needs, how might we systematically select, collect, and analyze school data to tailor and individualize behavioral interventions?

Indeed, we recognize that behavior is embedded in diverse and complex school contexts. Perhaps because of this, we often seek out informants like teachers or construct FBAs *on purpose*. We take measurements in relevant contexts, or we seek out the ratings of key sources embedded in these contexts. We can even point to scholars who make explicit the value of these measurements for attaining a context-sensitive understanding of students' behavior. For instance, Carr (1994) originally described contexts for behavioral assessment as including the "countless stimulus parameters, both environmental and biological, that operate both within and outside of assessment sessions that can interact with one another" (p. 395). School contexts experienced by youth with behavior problems can vary widely with regard to: (a) degree of structure, ranging from whole class, small group, individual instruction, and down time; (b) focus, on academic or non-academic activities with opportunities to engage and respond; and (c) management, with features of positive teacher-student relationships and evidence based behavior management (Baker et al., 2008a; Baker et al., 2008b; Maggin et al., 2011; Oliver et al., 2011).

Within the range of diverse school contexts, comprised of settings that are more and less structured, are continuous interactions between adults and students that influence behavior in the moment, all of which set the stage for behavioral contingencies (De Los Reyes et al., 2019a; Dunlap & Kern, 2018; Stichter & Conroy, 2005). Thus, the contexts within which students' needs manifest have a discernable structure. As such, there exists the potential for drawing explicit, systematic connections between these structures, school-based services, and the measures used to not only inform the development of these services but also their implementation with individual students. In fact, paradigms exist for facilitating these connections. Specifically, recent work has illustrated the value of SCD research in identifying patterns in teacher ratings and SDOs for progress monitoring in the context of diverse school contexts (i.e., Fabiano et al., 2017; Miller et al., 2017; Sims et al., 2017). SCD offers a dynamic intra-individual approach to analyzing informant ratings collected alongside SDOs in these settings. Researchers can simultaneously collect data on the contingencies for behavior and use those to interpret convergence and divergence in data from informants and observations. The central thesis of this chapter is the micro-level analytic structure provided by SCD may be the *lynchpin*: the final piece in the puzzle of how to leverage multi-source measurement to optimize the implementation of evidence based practices and generate clear and important directions for future research.

We propose the following approach in this chapter to address critical gaps in evidence based practice, specifically evidence based assessment and analysis in behavioral decision making. First, we describe a framework designed to guide the selection of data sources for decision making about student behavior (Kraemer et al., 2003), which the developers based on well established patterns of cross-source disagreement observed across decades of research (e.g.,

Achenbach et al., 1987). Work conducted over the last decade demonstrates that disagreements among these data sources reflect variations among informants in the specific contexts in which they observe behavior, as well as the perspectives through which they observe behavior (for a review, see De Los Reyes et al., 2019b). Second, we describe how SCD facilitates analyzing data from different sources in a prospective fashion, alongside descriptions of the contexts and contingencies for behavior associated with those data. Third, we apply a framework (Operations Triad Model [OTM]; De Los Reyes et al., 2013) designed to guide evidence based interpretations of findings from school informants and SDOs, namely the degree to which these sources converge or diverge in their ratings and what these patterns of convergence/divergence reflect. We conclude by recommending future directions for research in behavioral decision making using multiple sources of data within SCD.

Contexts and Perspectives for Rating and Observing Behavior

Kraemer and colleagues (2003) have proposed a mix-and-match approach to guide the selection of evidence based decision making. Kraemer's approach represents a significant advance from previous approaches, which were based on few guidelines from research and frequently resulted in researchers and clinicians discarding discrepant sources of data (e.g., youth reports) in favor of data from a single informant (e.g., parent reports; Makol et al., 2020; Marsh et al., 2020). In their approach to identifying appropriate informants, Kraemer and colleagues take into consideration the broad contexts for behavior (i.e., school and home) and the perspectives of raters (i.e., self and other) to guide the selection of optimal informants to describe behavior. In Table 1, we provide an illustration of how data typically collected in school contexts for behavioral decision making fit within Kraemer's model. <Insert Table 1 here.> Our adaptation of this model includes the various contexts within school settings, which include more

structured (e.g., different types of classes) and less structured (e.g., hallway, lunch, and playground) settings, along with a range of data sources. These data sources include the perspectives of diverse informants (e.g., youth and teacher), those who conduct SDOs, and diverse sources of behavioral data (e.g., ODRs) and academic data (e.g., grade point averages, high stakes test scores, curriculum-based measurement)—all of which can be used for intervention decision making (see Lane et al., 2014). Within these various school contexts, interactions between adults and youth (including peer interactions) and the application of more and less effective interventions establish contingencies for behavior (Kazdin, 2013).

Using Kramer's (2003) approach, we have intentionally identified a subset of data sources (informant ratings and SDOs) to rate, score, and describe behavior in particular contexts. Alongside or within these data sources, we can collect data on the antecedents, behavior, and consequences for behavior (ABC; Hanley et al., 2003) to describe specific contexts and contingencies that maintain behavior. In our application of Kraemer's framework, the informant and/or observer selected is considered an optimal choice to describe the child's behavior in a given context. The ABC analysis provides additional descriptive information to illuminate contingencies for behavior. Then, as informants and observers converge and diverge in their ratings and observations, this creates the opportunity for individual adaption and tailoring of interventions, along with providing feedback and coaching to teachers to support their use of evidence based interventions.

Operations Triad Model (OTM): A Framework for Interpretation

If the framework developed by Kraemer and colleagues (2003) facilitates selecting sources to provide ratings to guide service delivery, the OTM (De Los Reyes et al., 2013) provides a framework to guide interpretation of data when two or more different data sources are

used to assess a child's behavior. In our application of this model, informant ratings and SDOs are the two sources of data collected during the same observation period in the same class setting. The OTM in Figure 1 provides a useful framework for interpreting patterns in the data whether informant ratings and observations converge or diverge. <Insert Figure 1 here.> Specifically, the OTM delineates several patterns or ways in which information sources might agree or disagree in their ratings, along with guidance on how to ensure that any disagreements observed can be done so meaningfully.

In particular, the OTM illuminates three distinct patterns of expected research findings with informant ratings. In the first pattern, *Converging Operations*, informant ratings align consistently with SDOs in the same context in which observations are conducted. This pattern reflects consistencies in the contingencies that prompt and maintain problem behavior across the information sources, in this case informant ratings and SDOs. Conversely, in the second pattern, *Diverging Operations* reflect instances in which informant reports diverge from the results of SDOs in the same context. This pattern reflects discrepancies in the contingencies that prompt and maintain problem behavior across informant ratings and SDOs.

The following example reveals both convergence and divergence in informant ratings and SDOs that reflect the contingencies in which observations occur (e.g., setting events [class structure], human interactions, and interventions). In this example, two teachers, one a general educator (head teacher) and the second a special educator (co-teacher), along with a research assistant (RA), observe and rate the academic engagement and disruptive behavior of a special education student as they (teachers) are engaged in typical class instruction and activities. In this example, ratings by the head teacher *converge* with the observations of the RA, but *diverge* from the ratings of the special educator. The special educator rates the student as showing consistently

higher rates of academic engagement and lower rates of disruptive behavior compared to ratings

and observations by the head teacher and the RA. As it turns out, the head teacher and the RA

completed their ratings and observations from the perspective of interacting (head teacher) and

observing (RA) the entire class. In contrast, the special educator's ratings occurred within the

class structure/setting event of interactions with a small group of special education students.

Thus, the nature of the assessment design allows us to meaningfully interpret the discrepancies

among information sources (i.e., Diverging Operations). Specifically, discrepancies between the

head teacher and special educator reflected meaningful variations in the ability of data sources to

capture various kinds of contingencies eliciting behavior in the same context (i.e., class-wide vs.

individual).

However, what if assessment designs are not conducive to meaningful interpretations of

discrepancies, such that the variation in measurement methods provide a parsimonious

explanation for any rating discrepancies? In these cases, a third pattern of interpretation using the

OTM might apply—*Compensating Operations*—wherein divergence in ratings and observations

can reflect measurement confounds. By "measurement confounds" we refer to discrepancies that

reflect characteristics of the measurement process. For example, it may be that the SDO captures

features of the environmental context that the informant rating is unable to capture, not because

of a meaningful variation in contingencies, but because of methodological features of the

assessment process. This might happen when teachers are asked to make an assessment of

students' behavior following a relatively lengthy observation period (e.g., as long as a class

period or half of the school day), compared to SDOs by an examiner which capture behavior

within a brief (e.g., 10-15 minute) observation period (Chafouleas et al., 2021). Essentially, the

"design" of the two measures contains a different amount of "items" or periods of observation,

and thus there would be reason to believe that the two measures might have provided similar data if they were based on the same observation period and within similar contingencies. The distinctions between Diverging and Compensating Operations are important to make, because an assessment process that is improperly designed could result in erroneous decision making, particularly if observed discrepancies are treated meaningfully when, in fact, they could simply manifest as a function of measurement confounds. Although the previous example provided the appropriate kinds of information sources, these sources varied too much in how they arrived at their ratings to draw meaningful inferences about the observed discrepancies. Taken together, the OTM provides a framework with which we can interpret results from SCD studies.

<h3 style="text-align:center">Single Case Design Research and the OTM</h3>

The OTM highlights the need to design assessment processes that result in the ability to meaningfully interepret discrepant ratings across information sources. In this respect, SCDs provide the opportunity for analysis of informant ratings and SDOs in the context of establishing causal links between interventions and behavior (Kazdin, 2011). SCDs also set the stage for examination of contingencies for behavior when data sources converge and diverge. Researchers who use SCD methods rely on the following experimental features: (a) repeated measurement of dependent variables collected over time; (b) repeated manipulation of experimental conditions; and (c) examination of intraindividual data using visual analysis techniques (Horner et al., 2005; Kazdin, 2011). As a result, SCD is a sensitive and readily interpretable means of evaluating a range of data characteristics in addition to intervention effects (Maggin & Bruhn, 2018).

In our application of SCD for evidence based assessment and analysis, data are collected from one informant (i.e., a teacher) using a progress monitoring tool at the same time a second person conducts SDOs of the same behavior in a given school setting. Because key features of a child's behavior problems include both the context(s) in which the child's behavior manifests

and the contingencies that maintain that behavior (De Los Reyes et al., 2019a), we recommend that observers also record patterns in behavior associated with setting events (class structure and corresponding behavioral expectations), student behavior in the context of human interactions, and contingencies for that behavior in order to explain and describe convergence and divergence in the two data sources. To do that, researchers engage in ABC analysis alongside or embedded within SDOs (Hanley, 2003; for an example see Carroll et al., 2006). The ABC analysis facilitates the assessment of patterns in behavior associated with the environmental contexts and contingencies alongside informant (teacher) ratings and SDOs. In this way, the ABC analysis essentially takes the role of an independent data source, or a means by which to collect data independent of the informant ratings and SDOs. This is a key element of the approach taken in OTM-based research conducted in the last decade (for reviews, see De Los Reyes et al., 2019b). As such, examining relations between data from an ABC analysis and the patterns observed between informant ratings and SDOs helps to explain convergence and divergence between these information sources.

For example, in the case of Converging Operations, a teacher and observer might agree in their ratings of a student's off-task verbal behavior. During a particular observation period, the ABC analysis may reveal that the teacher became engaged in an interaction with the student about work completion, which consisted of persistent verbal reminders on the part of the teacher and arguing on the part of the student. The student behavior in this example is not disruptive, but the teacher-student interaction resulted in high ratings of off-task behavior for the student and both data sources captured this. On the other hand, in the case of Diverging Operations, disagreement between informant ratings and observations might occur if the teacher plays a key role in the contingencies that govern the behavior being rated. For instance, perhaps teacher

behavior during class time (e.g., moving away from the student who attempts to engage in off-task arguing, openly praising other students who are on-task) prompts the off-task student to reengage and thus decrease their off-task behavior. In this contingency situation, the teacher's actions decreased the student's off-task behavior but in the process, this created circumstances that the teacher was unable to observe. At the same time, the ABC analysis captured the contingences of the interaction and the observer's rating showed relatively low rates of off-task behavior. Not surprisingly, the teacher continued to rate the student's off-task behavior as high. This discrepancy between the teacher and observer rating reflects meaningful variations in opportunities to observe contingencies governing the student's behavior. When treated as such (i.e., as opposed to interpreting it as a "rater bias" on the part of the teacher), this creates an opportunity for behavioral coaching and feedback to the teacher regarding the effective management of this student's off-task behavior.

In the case of Compensating Operations, measurement confounds may also be related to the type of behavior being assessed, in addition to measurement features described in Table 1. For example, what if the teacher and observer were assigned to provide ratings, but they got the scheduling wrong, such that the teacher made ratings of the student's off-task behavior, and the observer made the ratings of a completely different target behavior (e.g., aggressive behavior during recess)? The "measurement confound" in this example is reflected in Kraemer's (2003) framework. It includes (a) the perspective of the teacher compared to an observer about two different target behaviors and (b) the specific context for behavior (i.e., structured class setting compared to less structured recess). Disagreements in this case are likely to be reduced by employing parallel instruments to measure the same behavior within settings. Even with parallel instruments, a long line of research indicates that we might still expect informants and SDOs to

disagree, given the various contingencies for behavior embedded within class structures (i.e.,

class compared to recess), interactions with peers and adults, and more or less effective

interventions (see also De Los Reyes et al., 2020).

Data Sources: Informant Ratings and SDOs

Assessment methods employed within an evidence based problem solving model,

particularly those used for progress monitoring, should be defensible, with adequate

psychometric characteristics; feasible to employ, for both researchers and practitioners; efficient,

or relatively quick and easy to use; and repeatable, useable over time (Chafouleas et al., 2011;

Christ et al., 2009). As Frick and colleagues (2020) have noted, SDOs have held a revered status

in clinical and behavioral assessment, considered by some to be a gold standard, unbiased form

of assessment. Yet, we know that there is no gold standard for behavioral assessment (see

Achenbach, 2011), and that SDOs have both their strengths and limitations (Frick et al., 2020).

That is, not all observation systems are defensible, created with careful attention to basic

psychometrics, and few of them have established reliability and validity in multiple samples

(Frick et al., 2020).

Informant ratings and SDOs conducted by diverse informants are both used for progress

monitoring; we have selected a handful of these tools for illustrative purposes that are

appropriate to use in the context of SCD. Our purpose is not to review the universe of these tools,

but to illustrate the interpretative approach we advocate through the use of these tools. The

National Center on Intensive Intervention (NCii, 2020) has identified three tools commonly used

in behavioral progress monitoring. The first is the Behavior Assessment System for Children

(BASC-3), which is designed to assess a range of adaptive and problem behavior in the school

setting. The BASC-3 includes informant ratings from parent, teacher, and youth perspectives,

along with a procedure for conducting SDOs (Reynolds & Kamphaus, 2015). The SDO procedure from the BASC-3 consists of 15-minute observation periods divided into 30-second intervals. In each interval, observers are charged with checking a box next to each category of behavior that occurred during that interval and noting whether the behavior was disruptive (Reynolds & Kamphaus, 2015). Examples of behavioral categories assessed with the BASC-3 include: response to the teacher/lesson; appropriate peer interactions; work on school subjects, appropriate behavior during class transitions, inappropriate behavior and vocalizations, inattention, somatization, repetitive behavior, aggression, self-injurious behavior, inappropriate sexual behavior, and bowel/bladder problems (Reynolds & Kamphaus, 2015). Observers are also charged with providing as much detail as possible about the teacher's interaction with the student, focusing on contingencies in the class that may affect the child's behavior.

The second method reviewed by the NCii is the Direct Behavior Rating-Single Item Scales (DBR-SIS), a measure which serves as a combination of direct observation and teacher rating. DBR-SIS requires teachers to rate the proportion of time a student is engaged in specific behavior (i.e., academic engagement, disruptive and respectful behavior) during a specified observation period. The observation period can be comprised of a relatively brief or longer period of time, such as part of a class or an entire class period or a larger portion of the school day (e.g., first half of the day; Chafouleas et al., 2021). Notably, the conceptual and empirical basis for the DBR-SIS is strong (i.e. demonstrating defensibility, usability, repeatability, and flexibility), with its psychometric characteristics well-established (Briesch et al., 2016). The

DBR-SIS is a progress monitoring tool used in three SCD studies alongside SDOs which we will review in a subsequent section of this chapter.

The third tool reviewed by the NCii is momentary time sampling, a non-continuous observation method of SDO which is commonly used in research (Harrop & Daniels, 1986). Regardless of whether the data collection method is continuous (e.g., frequency count, duration or latency of behavior) or non-continuous (e.g., time sampling), the essential components of developing SDOs include identifying the what, where, how, and by whom of an observation system (Frick et al., 2020). Researchers begin by defining *what* to observe, followed by *where* to observe (settings and contexts), *how* the observations will be conducted, and *by whom*. The observation system includes operational definitions of the target behavior to be observed along with a method of ongoing assessment of the reliability of the measurement system (Horner et al., 2005). To be considered reliable and valid, observation systems need to include the following: (a) comprehensive training, monitoring, and evaluation of observers; (b) graduated implementation of the observation systems (i.e., implemented in stages); (c) transition from the training setting to the applied setting, contingent on reaching a criterion level of observer agreement and accuracy; (d) recalibration of accuracy and agreement of observer recordings with identification and correction of any breakdowns in fidelity; (e) periodic assessment of the observation system and retraining of observers as necessary (Frick et al., 2020). Ensuring psychometric rigor in SDOs requires ongoing training of observers and systematic, ongoing assessment of interobserver agreement.

Observational systems that focus on the behavior and interactions of fellow students and teachers, in addition to the behavior of target students, help to identify the ABCs of behavior: those antecedents (stimuli) that evoke a behavior and those consequences (responses) by adults

and other children that serve to maintain the behavior (Carroll et al., 2006; Frick et al., 2020).

This systematic assessment of antecedents and consequences that reliably predict and maintain

behavior are key features of evidence-based assessment (Riley-Tilman & Johnson, 2017). As

data from these and comparable tools are recorded and graphed over time, researchers observe

convergence and divergence in informant ratings compared to SDOs.

Application of Single Case Design Research

In a recent special issue of *Assessment for Effective Intervention*, authors of three studies

(Fabiano et al., 2017; Miller et al., 2017; Sims et al., 2017) employed rigorous SCD methods

(What Works Clearinghouse, WWC, Institute of Education Sciences, U.S. Department of

Education, 2014) to examine the effectiveness of the Daily Report Card (DRC) intervention

(Chafouleas et al., 2005) on two different outcome measures: (a) teacher ratings using the DBR-

SIS and (b) SDOs conducted by an external observer. These studies provide an excellent

opportunity to analyze evidence based assessment in the context of SCD from two different

sources. We examined the convergence and divergence in data collected using teacher ratings

and SDOs in these three studies.

In each of the studies, researchers employed a multiple baseline design to systematically

introduce the DRC. In two of the three studies (Fabiano, 2017; Sims et al., 2017) researchers

delivered the intervention to elementary students and in one (Miller et al., 2017) researchers

delivered the intervention to both elementary and secondary students. Both Fabiano and Sims

examined the effects of the DRC intervention on academic engagement, disruptive behavior, and

respectful behavior, whereas Miller examined the effects of the DRC for academic engagement

and disruptive behavior only. Researchers in all three studies used the same definitions for the

target behaviors on the DBR-SIS as the SDOs. Methods of observation were comparable across

the three studies, with observations conducted during 15-20 minute sessions using momentary time sampling (Harrop & Daniels, 1986). In each of the three studies, the SDOs reflected the percentage of the observation period in which behavior was observed. Likewise, the DBR-SIS data reflected the percentage of the observation period in which the target behavior was displayed according to teachers. Although teachers completed the DBR-SIS every day, SDOs were conducted less often (e.g., in Miller's study, once per week). We examined data from the three studies for convergence and divergence, applying visual analysis to each of the data points for the 15 students across the three target behaviors assessed (academic engagement, disruptive behavior, and respectful behavior).

Visual Analysis

Visual analysis is the traditional approach used to determine the effects of experimental control in SCD research; it can also be used to observe patterns in the data measured by one or more dependent variables (Kazdin, 2011). Visual analysis of data from these three studies provides the opportunity to apply the OTM framework for interpretation within SCD (identifying occasions where the data sources converge and diverge), while illustrating the need for future research to examine the contexts and contingencies for behavior, particularly when data sources disagree. For example, the Fabiano et al. (2017) study revealed that data from teacher ratings and SDOs consistently diverged (i.e., data points differed by more than one percentage point[1]) for all three students in the measurement of all three target behaviors. For academic engagement, just 3 of 15 data points (33%) converged within one point for disruptive behavior, 7 of 16 data points (44%) converged, and for respectful behavior, 6 of 18 data points (33%) converged within one

---

[1] 1% was the criterion we applied to examine these data; that criterion may have differed from the authors.

point. The authors reported that teacher ratings appeared to provide different estimates of response to intervention compared to SDOs, a pattern which also varied across student participants (Fabiano et al., 2017).

Similarly, Miller et al. (2017) examined data for academic engagement and disruptive behavior rated by teachers and observers for four elementary students and found that just 3 of 34 data points (9%) converged within one percentage point for engagement and 2 of 13 data points (15%) converged for disruptive behavior. Miller also reported that for these elementary students, estimates of the duration of academic engagement and disruptive behavior tended to be higher using teacher ratings compared to SDOs, with both data sources revealing improvement in disruptive behavior due to the intervention. For secondary students' academic engagement, 6 of 28 data points (21%) converged within one point. In contrast to the elementary students, Miller found that estimates of the duration of academic engagement and disruptive behavior tended to be higher using SDOs compared to teacher ratings (Miller et al., 2017). Secondary students' response to intervention also differed depending on the data source, with teacher ratings indicating modest improvement in academic engagement and SDOs revealing that academic engagement actually became worse following the intervention (Miller et al., 2017).

There was greater convergence in teacher ratings and observations from Sims' (2017) study with four elementary students, wherein researchers found that 23 of 45 data points (51%) converged within one point for academic engagement; 20 of 45 data points (44%) converged for disruptive behavior; and 24 of 45 data points (53%) converged for respectful behavior. Sims used only the teacher ratings to evaluate the effectiveness of the intervention, and concluded that when the two sources of data diverged by more than three percentage points, interpretation of results was challenging.

Interpretation of results from the two data sources is facilitated by collecting additional and more frequent observation data within SCD to determine the contexts and contingencies for behavior. The purpose of such data collection is to determine *specific cases* in which (a) divergence in the data reflects different perspectives on student behavior that in turn reflect the context for observation (i.e., class structure, interactions, and interventions) and/or (b) divergence in the data reflects measurement confounds associated with Compensating Operations. Miller's (2017) study with secondary students illustrates the critical need for such analysis. In particular, what happened *in context and contingencies* to result in such highly discrepant results, wherein teachers rated behavior as improving under the intervention condition and observers rated behavior as getting worse? Two forms of data collection can be used to determine Diverging compared to Compensating Operations: (a) collection of ABC analysis data to clarify the specific contingencies for behavior in context, illuminating the characteristics of behavior detected by the two data sources; and (b) more frequent collection of SDO data alongside the teacher ratings to determine whether measurement confounds (e.g., the extent to which the two measures are parallel) are affecting the results. The outcome of these additional data collection and analysis procedures is a more informed process designed to improve upon the work of these and other authors who employ SCD research to intensify, adapt, and change interventions to meet students' needs.

Future Directions for Research

Our examination of teacher ratings and SDOs across these three studies meeting rigorous WWC standards and using psychometrically sound dependent measures mirrors, broadly, the many studies of agreement and disagreement among various informants who rate student externalizing and internalizing behavior (Achenbach et al., 1987; De Los Reyes et al., 2015). It

confirms the unique role that SCD methods play in the study of behavior rated by multiple informants and SDOs in complex school settings. Further targeted analysis of data collected in context can lead to improvement in the delivery and adaptation of evidence based practices for individual students with disabilities. In the following sections, we describe the most pressing research needs, along with recommendations for directions for future research research.

Evidence Based Assessment: The Contributions of Diverse School Informants

One needs only to review the plethora of data educators collect, examples of which we have provided in Table 1, to see the challenges they face in behavioral decision making. For example, federally funded organizations such as the NCii and the Center on Positive Behavior Interventions and Supports, PBIS (www.pbis.org) provide recommendations about what data sources to select for screening, progress monitoring, and diagnostic assessment at Tiers 2 and 3, but offer little to no empirical guidance about how to select and analyze different data sources to adapt and tailor interventions, especially when those data diverge. We have proposed employing a logical framework to select data sources (Kraemer et al., 2003) and SCD to analyze and interpret the data as they converge and diverge.

Obtaining data from multiple informants is considered an evidence based practice in assessment (Mash & Hunsley, 2005), yet there is little empirical guidance for the selection of one type of school informant compared to another, or one type of assessment (e.g., informant ratings) compared to another (SDOs). In addition, there is a clear need for research with multiple teacher informants in diverse school contexts, including the interactions between students and adults that occur moment by moment throughout the school day. Few researchers have compared the ratings of different teachers who rate the same students (see Talbott et al., 2021), yet these efforts are critical for moving the field forward.

Similarly, very few educational researchers systematically seek the perspectives of young people in the school evidence based assessment process, despite evidence of persistent challenges in engaging youth in mental health interventions, including those offered at school (Kim et al., 2015). In order to determine what interventions work, in what context and for what purpose, as well as those interventions in need of further study, we must engage youth in the assessment process associated with the delivery of school mental health services (Becker et al., 2018). In fact, engaging youth in identification and treatment of their "top problems" can help to focus attention on youth concerns; identify specific priorities for treatment; give youth a voice in their own treatment; enhance rapport between school mental health providers and youth; provide a focus for progress monitoring; and help to inform decisions about when to end treatment (Weisz et al., 2011).

Need for Behavioral Progress Monitoring Tools

Unlike the development of progress monitoring tools for academic interventions (see Fuchs & Fuchs, 1999), research on behavioral progress monitoring has not kept pace with intervention development (Owens & Evans, 2018; Riley-Tillman & Johnson, 2017). This is an urgent area for future research, as the progress of youth in response to evidence based intervention depends upon the effectiveness of the measurement systems developed to monitor their progress. There is a clear need for psychometrically strong measures that capture brief informant ratings of target behaviors following a prespecified observation period, as well as SDOs which include assessments of context and contingencies. For example, Carroll and colleagues have developed the Responses to Interpersonal and Physically Provoking Situations (RIPPS) to capture contexts and contingencies for behavior in class observations of students with Attention Deficit Hyperactivity Disorder, ADHD (Carroll et al., 2006; Houghton et al., 2005).

Within the RIPPS, the observer records the following about the class context: (a) whether the

class activity is initiated by students or by teachers; (b) whether the activity is conducted

independently or with the assistance of others, and (c) whether it is conducted at the student's

desk or elsewhere (Carroll et al., 2006). Observers also record whether the student is required to

listen to instructions or wait before beginning the assigned task and whether the student's

responses (and the antecedent events that trigger such responses) occur during a period of class

transition (Carroll et al., 2006). In addition to recording these features of context and

contingency, observers record the occurrence and severity of behavior and note what triggered

student responses (Carroll et al., 2006). The RIPPS has face validity and appears to meet the

criteria identified previously (i.e., defensible, feasible, efficient, and repeatable; Chafouleas et

al., 2011; Christ et al., 2009). It is one example of the type of systematic direct observation that

researchers need to employ so that evidence based interventions can be adapted and tailored to

meet the needs of these individual students in diverse school contexts, such as those we identify

in Table 1.

Evidence Based Analysis: What Do the Data Tell Us? What Works?

SDOs are commonly used to measure intervention effectiveness through SCD. What does

this say about the effectiveness of interventions if dependent measures include teacher or youth

reports or SDOs? What if different sources of data lead to a different conclusion about

intervention effectiveness (i.e., Miller et al., 2017)? On which data do researchers then rely?

Perhaps a given intervention is more sensitive to behavior change measured by one approach

compared to another, as informants are likely to see behavior from a different perspective

compared to observers. This increased sensitivity of one approach relative to another might not

reflect a measurement confound, but rather the exact instantiation of the Kraemer et al. (2003)

approach to selecting information sources to rate students. Indeed, in taking this approach, a user makes a fundamental assumption: One optimizes accuracy in measurement insofar as multiple, distinct data sources triangulate on the rating target (e.g., students delivered an intervention). In such a design, one *forces* disagreement among information sources to occur. It logically follows that purposeful decisions designed to force disagreement among information sources ought to result in discrepant outcomes, even when in comes to estimating the effects of interventions. Yet, use of such an approach requires particular attention to the kinds of measures informants and observers use in the context of SCD, with particular attention to the timing of data collection (i.e., between informant ratings and SDOs). In these respects, future research documenting contexts and contingencies for behavior alongside an assessment of intervention effectiveness can help to answer these questions and shape the focus of evidence based assessment going forward.

Conclusion

In this chapter, we have shown how Kraemer et al.'s (2003) framework can be employed to guide the selection of data sources to describe student behavior in the school context and single case research (SCD) used to analyze data from these sources. The OTM (De Los Reyes et al., 2013) is a practical framework with which to interpret results. Together, these processes represent an advanced and evidence based approach to behavioral assessment, which is a cornerstone of evidence based practices designed to improve outcomes for youth.

References

Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213-232.

Achenbach, T. M. (2011). Commentary: Definitely more than measurement error: But how should we understand and deal with informant discrepancies? *Journal of Clinical Child and Adolescent Psychology, 40*, 80-86.

Baker, J. A., Clark, T. P., Maier, K. S., & Viger, S. (2008a). The differential influence of instructional context on the academic engagement of students with behavior problems. *Teaching and Teacher Education*, *24*, 1876-1883.

Baker, J. A., Grant, S., & Morlock, L. (2008b). The teacher-student relationship as a developmental context for children with internalizing or externalizing behavior problems. *School Psychology Quarterly*, *23*, 3–15.

Becker, K. D., Boustani, M., Gellatly, R., & Chorpita, B. F. (2018). Forty years of engagement research in children's mental health services: Multidimensional measurement and practice elements. *Journal of Clinical Child and Adolescent Psychology*, *47*, 1–23.

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

Carr, E. G. (1994). Emerging themes in the functional analysis of problem behavior. *Journal of Applied Behavior Analysis*, *27*, 393-399.

Carroll, A., Houghton, S., Taylor, M., Hemingway, F., List-Kerz, M., Cordin, R., & Douglas, G. (2006). Responding to interpersonal and physically provoking situations in classrooms:

Emotional intensity in children with Attention Deficit Hyperactivity Disorder. *International Journal of Disability, Development and Education*, *53*, 209–227.

Chafouleas, S. M. (2011). Direct behavior rating: A review of the issues and research in its development. *Education and Treatment of Children*, *34*, 4.

Chafouleas, S. M., Johnson, A. H., Riley-Tillman, T. C., & Iovino, E. A. (2021). *School-based behavioral assessment: Informing prevention and intervention* (2nd Ed). Guilford.

Chafouleas, S. M., McDougal, J. L., Riley-Tillman, T. C., Panahon, C. J., & Hilt, A. M. (2005). What do daily behavior report cards (DBRCs) measure? An initial comparison of DBRCs with direct observation for off-task behavior. *Psychology in the Schools*, *42*, 669–676.

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

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin, 141*, 858-900.

De Los Reyes, A., Drabick, D. A. G., Makol, B. A., & Jakubovic, R. (2020). Introduction to the special section: The Research Domain Criteria's units of analysis and cross-unit correspondence in youth mental health research. *Journal of Clinical Child and Adolescent Psychology*, *49*, 279-296.

De Los Reyes, A., Cook, C. R., Gresham, F. M., Makol, B. A., & Wang, M. (2019a). Informant discrepancies in assessments of psychosocial functioning in school-based services and research: Review and directions for future research. *Journal of School Psychology*, *74*, 74-89.

De Los Reyes, A., Henry, D. B., Tolan, P. H., & Wakschlag, L. S. (2009). Linking informant

discrepancies to observed variations in young children's disruptive behavior. *Journal of

Abnormal Child Psychology*, *37*(5), 637-652.

De Los Reyes, A., Lerner, M. D., Keeley, L. M., Weber, R., Drabick, D. A. G., Rabinowitz, J., &

Goodman, K. L. (2019b). Improving interpretability of subjective assessments about

psychological phenomena: A review and cross-cultural meta-analysis. *Review of General

Psychology*, *23*(3), 293-319.

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. A. (2013). Principles

underlying the use of multiple informants' reports. *Annual Review of Clinical

Psychology, 9*, 123-149.

Dirks, M. A., De Los Reyes, A., Briggs-Gowan, M., Cella, D., & Wakschlag, L. (2012).

Embracing not erasing contextual variability in children's behavior: Theory and utility in

the selection and use of methods and informants in developmental psychopathology.

Journal of Child Psychology and Psychiatry, 53, 558-574.

Dunlap, G., & Kern, L. (2018). Perspectives on (functional) behavioral assessment. *Behavioral

Disorders, 43*, 316-321.

Fabiano, G. A., Pyle, K., Kelty, M. B., & Parham, B. R. (2017). Progress monitoring using

Direct Behavior Rating Single Item Scales in a multiple-baseline design study of the

Daily Report Card Intervention. *Assessment for Effective Intervention*, *43*, 21–33.

Frick P.J., Barry C.T., Kamphaus R.W. (2020). Behavioral observations. In *Clinical

Assessment of Child and Adolescent Personality and Behavior* (4th ed.), 151-169.

Springer.

Fuchs, L. S., & Fuchs, D. (1999). *Using CBM for progress monitoring in reading*.

https://files.eric.ed.gov/fulltext/ED519252.pdf

Hanley, G. P., Iwata, B. A., & McCord, B. E. (2003). Functional analysis of problem behavior: A

review. *Journal of Applied Behavior Analysis*, *36*, 147-185.

Harrop, A., & Daniels, M. (1986). Methods of time sampling: A reappraisal of momentary time

sampling and partial interval recording. *Journal of Applied Behavior Analysis*, *19*, 73-77.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of

single-subject research to identify evidence-based practice in special education.

Exceptional Children, *71*, 165–179.

Houghton, S., Carroll, A., West, J., Taylor, M., & List-Kerz, M. (2005). *The Responses to*

Interpersonal and Physically Provoking Situations Observation Schedule. Unpublished

observation schedule. Perth, WA: Centre for Attention and Related Disorders, University

of Western Australia.

Kazdin, A.E. (2013). *Behavior modification in applied settings* (7th ed.). Waveland Press.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings*

(2nd ed.). Oxford University Press.

Kim, R. E., Becker, K. D., Stephan, S. H., Hakimian, S., Apocada, D., Escudero, P. V., &

Chorpita, B. F. (2015). Connecting students to mental health care: Pilot findings from an

engagement program for school nurses. *Advances in School Mental Health Promotion*, *8*,

87–103.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J.

(2003). A new approach to integrating data from multiple informants in psychiatric

assessment and research: Mixing and matching contexts and perspectives. *American*

Journal of Psychiatry, *160*(9), 1566-1577.

Lane, K. L., Oakes, W. P., Ennis, R. P., & Hirsch, S. E. (2014). Identifying students for

secondary and tertiary prevention efforts: How do we determine which students have Tier

2 and Tier 3 needs? *Preventing School Failure: Alternative Education for Children and

Youth*, 58(3), 171-182.

Makol, B. A., Youngstrom, E. A., Racz, S. J., Qasmieh, N., Glenn, L. E., & De Los Reyes, A.

(2020). Integrating multiple informants' reports: How conceptual and measurement

models may address long-standing problems in clinical decision-making. *Clinical

Psychological Science*, *8*, 953-970.

Maggin, D. M., & Bruhn, A. (2018). Evidence-based assessment and single-case research: A

commentary for the special issue on Direct Behavior Ratings. *Assessment for Effective

Intervention*, 43, 71-78.

Maggin, D. M., Wehby, J. H., Moore Partin, T. C., Robertson, R., & Oliver, R. M. (2011). A

comparison of the instructional context for students with behavioral issues enrolled in

self-contained and general education classrooms. *Behavioral Disorders*, *36*, 84-99.

Marsh, J. K., Zeveney, A., & De Los Reyes, A. (2020). Informant discrepancies in judgments

about change during mental health treatments. *Clinical Psychological Science*, *8*, 318-

332.

Mash, E. J., & Hunsley, J. (2005). Developing guidelines for the evidence-based assessment of

child and adolescent disorders. *Journal of Clinical Child and Adolescent Psychology, 34*,

326-379.

Miller, F. G., Crovello, N. J., & Chafouleas, S. M. (2017). Progress monitoring the effects of

    Daily Report Cards across elementary and secondary settings using Direct Behavior

    Rating: Single Item Scales. *Assessment for Effective Intervention*, *43*, 34–47.

National Center on Intensive Intervention. (2020). *Behavior Progress Monitoring Tools Chart*.

    U.S. Department of Education, Office of Special Education Programs.

    https://intensiveintervention.org/about-charts-resources

Oliver, R. M., Wehby, J. H., & Reschly, D. J. (2011). Teacher classroom management practices:

    Effects on disruptive or aggressive student behavior. *Campbell Systematic Reviews*, *4*, 3-

    55. https://onlinelibrary.wiley.com/doi/pdf/10.4073/csr.2011.4

Owens, J. S., & Evans, S. W. (2018). Progress monitoring change in children's social, emotional,

    and behavioral functioning: Advancing the state of the science. *Assessment for Effective*

    Intervention, *43*, 67–70.

Reynolds, C. R., & Kamphaus, R. W. (2015). *BASC-3: Behavior Assessment for Children (Third*

    Edition). Pearson.

Riley-Tillman, T. C., Chafouleas, S. M., Sassu, K. A., Chanese, J. A. M., & Glazer, A. D.

    (2008). Examining the agreement of direct behavior ratings and systematic direct

    observation data for on-task and disruptive behavior. *Journal of Positive Behavior*

    Interventions, *10*, 136-143.

Riley-Tillman, T. C., & Johnson, A. H. (2017). Current advances and future directions in

    behavior assessment. *Assessment for Effective Intervention*, *42*, 77–80.

Sims, W. A., Riley-Tillman, C., & Cohen, D. R. (2017). Formative assessment using Direct

    Behavior Ratings: Evaluating intervention effects of Daily Behavior Report Cards.

    Assessment for Effective Intervention, *43*, 6–20.

Stichter, J. P., & Conroy, M. A. (2005). Using structural analysis in natural settings: A responsive functional assessment strategy. *Journal of Behavioral Education*, *14*(1), 19-34.

Talbott, E., Karabatsos, G., & Zurheide, J. L. (2018). Informant similarities, twin studies, and the assessment of externalizing behavior: A meta-analysis. *Journal of School Psychology*, *67*, 31-55.

Talbott, E., Zurheide, J. L., Karabatsos, G., & Kumm, S. (2021). Teacher similarities in the assessment of externalizing behavior among twins: A meta-analysis. *Behavioral Disorders*, *46*, 80-95.

Weisz, J. R., Chorpita, B. F., Frye, A., Ng, M. Y., Lau, N., Bearman, S. K., Ugeto, A. M., Langer, D. A., & Hoagwood, K. E. (2011). Youth top problems: Using idiographic, consumer-guided assessment to identify treatment needs and to track change during psychotherapy. *Journal of Consulting and Clinical Psychology*, *79*, 369–380.

What Works Clearinghouse, Institute of Education Sciences, U.S. Department of Education (2014). *What Works Clearinghouse: Procedures and standards handbook version 3.0*. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v3_0_standards_ handbook.pdf

Table 1

School Contexts, Perspectives, and Assessment Methods for Behavioral Decision Making

	SCHOOL CONTEXTS[a]		
	Structured Class 1	**Structured Class 2**	**Less Structured[b]**
PERSPECTIVE			
Self	Child/adolescent	Child/adolescent	Child/adolescent
Teacher	Teacher 1	Teacher 2	Teacher, school leader, paraprofessional, hall monitor, cafeteria worker, coach, bus driver
Systematic Direct Observation	Teacher, paraprofessional, school mental health provider	Teacher, paraprofessional, school mental health provider	Teacher, paraprofessional, school mental health provider
Academic & Behavioral Data	Curriculum-based measurement (CBM); Grade point averages (GPA); standardized test scores (state & national); office disciplinary referrals (ODR)	CBM; ODR; GPA; test scores	ODR

[a] *School contexts* include setting events (physical arrangements, degree of structure & corresponding behavioral expectations), human interactions (with adults and peers), and intervention effects. Contingencies for behavior occur within these contexts and interactions. Different types of behavior (externalizing and internalizing) are observed and rated by informants within these contexts.

[b]Less structured school settings are non-academic. Examples include class down time/free time, transitions between classes and activities (e.g., hallway interactions), lunch, recess, club activities, sports, etc.

Figure 1
The Operations Triad Model (De Los Reyes et al., 2013)