RUNNING HEAD: INFORMANT DISCREPANCIES IN SCHOOL-BASED RESEARCH

Informant Discrepancies in Assessments of Psychosocial Functioning in School-Based Services and Research: Review and Directions for Future Research

Andres De Los Reyes[1] Clayton R. Cook[2] Frank M. Gresham[3]
Bridget A. Makol[1] Mo Wang[4]

[1]University of Maryland at College Park [3]Louisiana State University
[2]University of Minnesota [4]University of Florida

CITATION

De Los Reyes, A., Cook, C.R., Gresham, F.M., Makol, B.A., & Wang, M. (2019, June). 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. doi: 10.1016/j.jsp.2019.05.005

Author Notes

Andres De Los Reyes and Bridget A. Makol, Comprehensive Assessment and Intervention Program, Department of Psychology, University of Maryland, College Park, MD. Clayton Cook, University of Minnesota, Minneapolis, MN. Frank M. Gresham, Louisiana State University, Department of Psychology, Baton Rouge, LA. Mo Wang, Department of Management, University of Florida, Gainesville, FL.

The authors' work on this paper was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A180032 to University of Maryland at College Park. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Correspondence regarding this manuscript should be addressed to Andres De Los Reyes, Comprehensive Assessment and Intervention Program, Department of Psychology, University of Maryland, Biology/Psychology Building, Room 3123H, College Park, MD 20742; Office: 301-405-7049; Fax: 301-314-9566; E-mail: adlr@umd.edu; Twitter: @JCCAP_Editor

Competing Interests Statement

The authors have no competing interests to declare

Abstract

Psychosocial functioning plays a key role in students' wellbeing and performance inside and outside of school. As such, techniques designed to measure and improve psychosocial functioning factor prominently in school-based service delivery and research. Given that the different contexts (e.g., school, home, community) in which students exist vary in the degree to which they influence psychosocial functioning, educators and researchers often rely on multiple informants to characterize intervention targets, monitor intervention progress, and inform the selection of evidence-based services. These informants include teachers, students, and parents. Across research teams, domains, and measurement methodologies, researchers commonly observe discrepancies among informants' reports. We review theory and research—occurring largely outside of school-based service delivery and research—that demonstrates how patterns of informant discrepancies represent meaningful differences that can inform our understanding of psychosocial functioning. In turn, we advance a research agenda to improve use and interpretation of informant discrepancies in school-based services and research.

Keywords: informant discrepancies; multiple informants; Operations Triad Model; schools

You are a school psychologist who is conducting a comprehensive evaluation for a student suspected of experiencing educational impairment due to an emotional disturbance. If found eligible, information gathered in the evaluation will be integral for developing educational programming as part of an individual education plan (IEP) for the student. Within the evaluation, you gather information via behavior rating scales administered to multiple adult authority figures in the student's life (one parent and their math and history teachers). You observe that ratings between the teachers and the parent do not agree with one another. In this scenario, are the teachers' reports "right" and the parent's report "wrong"? Alternatively, is the parent's report "right" and the teachers' reports "wrong"? Is the student eligible for special education services if there is disagreement in observations of the student's behavior between key authority figures in the student's life? If found eligible, what should the behavioral goals and programming entail based on this discrepant information? The truth is: most school psychologists and educators do not have the data available to facilitate answering these questions.

The above case illustration involves use of ratings from multiple informants to characterize a student's psychosocial functioning to guide educational programming. This is a common assessment strategy in school-based services, especially for students with disabilities. When using this strategy, discrepancies commonly arise among informants' ratings. Yet, our current technologies for assessing students' psychosocial functioning do not allow us to definitively interpret the meaning (or lack thereof) of these *informant discrepancies*. Conventional wisdom suggests that informant discrepancies represent measurement error or rater biases, rendering one or more of the informants' ratings invalid (De Los Reyes et al., 2015). In fact, common approaches for dealing with these discrepancies include analytic techniques that: (a) focus on shared variance in latent space (i.e., emphasis on the degree to which ratings

between informants overlap), (b) aggregate data from multiple informants into a single measured score, or (c) avoid multivariate data entirely in favor of a single "primary measure" completed by one, "optimal" informant (see De Los Reyes, Kundey, & Wang, 2011).

The uncertainty about how to interpret informant discrepancies represents a key barrier to educational diagnoses and identifying more precise and effective evidence-based programming tailored to student needs and the primary contexts in which they exist. Yet, what if these discrepancies ceased serving as barriers to interpreting assessment outcomes, and started serving as interpretive tools? In this paper, we discuss the impact of informant discrepancies on school-based services and research, and new areas of research and theory on how to improve their interpretability. We then advance a four-phase research agenda for studying how informant discrepancies might *inform* characterizing students' psychosocial functioning and facilitate identification of more precise and effective evidence-based services.

Conceptual Foundations of Assessments of Psychosocial Functioning

Educational assessments often consist of an array of modalities (e.g., surveys, interviews, performance-based tasks) that tap into myriad content domains, such as intellectual functioning, impulsivity, reading level, and academic achievement (Merrell, 2008). Among these content domains, psychosocial functioning comprises a set of important determinants of outcomes across a range of disciplines (American Psychiatric Association, 2013; National Institute of Mental Health [NIMH], 2015). The World Health Organization (2018) defines *psychosocial health* as "a state of complete physical, mental, and social well-being, and not merely the absence of disease and infirmity." In our paper, we conceptualize psychosocial functioning as comprising domains that connote adaptive functioning (i.e., promotive or protective factors that enhance positive outcomes) as well as domains that connote challenges to adaptive functioning (i.e., problem or

risk factors that minimize positive outcomes). Along these lines, we review research on informant discrepancies in assessments across multiple domains of psychosocial functioning.

Psychosocial functioning plays a key role in students' wellbeing and performance inside and outside of school (e.g., Reynolds, Livingston, & Wilson, 2006). The different contexts in which students exist (e.g., school and home) also serve to influence and modify psychosocial functioning. As such, a key component of evidence-based assessments in school-based services and research involves reliably and validly assessing domains of psychosocial functioning (Cook, Volpe, & Delport, 2014). The ways in which students function psychosocially are not constrained by the start and stop of their school day. Social contexts *inside* and *outside* of the school system influence students' psychosocial functioning (e.g., school, home, peer interactions; Cicchetti, 1984; Luthar et al., 2000). Further, any one student may display behaviors indicative of psychosocial functioning in some contexts, such as the classroom or peer interactions, to a greater degree than other contexts, such as the home or community (Dirks, De Los Reyes, Briggs-Gowan, Cella, & Wakschlag, 2012). In fact, contextual variations in displays of psychosocial functioning occur within a variety of domains relevant to academic performance, such as attention and hyperactivity, conduct problems, social anxiety, and social competence (e.g., De Los Reyes, Henry, Tolan, & Wakschlag, 2009; Deros et al., 2018; Dirks et al., 2012). For example, social contexts may differ in the degree to which they elicit displays of oppositional behaviors (e.g., noncompliance, defiance) due to differences across contexts in social structure, instructional demands, and emotional support (Dretzke et al., 2009). Thus, identifying the contexts in which children would benefit from improvements in their psychosocial functioning may facilitate more precise intervention programming and boost programming efficacy (e.g., NIMH, 2015).

If a student varies in levels of psychosocial functioning across social contexts, it logically follows that no one person in that student's life has the capacity to observe and accurately rate their functioning across all contexts. In fact, collecting reports from multiple informants about students' psychosocial functioning is a common practice that happens in nearly every public school system (Reynolds et al., 2006; Salvia, Ysseldyke, & Bolt, 2012). In school-based services and research, data from multiple informants' reports inform several areas of decision-making regarding educational programming, including special education eligibility, goals included in students' IEPs, selection and delivery of supports, monitoring of student response to intervention, and placement decisions (Merrell, 2008). Indeed, federal law requires "Protection in Evaluation" procedures and defines a comprehensive evaluation as one that includes multiple methods derived from multiple informants in order to identify a student as eligible for special education services (Individuals with Disabilities Education Act, 2004). By gathering reports from different informants who vary in the specific contexts in which they observe and interact with students, school professionals may be able to gain a better understanding as to how students' psychosocial functioning varies within and across contexts (see also Hunsley & Mash, 2007). In school-based services and research, the multi-informant approach typically involves soliciting reports from informants who presumably observe students' behaviors as displayed within the school system (e.g., teachers), outside of the school system (e.g., parents), and across systems (e.g., students; see Kraemer et al., 2003).

To provide data about students, informants may complete standardized instruments that ask them to rate how often a behavior has occurred over a specific period. The most common of these broadband instruments include behavior checklists such as the Achenbach System of Empirically Based Assessments (Achenbach & Rescorla, 2001), Behavioral Assessment System

for Children (Reynolds & Kamphaus, 2004), and Social Skills Improvement System-Rating Scales (Elliott & Gresham, 2008). Moreover, there are several narrowband instruments that educators and researchers regularly administer to gather reports from teachers and parents, including the Connors Rating Scales to assess hyperactivity, inattention, and disruptive behavior (Connors, 1997), Vineland Adaptive Behavior Scales (VABS) to assess adaptive behaviors (Sparrow, Balla, & Cicchetti, 1984), and Childhood Autism Rating Scale (CARS) to assess symptoms of autism (Schopler, Reichler, & Renner, 1988). These broadband and narrowband instruments are widely used because of their ease of administration and relative low cost compared to other more intensive assessment methods. Behavior rating scales, however, are indirect assessment methods, in that they do not yield indices of behavior at the time and place of their actual occurrence. Indeed, as we describe below, a key element of emerging research and theory on multi-informant assessments involves understanding links between informants' reports and behavioral data collected on independent assessments (e.g., naturalistic observations and official records; De Los Reyes, Thomas, Goodman, & Kundey, 2013).

In both service and research settings, collecting multiple informants' reports generates a great deal of information about students' psychosocial functioning. However, when compared, the individual reports from separate informants often yield inconsistent conclusions (De Los Reyes, 2011). These informant discrepancies are some of the most robust findings in the social sciences (Achenbach, 2006). Over the last 30 years, several large-scale epidemiological studies and meta-analytic reviews demonstrate that informant discrepancies occur across development (i.e., childhood, adolescence, and adulthood; Achenbach, Krukowski, Dumenci, & Ivanova, 2005; Achenbach, McConaughy, & Howell, 1987; Duhig, Renk, Epstein, & Phares, 2000), cultures (De Los Reyes, Lerner, et al., 2019; Rescorla et al., 2014, 2017), psychosocial domains

(e.g., autism, child maltreatment, parenting, suicide, social competence; Gresham et al., 2018; Jones et al., 2019; Korelitz & Garber, 2016; Renk & Phares, 2004; Romano, Weegar, Babchishin, & Saini, 2018; Stratis & Lecavalier, 2015), and measurement modalities (i.e., dimensional and categorical measures; De Los Reyes et al., 2015). In fact, these informant discrepancies are quite consistent across samples and studies. For instance, in a meta-analytic review of 341 studies on cross-informant correspondence in mental health assessments (De Los Reyes et al., 2015), the 95% confidence interval of correspondence levels using the *r* metric [.22, .33] includes magnitudes for which both the upper and lower bounds would have been classified by Cohen (1988) as falling in the "moderate" range. To be clear, this is not to say that variability in cross-informant correspondence levels does not exist. Indeed, correspondence levels among informants who observe behavior in the same context (e.g., pairs of teachers, pairs of parents) hovers in the .40s-.60s range (Achenbach et al., 1987; De Los Reyes et al., 2015). Yet, even these magnitudes of correspondence do not rise to a level that, in a given sample, any two informant's reports become redundant with one another or yield the same conclusions (see also Achenbach, 2006). In sum, scores of research conducted over the last few decades speak to how robustly informant discrepancies manifest in assessments of numerous domains of psychosocial functioning.

Given their common occurrence, it should come as no surprise that informant discrepancies have historically created uncertainty when interpreting assessment outcomes. For instance, these discrepancies often make it difficult to determine whether an individual meets diagnostic criteria for externalizing or internalizing disorders. In an outpatient sample of children receiving services for externalizing concerns, prevalence estimates of DSM-IV subtypes (Primarily Inattentive, Primarily Hyperactive Impulsive, or Combined Type) of Attention

Deficit/Hyperactivity Disorder (ADHD) varied widely from zero children in the sample to the grand majority of children in the sample, depending on how one combined parent and teacher reports (i.e., "and" vs. "or" rule) and the instrument used to gather parent and teacher reports (i.e., interview vs. rating scale; see Table 5 of Valo & Tannock, 2010). Informant discrepancies also impact judgments about the efficacy of interventions in that intervention effects may range from "small" to "large" depending on the informant (for reviews, see De Los Reyes & Kazdin, 2006, 2009). Moreover, researchers observe informant discrepancies when assessing severity of symptoms of students with autism spectrum disorders, leading some to conclude that their research "support[s] the role of environmental context in psychiatric symptom expression in children affected by autism and suggest[s] that informant discrepancies may provide critical cues for these children via specific environmental modifications" (Kanne, Abbacchi, & Constantino, 2009). Clearly, these discrepancies greatly influence intervention programming and interpretation of intervention outcomes and may lead to differential conclusions regarding the efficacy of interventions. Importantly, informants within (e.g., across teachers) versus outside (e.g., parents) of the school system often vary as to whether they perceive school interventions as yielding beneficial effects in children's behavior, and these informants also comprise key stakeholders in the administration and outcomes of school-based interventions. Thus, failing to properly attend to informant discrepancies can have profound implications for school-based services and research.

Despite the historic difficulty in understanding and interpreting informant discrepancies, their robust presence in school-based services and research makes intuitive sense. As mentioned previously, students' social contexts (e.g., classrooms, home environment) may vary considerably in eliciting the behaviors that impact school performance. Given these variations,

informants embedded in different social contexts (e.g., parents vs. teachers) should also differ in how they observe and thus rate students' behavior. The intuitive nature of informant discrepancies and their ubiquity in school-based services and research makes it even more surprising that in practice, we lack consensus guidelines on how to interpret informant discrepancies (see also Beidas et al., 2015). The lack of guidelines for this crucial element of service provision likely stems from long-held traditions in measurement construction (e.g., pursuit for agreement across raters) and interpretation of multivariate data (for a review, see De Los Reyes et al., 2015). That being said, an emerging body of research—occurring in disciplines largely outside of school-based service delivery and research—reveals new insights into these informant discrepancies. Specifically, the degree of discrepancy among informants' reports about behavior yields important information about that very behavior. In fact, informant discrepancies may reflect individual differences in displays of behavior based on the context(s) in which they occur and the psychosocial phenomena they reflect. Theoretical models exist to explain these discrepancies (De Los Reyes, Thomas et al., 2013), and empirical paradigms offer ways to model or test these discrepancies (e.g., Lerner, De Los Reyes, Drabick, Gerber, & Gadow, 2017). Below we review recent advancements in theory on informant discrepancies, as well as an emerging evidence base that supports this theoretical work. We then lay the groundwork for a four-phase research agenda for which the central goal is to improve the interpretability of informant discrepancies in assessments of psychosocial functioning in school-based services and research.

Recent Theoretical Work on Informant Discrepancies: Operations Triad Model

Many challenges arising from use of multi-informant assessments stem from the historic lack of conceptual models for interpreting informant discrepancies (for a review, see De Los

Reyes, Thomas, et al., 2013). Specifically, Converging Operations is the dominant conceptual model for interpreting multivariate data (Garner, Hake, & Eriksen, 1956). Converging Operations holds that one interprets the veracity of a study's findings—and by extension, multiple data points collected for an individual assessment—based on the extent to which findings from multiple methodologically distinct data sources (e.g., different informants' reports) converge and point to the same conclusion. In fact, data triangulation through obtaining reports from multiple informants is a key component of "best practices" advocated in school psychology assessment textbooks (e.g., Sattler & Hoge, 2006). As a framework, Converging Operations drives its users toward removing or otherwise discounting source-specific variance (e.g., discounting a parent's report based on the judgment of the school psychologist; see also De Los Reyes et al., 2015). This can be seen in empirical approaches described previously that focus on shared variance and eschew, dismiss, or otherwise nullify use of unshared or informant-specific variance (De Los Reyes, Kundey, & Wang, 2011).

Importantly, these approaches informed by Converging Operations lie in stark contrast to the very reasons why one collects multi-informant data. Indeed, in school-based services and research, one takes a multi-informant approach to assessment based on the ideas that (a) students behave differently, depending on the social context or environmental demands; and (b) informants often vary in where they observe the students about whom they provide behavioral reports (Achenbach et al., 1987). Classrooms, for example, are performance settings in which students are expected to arrive on time, be prepared with materials, and engage in academic instruction and activities. In contrast, at home students have access to their belongings and often spend a great deal of time at rest or engaging in leisure and recreational activities. These assessment conditions thus beg the question: Why would one expect multiple informants' reports

of the same student's behavior to yield the same conclusion? Why would researchers place exclusive focus on estimating common variance? If multiple informants' reports potentially yield contextually sensitive data from discrepancies among these reports, then focusing exclusively on estimating common variance may result in (a) loss of valuable information about the behaviors targeted for assessment, (b) erroneous eligibility decisions, and thus (c) less effective programming.

One organizing framework for using and interpreting multi-informant assessments is the Operations Triad Model (De Los Reyes, Thomas, et al., 2013). We graphically depict this framework in Figure 1. The Operations Triad Model expands upon the Converging Operations concept (Figure 1a) by delineating conditions for identifying patterns of multi-informant assessment outcomes that reflect Converging Operations, as well as two alternative concepts. First, the Operations Triad Model specifies conditions for Diverging Operations (Figure 1b), a set of measurement conditions by which multiple informants' reports yield discrepant findings, and the discrepancies reflect meaningful variation in the behaviors being assessed. An example of a Diverging Operations scenario might involve a teacher reporting hyperactivity and noncompliance concerns in a student that go uncorroborated by a parent's report, with the discrepancies occurring because the student primarily displays hyperactivity and noncompliance when interacting with peers in the context of classroom activities (i.e., few concerns displayed at home). Second, the Operations Triad Model specifies conditions for Compensating Operations (Figure 1c), a set of measurement conditions by which multiple informants' reports yield discrepant findings, and the discrepancies reflect mundane methodological features of the assessment procedures. For instance, Compensating Operations might manifest if a teacher and student provide reports using measures of a psychosocial domain like social competence that

differ in item content and scaling. Consequently, the two reports may have diverged because their measures differed in terms of their psychometric properties. The ability of the Operations Triad Model to distinguish informant discrepancies that reflect meaningful versus mundane reasons (i.e., Diverging Operations vs. Compensating Operations) is a key element of the framework. Indeed, this distinction provides the conceptual foundation to guide direct empirical tests of informant discrepancies, as well as tests of the assumptions that often underlie use of commonly leveraged multivariate techniques (e.g., structural equation modeling and use of composite scores to aggregate data).

More broadly, embedded in the Operations Triad Model are a set of measurable, testable conditions for distinguishing assessment outcomes that reflect Converging Operations, Diverging Operations, or Compensating Operations, which we present in Figure 2. Specifically, the Operations Triad Model facilitates the process by which one poses a priori hypotheses as to whether they expect to observe converging findings or diverging findings among a set of multiple informants' reports (Figure 2a). Empirical questions outlined in the figure can then guide tests of these expectations. For instance, these questions can be used to determine if the evidence supports the a priori expectation of converging findings (i.e., Converging Operations; Figure 2b). Questions can also be posed to determine whether the evidence supports the a priori expectation of diverging findings as yielding meaningful information about behavior (i.e., Diverging Operations; Figure 2c). If the evidence fails to support either of these hypotheses, one can proceed to testing whether the observations are best explained by measurement error (i.e., Compensating Operations; Figure 2d).

In sum, the Operations Triad Model provides researchers with an evidence-based approach for using and interpreting multi-informant assessments. Using this framework,

researchers can test whether their multi-informant approach yields data that converge on a common outcome or diverge toward different outcomes for meaningful or mundane reasons (cf. Figures 1b and 1c). Further, the Operations Triad Model has heuristic value. As evidence of this, consider the recent development of modified versions of the Operations Triad Model for interpreting (a) physiological data in mental health assessments (De Los Reyes & Aldao, 2015) and (b) developmental assessments of family functioning (De Los Reyes & Ohannessian, 2016; De Los Reyes, Ohannessian, & Racz, 2019).

Empirical Work Supporting the Operations Triad Model

An emerging body of empirical work across multiple disciplines supports central tenets of the Operations Triad Model. The first tenet is that, if informant discrepancies yield systematic information about psychosocial domains, then magnitudes or levels of these discrepancies should co-vary with basic characteristics of both the constructs/domains assessed by informants' reports as well as the social contexts (i.e., where informants observe behavior). For example, basic psychometric work indicates that when collecting multiple raters' reports of observed behavior, one should observe the greatest inter-rater agreement for behaviors that are relatively easier to observe (e.g., aggression vs. low mood; Groth-Marnat & Wright, 2016; Hunsley & Lee, 2010). Consistent with this expectation, two large-scale meta-analyses that collectively reviewed 50 years of research on cross-informant correspondence in reports of youth mental health observed greater correspondence levels for reports about externalizing difficulties relative to internalizing difficulties (Achenbach et al., 1987; De Los Reyes et al., 2015). Another expectation is that informants rating behaviors based on observations from the same context should correspond in their reports to a greater degree than informants who base their ratings on observations from different contexts (Kraemer et al., 2003). In fact, the two meta-analyses described previously

also revealed greater correspondence between reports by informants from the same context (e.g., pairs of parents; pairs of teachers) relative to reports by informants from different contexts (e.g., mother and teacher; Achenbach et al., 1987; De Los Reyes et al., 2015).

A second tenet is that, for informant discrepancies to yield meaningful information about psychosocial domains, levels of these discrepancies need to *vary* among informant pairs. That is, low correspondence characterizes multiple informants' reports in general. Nevertheless, in addition to pairs who provide very different reports, there should also exist some informant pairs who provide quite similar reports. To test these ideas, researchers often leverage person-centered models of data analysis (e.g., latent class analysis; McCutcheon, 1987) to examine whether samples with multiple informants' reports contain subgroups of informant pairs who vary in the levels of discrepancies between their reports. In fact, among many informant pairs (e.g., parent-child, teacher-parent, mother-father) and domains (e.g., internalizing and externalizing difficulties, parenting and parental monitoring, family conflict, social competence), prior work identifies at least three kinds of reporting patterns: (a) converge on reports of high levels of psychosocial functioning; (b) diverge such that one informant reports lower levels of psychosocial functioning, relative to the other informant; or (c) converge on reports of low levels of psychosocial functioning (e.g., De Los Reyes et al., 2015). For example, in two studies about assessments of children's conduct problems, researchers identified four groups of teacher-parent dyads: (a) both teacher and parent agree on low conduct problems, (b) teacher (but not parent) reports high conduct problems, (c) parent (but not teacher) reports high conduct problems, and (d) both teacher and parent agree on high conduct problems (Fergusson, Boden, & Horwood, 2009; Sulik et al., 2017).

A third tenet is that, not only should there exist individual differences in levels of

informant discrepancies, but that these individual differences should convey meaningful information about the psychosocial domains for which informants provide reports. Recent work reveals how informant discrepancies convey these kinds of meaningful information. Specifically, informant discrepancies may inform how one characterizes assessed behaviors. That is, beyond the information provided by individual informants' reports of psychosocial functioning, understanding patterns of agreement and disagreement across informants' reports reveal important aspects of such functioning that cannot be obtained by simply interpreting reports individually or additively. For example, in assessments of disruptive behavior in early childhood, when parents endorse disruptive behavior that teachers do not endorse, these discrepancies tend to point to children displaying observed disruptive behavior with parents but not non-parental adults on independent, contextually sensitive laboratory tasks (De Los Reyes et al., 2009). In this study, teacher-endorsed disruptive behavior that went uncorroborated by parent report pointed to children displaying observed disruptive behavior with non-parental adults and not parents, and children for whom both teacher and parent endorsed disruptive behavior tended to display observed disruptive behavior across interactions with parents and non-parental adults. In short, this study found that informant discrepancies provided information about contextual variations in the assessed behavior. That is, informant discrepancies may not reflect measurement error but rather meaningful differences in displays of the child's behavior across social environments. Similar findings manifest between (a) parent and teacher reports in assessments of aggressive behavior in relatively older children (Hartley, Zakriski, & Wright, 2011); (b) parent and teacher reports of childhood autism spectrum symptoms (Lerner et al., 2017); (c) parent and adolescent reports in assessments of adolescent social anxiety (Deros et al., 2018; Glenn et al., 2019); (d) parent and youth reports in assessments of youth depression collected in school settings (Makol

& Polo, 2018); (e) parent and adolescent reports of adolescent psychosocial functioning in the context of screening assessments for pediatric asthma (Al Ghriwati, Winter, Greenlee, & Thompson, 2018); (f) parent and adolescent reports of adolescent internalizing concerns completed at admission to an inpatient unit (Makol, De Los Reyes, Ostrander, & Reynolds, 2019); and (g) cross-parental reports in non-clinic, community assessments of adolescent psychosocial functioning (De Los Reyes, Alfano, Lau, Augenstein, & Borelli, 2016).

Research Agenda for Informant Discrepancies in School-Based Services and Research

The body of research and theory reviewed previously supports the idea that informant discrepancies in reports of students' psychosocial functioning may contain meaningful information about this functioning. Yet, a hallmark principle of evidence-based assessment is that one should not assume that the psychometric properties of assessments generalize to all assessment occasions or service populations (Groth-Marnat & Wright, 2016; Hunsley & Mash, 2007). In line with this, we propose a four-phase agenda for improving the interpretability of informant discrepancies in assessments of students' psychosocial functioning in school-based services and research. This four-phase agenda reflects the translational research spectrum (see Fishbein, Ridenour, Stahl, & Sussman, 2016) beginning with more basic science and culminating with implementation science to incorporate findings into routine practice. In Figure 3 we summarize each phase in the agenda. Below and for each phase in the agenda, we either cite prior work that provides examples of studies that meet the aims and scope of research conducted in the phase, or explain how one might conduct research in a phase within school-based populations.

Phase 1: Generalizability of the Operations Triad Model to School-Based Assessments

Phase 1 of the research agenda involves testing whether the Operations Triad Model

generalizes to school-based assessments of students' psychosocial functioning. That is, low-to-moderate levels of cross-informant correspondence manifest in school-based assessments of a host of psychosocial domains (e.g., Lee, Elliott, & Barbour, 1994). Further, for several measures of psychosocial functioning there exists documented evidence of the presence of informant discrepancies when these measures are used in school-based assessments. Specifically, Table 1 includes a list of several of these instruments. To identify these instruments, we reviewed the 341 articles examined in a recent meta-analysis of child and adolescent mental health (De Los Reyes et al., 2015). The measures reported in Table 1 were those used in studies tested in the meta-analysis that (a) collected multi-informant data on the measure, (b) estimated levels of cross-informant correspondence in reports on that measure, and (c) did so using a school-based sample.

However, as mentioned previously research informed by the Operations Triad Model finds that underlying low levels of cross-informant correspondence there exist considerable between-dyad variations or subgroups of multi-informant reports in terms of the level (e.g., convergence vs. divergence) and direction (e.g., teacher > parent and parent > teacher) of informant discrepancies (see also De Los Reyes et al., 2019). Similarly, do subgroups or patterns of multi-informant reports described previously also manifest when assessing psychosocial functioning in school-based settings? For example, in large samples of school-based assessments of students' specific externalizing behaviors (e.g., aggression, noncompliance, hyperactivity), can one reliably identify "pockets" of teacher-parent dyads in these samples who converge in reports of high levels of externalizing concerns; diverge such that teachers report higher concerns relative to parents, and vice versa; or converge in reports of low levels of externalizing concerns? If school-based assessments yield between-dyad differences in patterns of multi-informant reports, the key question would be: Do these differences meaningfully relate to variations in

psychosocial phenomena relevant to school-based services and research?

Several recent studies cited previously from research groups in clinical psychology, developmental psychology, and psychiatry reveal strategies for addressing these questions (Al Ghriwati et al., 2018; Lerner et al., 2017; Makol & Polo, 2018). Each of these studies had one thing in common: They tested links between patterns of multi-informant reports and independent measures of psychosocial phenomena relevant to the instruments on which informants provided reports. By "independent measures," we mean data derived from modalities and/or information sources that do not overlap with the informants completing ratings about students' psychosocial functioning. In this way, one can leverage these independent instruments as criterion measures to test in relation to patterns of multi-informant reports, without the confound of shared method variance between predictor and criterion variables (i.e., *criterion contamination*; see De Los Reyes et al., 2015). For instance, consider a study for which a key aim was to test whether discrepancies between parent and teacher survey reports of social competence reflected variations in the degree to which students displayed low competence in home and/or school contexts. If parents and teachers also completed the criterion measures (e.g., number of friends with whom the student interacts), then any links between patterns of multi-informant reports and the criterion measures could be parsimoniously explained by the same two informants providing data for both the predictor and criterion measures.

Needless to say, addressing issues of criterion contamination also raises issues of feasibility for testing the generalizability of the Operations Triad Model in school-based settings. Indeed, previous studies testing elements of the Operations Triad Model often used scores taken from performance-based tasks or independent observers' ratings of behavior on laboratory-controlled social interaction tasks (e.g., De Los Reyes et al., 2009; De Los Reyes, Alfano, Lau,

Augenstein, & Borelli, 2016; De Los Reyes, Bunnell, & Beidel, 2013; Glenn et al., 2019; Lerner et al., 2017). These laboratory tasks might be difficult to implement in school-based settings. Importantly, tests of the Operations Triad Model are not limited to use of these laboratory-based criterion measures. The key issues regarding research on the Operations Triad Model are two-fold. The first is minimizing or avoiding the criterion contamination issues described previously. The second is identifying criterion measures that reflect variations in psychosocial phenomena relevant to the constructs about which informants provide reports. Regardless of the criterion measures one implements, the purpose is to test whether patterns of multi-informant reports "match" variations in behavior observed on the criterion measures.

An example here may be helpful. Consider a study focused on testing the meaning of informant discrepancies in school-based assessments of students' externalizing behaviors. For those students for whom parents report externalizing concerns that teachers' reports do not corroborate, are those students also likely to experience associated features of externalizing concerns at home (e.g., maladaptive parenting; Hunsley & Lee, 2010)? Further, are those students, at the same time, experiencing protective factors at school (e.g., teacher displays proper classroom management skills) or the absence of associated features of externalizing concerns at school (e.g., rejection by peers during classroom activities or social isolation)? For those students for whom teachers report externalizing concerns that parents' reports do not corroborate, are they also likely to experience associated features of externalizing concerns at school to a greater degree than at home? Do students for whom both teachers *and* parents report externalizing concerns experience associated features of such concerns across home and school contexts? The answers to these questions allow one to discern whether patterns of multi-informant reports of students' externalizing behaviors contain meaningful information. In this way, Phase 1 addresses

whether informant discrepancies have the *potential* to inform school-based assessments of students' psychosocial functioning for the purposes of eligibility determination and the selection and delivery of evidence-based programming.

Phase 2: Strategies for Optimizing Multi-Informant Assessments to Yield Interpretable Data about Informant Discrepancies in School-Based Services and Research

In many respects, Phase 1 of this proposed research agenda focuses on the basic science of informant discrepancies in school-based settings. That is, the work reviewed previously, conducted largely outside school-based service and research settings, reveals both the presence of informant discrepancies and the prospect that they yield key indices reflecting psychosocial functioning. However, all of this work involved leveraging instruments that were not designed to sensitively assess informant discrepancies. In fact, researchers developed current multi-informant instruments as one would from a Converging Operations perspective: A parallel format with the same items, responses options, and item content largely held constant across informants.

In prior work, use of parallel formats across informants' reports greatly facilitated study of informant discrepancies. Indeed, parallel formats allow researchers to parsimoniously rule out changes in instrumentation as an explanation for discrepancies observed between informants' reports (De Los Reyes, Thomas, et al., 2013). However, the foundation for use of parallel construction of multi-informant measures largely rests on the idea that the constructs about which informants provide reports manifest in much the same way across informants—and by extension—across contexts; tests of this foundational assumption are often referred to as tests of *measurement invariance* (Dirks et al., 2014; Olino, Finsaas, Dougherty, & Klein, 2018; Russell, Graham, Neill, & Weems, 2016). Yet, decades of research on informant discrepancies prove otherwise: Several meta-analyses published over the last 30 years reveal that informants often

rate behavior in fundamentally different ways across social contexts (e.g., Achenbach et al., 1987; De Los Reyes et al., 2015; Duhig et al., 2000). Further, *exclusively* taking a parallel format approach to multi-informant assessment translates into developing measures designed to minimize informant discrepancies, by deleting items upon which informants systematically differ in item responses. It is no wonder why anyone using these measures would often find themselves confronted with perhaps a misguided task: Determining which informant is providing the "valid report" and which ones are providing "biased or skewed reports" (see also De Los Reyes, Youngstrom, et al., 2011).

The findings of prior work, as robust as they are, may be *grossly underestimating* the impact of informant discrepancies on services and research geared toward understanding and intervening to improve psychosocial functioning. Thus, for informant discrepancies research conducted at the sample or group level to translate to routine assessments conducted in school-based settings, we require a second phase of work (i.e., Phase 2) to focus on measurement development and/or refinement of existing measures of students' psychosocial functioning.

In Phase 2 of our agenda, we call for a significant *paradigm shift* in how we conceptualize and approach measurement development and testing of multi-informant reports in school-based services and research. Multi-informant assessments may, in theory, allow school professionals to identify the specific contexts in which students display behaviors indicative of psychosocial functioning. Yet, in practice existing multi-informant instruments may be ill-suited to sensitively identify, on an item-specific basis, behaviors that an individual student might display specifically at home, at school, or consistently across contexts. We require this item-level certainty in measurement in order to ensure that informant discrepancies can meaningfully inform service delivery and research.

We propose that school-based service and research settings require measures of psychosocial functioning that contain items that vary as to their context specificity. That is, measures that contain two kinds of items. The first, *discrepant items*, include those items upon which informants are unlikely to agree, and are more likely to reflect *context-specific* behaviors. An example of such an item might be "avoids speaking in groups," in which the behaviors described may systematically manifest to a greater degree in school, a setting in which school-based observers (e.g., teachers or school counselors) have more opportunities to observe students behave in social contexts relevant to these behaviors (e.g., group activities during class). The second, *nondiscrepant items*, include those items upon which informants are likely to agree, and may reflect *cross-contextual* behaviors. An example of such an item might be "helps others when asked." Here, the behaviors described reflect a construct (e.g., prosocial behavior) that, when displayed, may systematically manifest to such a degree that multiple informants, regardless of social context, would have sufficient observations or "samples" of behavior about which to provide reports (e.g., helping fellow students at school, helping siblings at home or relatives at family gatherings). Equipped with these items, service providers and researchers could construct scales that assess behaviors that are bound to a given context (i.e., discrepant), or behaviors that are contextually unbound and thus displayed across contexts (i.e., nondiscrepant).

Importantly, the approach we advance here and describe in further detail below seeks to unify the various traditions for integrating multi-informant data described previously. Specifically, by calling for measures that include scales with nondiscrepant items, we embrace the importance of tests of measurement variance, which have the key goal of identifying and retaining data about behaviors that manifest across informants' perspectives

(see Dirks et al., 2014, Olino et al., 2018; Russell et al., 2016). At the same time, the idea to create scales that include items for which informants provide differential responses (i.e., discrepant items) embraces the need to retain information germane to individual informants' unique perspectives (see also Becker-Haimes, Jensen-Doss, Birmaher, Kendall, & Ginsburg, 2018; Kraemer et al., 2003; Laird & De Los Reyes, 2013; Laird & Weems, 2011).

The combined, integrative approach we advance here would enable the development of specific interpretative guidelines for the purposes of diagnosis and evidence-based programming. In particular, data from discrepant scales could triangulate on characterizing and indexing changes in context-specific behaviors in response to school and/or home supports. In contrast, data from nondiscrepant scales might characterize and index changes in behaviors that cut across contexts or are not necessarily bound to any given social environment. Both of these scales could be used to pinpoint the selection and delivery of interventions, develop behavioral goals, and monitor intervention response in order to inform data-driven decisions. We see many possibilities for use of such scales in school-based services and research.

What techniques might we leverage to identify context-sensitive items of students' psychosocial functioning? The answer may lie in an innovative application of existing analytic techniques. Specifically, for decades researchers have leveraged indices of inter-rater agreement (IRA) and differential item functioning (DIF) to test whether scores obtained from different informants are interchangeable or equivalent in terms of their absolute value (IRA) or response probability (DIF; see Andrich & Hagquist, 2012; LeBreton, Burgess, Kaiser, & James, 2003). The typical application of these techniques involves identifying items that are subject to differential response. Predicated on the notion that differential response signals rater or statistical biases, measurement development has historically involved *pruning out* those items that respond

differentially across informants or raters. In stark contrast, we call for leveraging these techniques to: (a) *explicitly identify* items on which different informants make differential responses (i.e., discrepant), (b) *distinguish* those items from items on which informants provide similar responses (i.e., nondiscrepant), and (c) *include both kinds of items* on scales developed to measure psychosocial functioning.

How might one apply these techniques to identify discrepant and nondiscrepant items? To illustrate, we describe below how one might implement three different quantitative metrics to identify discrepant and nondiscrepant items: (a) inter-rater agreement r_{WG} index, (b) standardized mean difference (SMD), and (c) DIF. First, one of the most popular estimates of IRA is James, Demaree, and Wolf's (1984) single-item r_{WG}. We can depict the r_{WG} index, which defines agreement in terms of the proportional reduction in error variance, by the equation:

$$r_{WG} = 1 - \frac{S_X^2}{\sigma_E^2} \qquad (1)$$

where S^2_X is the observed variance on the variable X (e.g., ratings of noncompliance or positive affect) taken over K different informants or raters and σ^2_E is the variance expected when there is a complete lack of agreement among the informants. One can calculate the degree of agreement by comparing the observed variance to the variance expected when informants respond randomly. Using this formula, when all informants are in perfect agreement, the observed variance among judges is 0, and $r_{WG} = 1.0$. In contrast, when informants are in total lack of agreement, the observed variance will asymptotically approach the error variance, which leads r_{WG} to incrementally approach 0.0. Consistent with guidelines for acceptable IRA based on correlational metrics (LeBreton et al., 2003), one can identify an item as nondiscrepant with $r_{WG} > .70$. Conversely, the decision rule to identify discrepant items might be $r_{WG} < .70$.

Overall, r_{WG} is a tool for identifying discrepant and nondiscrepant items in terms of

absolute agreement. Thus, this index allows us to identify candidate discrepant and nondiscrepant items. However, by construction this index yields no information on the *direction* of any discrepant item identified. For this approach to yield useful data, we require the identification of discrepant items that largely index behaviors constrained to one context and not others (e.g., school but not home), and vice versa. This necessitates a second index that captures the direction of discrepant items. Specifically, one can identify discrepant and nondiscrepant items using the SMD effect size, using the following formula:

$$\frac{X_{r1} - X_{r2}}{S_{pooled}} \qquad (2)$$

where X_{r1} refers to the score for informant one (e.g., teacher) and X_{r2} refers to the value for informant two (e.g., parent) and S_{pooled} reflects the pooled standard deviation across informants' scores on the particular item. For the SMD, one can use Cohen's (1988) criteria as a guideline: small SMD = 0.20 to 0.49; moderate SMD = 0.50 to 0.79; large SMD > 0.80. Based on these guidelines, one can identify nondiscrepant items as those that have a SMD in the small range (-0.20 >SMD< +0.20), indicating low discrepancy. The decision rule to identify discrepant items would be a SMD in the large range (above +0.80 or below ─0.80). Discrepant items would be further decomposed by items for which informant one (e.g., teacher) had the larger score (SMD > +.80), and items for which informant two (e.g., parent) had the larger score (SMD < ─.80).

Using the first two formulas, one can see item identification begins with r_{WG} to identify candidate discrepant and nondiscrepant items, and proceeds with SMD to accomplish two goals: (a) confirm identification of nondiscrepant items observed using r_{WG} (i.e., items with a r_{WG} > .70 and -0.20 >SMD< +0.20); and (b) identify context-specific discrepant items (i.e., items with scores for informant 1 > informant 2, and vice versa). However, a key limitation with relying on just these two indices is the lack of a confirmatory index for identifying directional discrepant

items. Thus, we propose a third stage to the item identification process through the use of DIF analyses within item response theory models, which provide information about the properties of items via individual responses to those items. DIF analyses determine whether informants' ratings of the same latent trait (behavior) have a different probability of giving a certain response on behavior rating items (Andrich & Hagquist, 2012). As with the first two methods, the DIF method can be applied at the item-level, namely comparing item characteristics across rating sources for each of the items previously identified as discrepant and nondiscrepant. Based on these tests, one can implement a transformed log-odds ratio with the Mantel-Haenszel procedure to quantify the DIF effect (see Andrich & Hagquist 2012), and using odds ratio thresholds for nondiscrepant (.80 – 1.20) and discrepant (< 0.80 or >1.20) items. Items that comprise the final pool to include in discrepant and nondiscrepant scales would consist of those items that "pass" identification thresholds across all three IRA indices.

Phase 3: Testing the Ability of Multi-Informant Assessments to Sensitively Detect Contextual Variations in Outcomes of School-Based Services

After developing multi-informant measures of psychosocial functioning that contain discrepant and nondiscrepant items, one key question arises: Are scales containing these items capable of detecting context-specificity in behaviors displayed by an individual student? To address this question, Phase 3 leverages widely used techniques that have recently gained currency in psychological assessment (for a review, see Youngstrom, 2013). In particular, receiver operating characteristic (ROC) methods may facilitate the kinds of *personalized* assessment outcomes that the item-sensitive scales described previously ought to yield (De Los Reyes et al., 2015; NIMH, 2015). One leverages ROC methods for a variety of reasons. In particular, ROC methods allow a user to identify specific scores on an instrument for detecting

two different kinds of cases. First, one might use ROC methods to "rule in" or identify students who display a characteristic, such as students achieving positive response to an IEP, in an effort to optimize *sensitivity* of case identification. Second, ROC methods allow a user to identify specific scores on an instrument that allow one to "rule out" non-cases, such as a screening measure designed to identify students who do not require services, in an effort to optimize *specificity* of case identification. ROC methods involve testing links between measures of interest and their ability to either sensitively or specifically distinguish discrete events on a criterion variable, such as a student's status on intervention response metrics or indicators of diagnostic status. Recent examples of applications of these ROC methods to psychological assessment exist elsewhere (e.g., Jarrett, Van Meter, Youngstrom, Hilton, & Ollendick, 2018).

Tests using ROC methods may facilitate use of the discrepant and nondiscrepant scales with individual students. In fact, initial uses of ROC methods may test the utility of the scales to facilitate decision-making during circumstances of their optimal implementation. For example, official records (e.g., grade retention, below-average test scores), tasks (e.g., family interaction tasks or naturalistic classroom observations), and standardized tests can all serve as criterion measures in ROC tests. Further, the kind of criterion measure one selects ought to dictate which of the discrepant or nondiscrepant scales would achieve the highest sensitivity and/or specificity. For example, a discrepant scale designed to assess intervention responses in students' behaviors bound to the school context should display relatively higher sensitivity for detecting changes in school-specific performance, relative to the sensitivity metrics yielded for discrepant scales bound to the home context. Conversely, discrepant scales bound to the home context should outperform discrepant scales bound to the school context for home-specific criterion measures. Further, nondiscrepant scales should outperform any discrepant scale for criterion measures that

index behaviors that display consistently across contexts. Should discrepant and nondiscrepant scales meet these expectations, specific scores could be identified that sensitively and/or specifically index students' cross-contextual and context-specific behaviors.

Phase 4: Translating Scientific Findings into Routine Practice

Implementation science. Building from the previous phases, Phase 4 of the research agenda focuses on transferring scientific findings into routine practice in schools. In particular, the transdisciplinary field of *implementation science* focuses on bridging the gap between research and practice through the development and testing of frameworks and strategies to improve the uptake of evidence-based assessment and intervention practices into service delivery settings (Lobb & Colditz, 2013). Generally, implementation science examines techniques and methods (Powell et al., 2015) that optimize both implementation outcomes (e.g., fidelity, reach, feasibility, sustainment; Proctor et al., 2011) and client outcomes in specific service settings (e.g., hospital, mental health clinics, schools; Aarons, Hurlburt, & Horwitz, 2011; Damschroder et al., 2009; Proctor et al., 2011; Proctor et al., 2009). In this respect, two efforts might be particularly fruitful to pursue.

First, in this paper we reviewed an emerging but compelling body of research demonstrating the value of informant discrepancies for improving the interpretability of multi-informant assessments of students' psychosocial functioning. In doing so, we hope to increase efforts to disseminate this knowledge to scholars and practitioners in the discipline of school psychology and seed ideas for new research. At the same time, *how* might one carry out this work in school-based service settings? *What* measurement paradigms and statistical techniques might one leverage to test research questions germane to issues surrounding informant discrepancies in school-based assessments? We might facilitate answering questions such as

these by improving the synergy among scholars in school psychology, school professionals in the discipline, and scholars and professionals in disciplines where this kind of informant discrepancies work already occurs. Consequently, a key first step in implementation science endeavors should involve creating multi-disciplinary spaces, perhaps via pre-conference summits organized and held at professional conferences. These summits might focus on the sharing of ideas, resources, and strategies among scholars and professionals who regularly use multi-informant approaches to assess psychosocial functioning. These summits might result in cross-disciplinary collaborative teams. Members of these teams might establish research networks focused on addressing common questions surrounding informant discrepancies. We expect that establishing these collaborative networks will optimize the value of data gathered on these issues in school-based service settings, and at the same time reduce redundancies among research teams testing questions germane to informant discrepancies.

Second, implementation research often involves conducting investigations of typical or usual practice. As we noted previously, *collecting* multi-informant data already comprises a key component of "best practices" in school-based assessments (e.g., Reynolds et al., 2006; Salvia et al., 2012). Yet, it remains an empirical question how professionals in school-based service settings *use* multi-informant data to make decisions regarding educational programming. That is, do school professionals as part of routine practice make these decisions in a way that is sensitive to the contexts in which students require services and/or improve as a consequence of receiving services? For example, when school professionals encounter discrepancies between a parent's and teacher's reports of a student's psychosocial functioning, do they interpret these discrepancies as reflecting differences between how that student behaves at home versus school? Alternatively, do professionals often make decisions running on the assumption that when they

observe informant discrepancies, this signals that one or more of the informants provided reports of questionable precision or accuracy? Research focused on attaining estimates of how school professionals currently use multi-informant data to make decisions about educational programming comprises a necessary step toward understanding potential barriers in developing multi-informant assessments along the lines that we propose.

Implementation practice. *Implementation practice* is closely tied to implementation science and involves applying the knowledge generated from research to effectively install quality practices as part of routine service delivery. We need innovative research that examines how to address the implementation gap, in order to maximize both the efficient use of evidence-based assessment and intervention practices, and the impact of these practices on the students who could benefit from research on informant discrepancies (O'Connell, Boat, & Warner, 2009). Indeed, although significant attention has been devoted to cataloging barriers/facilitators (Damschroder et al., 2009) and identifying implementation strategies (Waltz et al., 2015), there is a need for research that develops and tests innovative strategies to enhance successful administration and use of evidence-based assessment practices. In schools, implementation initiatives often occur as top-down mandates without attention paid to the individual factors found to impact practitioner behavior change (e.g., perceptions and intentions influencing implementation fidelity). In the case of multi-informant assessments, failing to address individual-level barriers (e.g., knowledge, motivation, self-efficacy) may be costly. That is, individual behavior change is ultimately required for practitioners—school psychologists, special educators, and educational diagnosticians—to incorporate scientific findings on informant discrepancies into practice, even when organizational factors such as evidence-informed policy, supportive leadership, and effective training are in place (Michie, van Stralen, & West, 2011).

Thus, we need to investigate ways of effectively supporting practitioner behavior change to ensure successful transfer of knowledge on informant discrepancies.

Regarding practitioner behavior change, we might consider recent efforts focused on addressing long-standing barriers in assessment practices, albeit outside the realm of informant discrepancies research. Specifically, scholars and professionals focused on assessing and diagnosing pediatric bipolar disorder have long raised concerns about the lack of standardized approaches to training professionals on how to assess and diagnose the condition (for a review, see Youngstrom et al., 2017). In fact, the wide variation in training might very well account for the relatively low rates of inter-rater reliability often observed among assessors' diagnostic decisions (Youngstrom, Halverson, Youngstrom, Lindhiem, & Findling, 2018). Some of the issues stemming from current practices resolve themselves, in part, using short workshop training programs focused on behavior change. One program involves providing professionals with tools for making informed diagnostic decisions based on standardized instruments (Jenkins, Youngstrom, Washburn, & Youngstrom, 2011). Another program from the same team focuses on reducing decision-making biases and errors commonly found when professionals diagnose the condition (Jenkins & Youngstrom, 2016). Researchers designed both of these programs to be (a) brief; (b) scalable for use either online (Jenkins & Youngstrom, 2016), or within the workshop structures of professional meetings (Jenkins et al., 2011); and (c) interactive such that attendees both learn procedures and practice their implementation within the same program sequence.

We review this example to highlight approaches that research teams might modify for use in dealing with potential barriers to using and interpreting multi-informant assessments. For instance, these training programs might focus on providing professionals with tools for how to test the meaning of informant discrepancies when they encounter them. The tools might involve

strategic use of independent assessments of behaviors (e.g., home and school observations) to decipher whether a particular instance of discrepant reports (e.g., teachers report student behavior that is not corroborated by parent reports) occurred because the behaviors manifest in one context but not the other. Essentially, a key next step in implementation practice might involve creating feasible (i.e., brief, inexpensive), scalable programs designed to facilitate application of the Operations Triad Model (De Los Reyes, Thomas, et al., 2013) in school-based service settings. In sum, we expect a combination of research endeavors designed to impact both implementation science and implementation practice to facilitate adoption and use of the innovative multi-informant assessments we propose.

Concluding Comments

Across multiple disciplines tasked to deliver and study services for assessing and intervening upon children's psychosocial functioning, a key challenge is that the foci of these services are *moving targets*. The constituent behaviors that reflect psychosocial functioning vary widely in their displays, from context-to-context, and in the school setting, from among different students. Professionals in school-based service and research settings take great strides to account for all of this complexity by, among other practices, collecting data about students' psychosocial functioning from multiple key figures in students' lives. We ask students how they feel, how they behave, what they need. We ask their teachers and parents for their thoughts on these matters. These data tell us a lot about students. Yet, when each piece of data points us in a different direction, we realize that we need to learn more in order to make the right decisions for students and the services required to meet their needs. Stated otherwise, we need a better "roadmap."

We reviewed a body of work—occurring largely outside of school-based service delivery

and research—focused on what we can learn from the discrepancies between informants tasked to report on students' psychosocial functioning. We used this overview to inform the development of a four-phase research agenda designed for direct application to understanding these informant discrepancies as they manifest in school-based service and research settings. We have a lot of work ahead of us. We need new studies. We need new measures. We perhaps need to rethink the findings of prior studies and refine existing measures. We are also very curious about what lies ahead regarding research on multi-informant assessments. After reading our paper, we hope you are too.

References

Aarons, G.A., Hurlburt, M., & Horwitz, S.M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*, 4-23. doi: 10.1007/s10488-010-0327-7

Achenbach, T.M. (2006). As others see us: Clinical and research implications of cross-informant correlations for psychopathology. *Current Directions in Psychological Science, 15,* 94-98. doi: 10.1111/j.0963-7214.2006.00414.x

Achenbach, T.M., Krukowski, R.A., Dumenci, L., & Ivanova, M.Y. (2005). Assessment of adult psychopathology: Meta-analyses and implications of cross-informant correlations. *Psychological Bulletin, 131,* 361-382. doi: 10.1037/0033-2909.131.3.361

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. doi: 10.1037/0033-2909.101.2.213

Achenbach, T.M., & Rescorla, L. (2001). *Manual for the ASEBA School-Age Forms and Profiles.* Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (*5th ed.*) (*DSM-V*). Washington, DC: American Psychiatric Association.

Al Ghriwati, N., Winter, M.A., Greenlee, J.L., & Thompson, E.L. (2018). Discrepancies between parent and self-reports of adolescent psychosocial symptoms: Associations with family conflict and asthma outcomes. *Journal of Family Psychology*, *32*, 992-997. doi: 10.1037/fam0000459

Andrich, D., & Hagquist, C. (2012). Real and artificial differential item functioning. *Journal of Educational and Behavioral Statistics, 37,* 387-412. doi: 10.3102/1076998611411913

Becker-Haimes, E.M., Jensen-Doss, A., Birmaher, B., Kendall, P.C., & Ginsburg, G.S. (2018). Parent–youth informant disagreement: Implications for youth anxiety treatment. *Clinical Child Psychology and Psychiatry*, *23*, 42-56. doi: 10.1177/1359104516689586

Beidas, R.S., Stewart, R.E., Walsh, L., Lucas, S., Downey, M.M., Jackson, K., Fernandez, T., & Mandell, D.S. (2015). Free, brief, and validated: Standardized instruments for low-resource mental health settings. *Cognitive and Behavioral Practice*, *22*, 5-19. doi: 10.1016/j.cbpra.2014.02.002

Cicchetti, D. (1984). The emergence of developmental psychopathology. *Child Development, 55,* 1-7. doi: 10.2307/1129830

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conners, C.K. (1997). *Conners' Rating Scales–revised technical manual*. North Tonawanda, NY: Multi-Health Systems.

Cook, C.R., Volpe, R.J., & Delport, J. (2014). Systematic progress monitoring of students with emotional and behavioral disorders: The promise of change-sensitive brief behavior rating scales. In H.M. Walker and F.M. Gresham (Eds.), *Handbook of evidence-based practices for emotional and behavioral disorders: Applications in schools* (pp. 211-228). New York: Guilford.

Damschroder, L.J., Aron, D.C., Keith, R.E., Kirsh, S.R., Alexander, J.A., & Lowery, J.C. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science*, *4*, 50. doi: 10.1186/1748-5908-4-50

De Los Reyes, A. (2011). Introduction to the special section. More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children

and adolescents. *Journal of Clinical Child and Adolescent Psychology, 40,* 1-9. doi: 10.1080/15374416.2011.533405

De Los Reyes, A., & Aldao, A. (2015). Introduction to the special issue: Toward implementing physiological measures in clinical child and adolescent assessments. *Journal of Clinical Child and Adolescent Psychology, 44,* 221-237. doi: 10.1080/15374416.2014.891227

De Los Reyes, A., Alfano, C.A., Lau, S., Augenstein, T.M., & Borelli, J.L. (2016). Can we use convergence between caregiver reports of adolescent mental health to index severity of adolescent mental health concerns? *Journal of Child and Family Studies, 25*, 109-123. doi: 10.1007/s10826-015-0216-5

De Los Reyes, A., Augenstein, T.M., Wang, M., Thomas, S.A., Drabick, D.A.G., Burgers, D., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin, 141*, 858-900. doi: 10.1037/a0038498

De Los Reyes, A., Bunnell, B.E., & Beidel, D.C. (2013). Informant discrepancies in adult social anxiety disorder assessments: Links with contextual variations in observed behavior. *Journal of Abnormal Psychology, 122,* 376-386. doi: 10.1037/a0031150

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. doi: 10.1146/annurev-clinpsy-050212-185617

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,* 637-652. doi: 10.1007/s10802-009-9307-3

De Los Reyes, A., & Kazdin, A.E. (2009). Identifying evidence-based interventions for children and adolescents using the range of possible changes model: A meta-analytic illustration.

Behavior Modification, 33, 583-617. doi: 10.1177/0145445509343203

De Los Reyes, A., & Kazdin, A.E. (2006). Conceptualizing changes in behavior in intervention research: The range of possible changes model. *Psychological Review, 113,* 554-583. doi: 10.1037/0033-295X.113.3.554

De Los Reyes, A., Kundey, S.M.A., & Wang, M. (2011). The end of the primary outcome measure: A research agenda for constructing its replacement. *Clinical Psychology Review, 31,* 829-838. doi: 10.1016/j.cpr.2011.03.011

De Los Reyes, A., Lerner, M.D., Keeley, L.M., Weber, R., Drabick, D.A.G., Rabinowitz, J., & Goodman, K.L. (2019). Improving interpretability of subjective assessments about psychological phenomena: A review and cross-cultural meta-analysis. *Review of General Psychology*. Advance online publication. doi: 10.1177/108926801983764

De Los Reyes, A., & Ohannessian, C.M. (2016). Introduction to the special issue: Discrepancies in adolescent-parent perceptions of the family and adolescent adjustment. *Journal of Youth and Adolescence, 45*, 1957-1972. doi: 10.1007/s10964-016-0533-z

De Los Reyes, A., Ohannessian, C.M., & Racz, S.J. (2019). Discrepancies between adolescent and parent reports about family relationships. *Child Development Perspectives*, *13*, 53-58. doi: 10.1111/cdep.12306

De Los Reyes, A., Youngstrom, E.A., Swan, A.J., Youngstrom, J.K., Feeny, N.C., & Findling, R.L. (2011). Informant discrepancies in clinical reports of youths and interviewers' impressions of the reliability of informants. *Journal of Child and Adolescent Psychopharmacology, 21,* 417-424. doi: 10.1089/cap.2011.0011

Deros, D.E., Racz, S.J., Lipton, M.F., Augenstein, T.M., Karp, J.N., Keeley, L.M., . . . De Los Reyes, A. (2018). Multi-informant assessments of adolescent social anxiety: Adding

clarity by leveraging reports from unfamiliar peer confederates. *Behavior Therapy, 49*, 84-98. doi: 10.1016/j.beth.2017.05.001

Dirks, M.A., De Los Reyes, A., Briggs-Gowan, M.J., Cella, D., & Wakschlag, L.S. (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. doi: 10.1111/j.1469-7610.2012.02537.x

Dirks, M.A., Weersing, V.R., Warnick, E., Gonzalez, A., Alton, M., Dauser, C., Scahill, L., & Woolston, J. (2014). Parent and youth report of youth anxiety: Evidence for measurement invariance. *Journal of Child Psychology and Psychiatry, 55*, 284-291. doi: 10.1111/jcpp.12159

Dretzke, J., Davenport, C., Frew, E., Barlow, J., Stewart-Brown, S., Bayliss, S., . . . Hyde, C. (2009). The clinical effectiveness of different parenting programmes for children with conduct problems: a systematic review of randomised controlled trials. *Child and Adolescent Psychiatry and Mental Health*, *3*:7. doi: 10.1186/1753-2000-3-7

Duhig, A.M., Renk, K., Epstein, M.K., & Phares, V. (2000). Interparental agreement on internalizing, externalizing, and total behavior problems: A meta-analysis. *Clinical Psychology: Science and Practice, 7,* 435-453. doi: 10.1093/clipsy.7.4.435

Elliott, S. N., & Gresham, F. M. (2008). *Social Skills Improvement System*. Minneapolis, MN: Pearson Assessments.

Fergusson, D.M., Boden, J.M., & Horwood, L.J. (2009). Situational and generalised conduct problems and later life outcomes: evidence from a New Zealand birth cohort. *Journal of Child Psychology and Psychiatry*, *50*, 1084-1092. doi : 10.1111/ j .1469-7610.2009.02070.x

Fishbein, D.H., Ridenour, T.A., Stahl, M., & Sussman, S. (2016). The full translational spectrum of prevention science: Facilitating the transfer of knowledge to practices and policies that

prevent behavioral health problems. *Translational Behavioral Medicine*, *6*, 5-16. doi: 10.1007/s13142-015-0376-2

Garner, W.R., Hake, H.W., & Eriksen, C.W. (1956). Operationism and the concept of perception. *Psychological Review, 63,* 149-159. doi: 10.1037/h0042992

Glenn, L.E., Keeley, L.M., Szollos, S., Okuno, H., Wang, X., Rausch, E., . . . De Los Reyes, A. 2019). Trained observers' ratings of adolescents' social anxiety and social skills within controlled, cross-contextual social interactions with unfamiliar peer confederates. *Journal of Psychopathology and Behavioral Assessment*, *41*, 1-15. doi: 10.1007/s10862-018-9676-4

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire (SDQ). *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*, 1337-1345. doi: 10.1097/00004583-200111000-00015

Gresham, F.M., Elliott, S.N., Metallo, S., Byrd, S., Wilson, E., & Cassidy, K. (2018). Cross-informant agreement of children's social-emotional skills: An investigation of ratings by teachers, parents, and students from a nationally representative sample. *Psychology in the Schools*, *55*, 208-223. doi: 10.1002/pits.22101

Groth-Marnat, G., & Wright, A.J. (2016). *Handbook of psychological assessment* (6th ed.). Hoboken, NJ: Wiley.

Hartley, A.G., Zakriski, A.L., & Wright, J.C. (2011). Probing the depths of informant discrepancies: Contextual influences on divergence and convergence. *Journal of Clinical Child and Adolescent Psychology, 40,* 54-66. doi: 10.1080/15374416.2011.533404

Hunsley, J., & Lee, C. M. (2010). *Introduction to clinical psychology*. Hoboken, NJ: Wiley.

Hunsley, J., & Mash, E.J. (2007). Evidence-based assessment. *Annual Review of Clinical*

Psychology, 3, 29-51. doi: 10.1146/annurev.clinpsy.3.022806.091419

Individuals with Disabilities Education Improvement Act, H.R. 1350, 108th Congress (2004).

James, L.R., Demaree, R.G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69,* 85-98. doi: 10.1037/0021-9010.69.1.85

Jarrett, M.A., Van Meter, A., Youngstrom, E.A., Hilton, D.C., & Ollendick, T.H. (2018). Evidence-based assessment of ADHD in youth using a receiver operating characteristic approach. *Journal of Clinical Child and Adolescent Psychology*, *47*, 808-820. doi: 10.1080/15374416.2016.1225502

Jarrett, M., Siddiqui, S., Lochman, J., & Qu, L. (2014). Internalizing problems as a predictor of change in externalizing problems in at-risk youth. *Journal of Clinical Child and Adolescent Psychology*, *43*, 27-35. doi: 10.1080/15374416.2013.764823

Jenkins, M.M., & Youngstrom, E.A. (2016). A randomized controlled trial of cognitive debiasing improves assessment and treatment selection for pediatric bipolar disorder. *Journal of Consulting and Clinical Psychology*, *84*, 323-333. doi: 10.1037/ccp0000070

Jenkins, M.M., Youngstrom, E.A., Washburn, J.J., & Youngstrom, J.K. (2011). Evidence-based strategies improve assessment of pediatric bipolar disorder by community practitioners. *Professional Psychology: Research and Practice*, *42*, 121-129. doi: 10.1037/a0022506

Jones, J.D., Boyd, R.C., Calkins, M.E., Ahmed, A., Moore, T.M., Barzilay, R., Benton, T.D., & Gur, R.E. (2019). Parent-adolescent agreement about adolescents' suicidal thoughts. *Pediatrics*, *143*, 1-10. doi: 10.1542/peds.2018-1771

Kanne, S.M., Abbacchi, A.M., & Constantino, J.N. (2009). Multi-informant ratings of psychiatric symptom severity in children with autism spectrum disorders: The

importance of environmental context. *Journal of Autism and Developmental Disorders, 39*, 856-864. doi: 10.1007/s10803-009-0694-7

Korelitz, K.E., & Garber, J. (2016). Congruence of parents' and children's perceptions of parenting: A meta-analysis. *Journal of Youth and Adolescence*, *45*, 1973-1995. doi: 10.1007/s10964-016-0524-0

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. *The American Journal of Psychiatry, 160,* 1566-1577. doi: 10.1176/appi.ajp.160.9.1566

Laird, R.D., & De Los Reyes, A. (2013). Testing informant discrepancies as predictors of early adolescent psychopathology: Why difference scores cannot tell you what you want to know and how polynomial regression may. *Journal of Abnormal Child Psychology*, *41*, 1-14. doi: 10.1007/s10802-012-9659-y

Laird, R.D. & Weems, C.F. (2011). The equivalence of regression models using difference scores and models using separate scores for each informant: Implications for the study of informant discrepancies. *Psychological Assessment*, *23*, 388-397. doi: 10.1037/a0021926

LeBreton, J.M., Burgess, J.R.D., Kaiser, R.B., & James, L.R. (2003). The restriction of variance hypothesis and interrater reliability and agreement: Are ratings from multiple sources really dissimilar? *Organizational Research Methods, 6,* 80-128. doi: 10.1177/1094428102239427

Lee, S.W., Elliott, J., & Barbour, J.D. (1994). A comparison of cross-informant behavior ratings in school-based diagnosis. *Behavioral Disorders, 19,* 87-97. Retrieved from http://www.jstor.org/stable/23887490

Lerner, M.D., De Los Reyes, A., Drabick, D.A.G., Gerber, A.H., & Gadow, K.D. (2017). Informant discrepancies define discrete, clinically useful autism spectrum disorder subgroups. *Journal of Child Psychology and Psychiatry, 58*, 829-839. doi: 0.1111/jcpp.12730

Lobb R, Colditz GA. (2013). Implementation science and its application to population health. *Annual Review Public Health*, *34*, 235-251. doi: 10.1146/annurev-publhealth-031912-114444

Luthar, S.S., Cicchetti, D., & Becker, B. (2000). The construct of resilience: A critical evaluation and guidelines for future work. *Child Development, 71,* 543-562. doi: 10.1111/1467-8624.00164

Makol, B.A., De Los Reyes, A., Ostrander, R., & Reynolds, E.K. (2019). Parent-youth divergence (and convergence) in reports of youth internalizing problems in psychiatric inpatient care. *Journal of Abnormal Child Psychology*. Advance online publication. doi: 10.1007/s10802-019-00540-7

Makol, B.A., & Polo, A.J. (2018). Parent-child endorsement discrepancies among youth at chronic-risk for depression. *Journal of Abnormal Child Psychology*, *46*, 1077-1088. doi: 10.1007/s10802-017-0360-z

McCutcheon, A.L. (1987). *Latent class analysis*. Newbury Park, CA: Sage Publications.

Merrell, K.W. (2008). *Behavioral, social, and emotional assessment of children and adolescents* (3rd ed.). New York: Erlbaum/Taylor & Francis.

Michie, S., Van Stralen, M.M., & West, R. (2011). The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implementation Science*, *6*, 42. doi: 10.1186/1748-5908-6-42

National Institute of Mental Health. (2015). *The National Institute of Mental Health strategic plan for research* (NIH Publication No. 15-6368). Retrieved from http://www.nimh.nih.gov/about/strategic-planning-reports/index.shtml

O'Connell, M.E., Boat, T., & Warner, K.E. (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities* (Vol. 7). Washington, DC: National Academies Press.

Olino, T.M., Finsaas, M., Dougherty, L.R., & Klein, D.N. (2018) Is parent–child disagreement on child anxiety explained by differences in measurement properties? An examination of measurement invariance across informants and time. *Frontiers in Psychology, 9*, 1295. doi: 10.3389/fpsyg.2018.01295

Powell, B.J., Waltz, T.J., Chinman, M.J., Damschroder, L.J., Smith, J.L., Matthieu, M.M., . . . Kirchner, J.E. (2015). A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. *Implementation Science*, *10*, 21. doi: 10.1186/s13012-015-0209-1

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., . . . Hensley, M. (2011). Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, *38*, 65-76. doi: doi.org/10.1007/s10488-010-0319-7

Proctor, E.K., Landsverk, J., Aarons, G., Chambers, D., Glisson, C., & Mittman, B. (2009). Implementation research in mental health services: an emerging science with conceptual, methodological, and training challenges. *Administration and Policy in Mental Health and Mental Health Services Research*, *36*, 24-34. doi: 10.1007/s10488-008-0197-4

Ratcliffe, B., Wong, M., Dossetor, D., & Hayes, S. (2014). Teaching social–emotional skills to school-aged children with Autism Spectrum Disorder: A treatment versus control trial in 41 mainstream schools. *Research in Autism Spectrum Disorders*, *8*, 1722-1733. doi:10.1016/j.rasd.2014.09.010

Renk, K., & Phares, V. (2004). Cross-informant ratings of social competence in children and adolescents. *Clinical Psychology Review, 24,* 239-254. doi: 10.1016/j.cpr.2004.01.004

Reynolds, C., & Kamphaus, R. (2004). Behavior Assessment System for Children, (BASC-2) handout. *AGS Publishing*, *4201*, 55014-1796.

Rescorla, L.A., Bochicchio, L., Achenbach, T.M., Ivanova, M.Y., Almqvist, F., Begovac, I., . . . Verhulst, F.C. (2014). Parent-teacher agreement on children's problems in 21 societies. *Journal of Clinical Child and Adolescent Psychology*, *43*, 627-642. doi: 10.1080/15374416.2014.900719

Rescorla, L.A., Ewing, G., Ivanova, M.Y., Aebi, M., Bilenberg, N., Dieleman, G.C., . . . Verhulst, F.C. (2017). Parent–adolescent cross-informant agreement in clinically referred samples: Findings from seven societies. *Journal of Clinical Child and Adolescent Psychology*, *46*, 74-87. doi: 10.1080/15374416.2016.1266642

Reynolds, C.R., Livingston, R.B., & Willson, V. (2006). *Measurement and assessment in education.* Upper Saddle River, NJ: Pearson.

Romano, E., Weegar, K., Babchishin, L., & Saini, M. (2018). Cross-informant agreement on mental health outcomes in children with maltreatment histories: A systematic review. *Psychology of Violence, 8*, 19-30. doi: 10.1037/vio0000086

Russell, J.D., Graham, R. A., Neill, E.L. & Weems, C.F. (2016). Agreement in youth-parent perceptions of parenting behaviors: A case for testing measurement invariance in reporter discrepancy research. *Journal of Youth and Adolescence*, *45*, 2094-2107. doi: 10.1007/s10964-016-0495-1

Salvia, J., Ysseldyke, J., & Bolt, S. (2012). *Assessment: In special and inclusive education*. New York, NY: Cengage.

Sattler, J.M., & Hoge, R.D. (2006). *Assessment of children: Behavioral, social, and clinical foundations* (5th ed.). San Diego, CA: Jerome M. Sattler Publisher Inc.

Schopler, E., Reichler, J., & Renner, B. (1988). *The Childhood Autism Rating Scale (C.A.R.S.)*. Los Angeles: Western Psychological Services.

Sparrow, S.S., Balla, D., & Cicchetti, D. (1984). *Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Service.

Stratis, E. A., & Lecavalier, L. (2015). Informant agreement for youth with autism spectrum disorder or intellectual disability: A meta-analysis. *Journal of Autism and Developmental Disorders, 45*, 1026-1041. doi: 10.1007/s10803-014-2258-8

Sulik, M.J., Blair, C., Greenberg, M., & Family Life Project Investigators. (2017). Child conduct problems across home and school contexts: A person-centered approach. *Journal of Psychopathology and Behavioral Assessment*, *39*, 46-57. doi:10.1007/s10862-016-9564-8

Valo, S., & Tannock, R. (2010). Diagnostic instability of DSM-IV ADHD subtypes: Effects of informant source, instrumentation, and methods for combining symptom reports. *Journal of Clinical Child and Adolescent Psychology, 39*, 749-760. doi: 10.1080/15374416.2010.517172

Voelker, S.L., Shore, D.L., Lee, C.H., & Szuszkiewicz, T.A. (2000). Congruence in parent and teacher ratings of adaptive behavior of low-functioning children. *Journal of Developmental and Physical Disabilities*, *12*, 367-376. doi: 10.1023/A:1009436230984

Waltz, T.J., Powell, B.J., Matthieu, M.M., Damschroder, L.J., Chinman, M.J., Smith, J.L., . . . Kirchner, J.E. (2015). Use of concept mapping to characterize relationships among implementation strategies and assess their feasibility and importance: results from the Expert Recommendations for Implementing Change (ERIC) study. *Implementation Science*, *10*, 109. doi: 10.1186/s13012-015-0295-0

Wolraich, M.L., Feurer, I., Hannah, J.N., Pinnock, T.Y., & Baumgaertel, A. (1998). Obtaining systematic teacher reports of disruptive behavior disorders utilizing DSM-IV. *Journal of Abnormal Child Psychology*, *26*, 141-152. doi: 10.1023/A:1022673906401

Wolraich, M. L., Lambert, W., Doffing, M. A., Bickman, L., Simmons, T., & Worley, K. (2003). Psychometric properties of the Vanderbilt ADHD diagnostic parent rating scale in a referred population. *Journal of Pediatric Psychology*, *28*, 559-568. doi: 10.1093/jpepsy/jsg046

Wolraich, M.L., Lambert, E.W., Bickman, L., Simmons, T., Doffing, M.A., & Worley, K.A. (2004). Assessing the impact of parent and teacher agreement on diagnosing attention-deficit hyperactivity disorder. *Journal of Developmental and Behavioral Pediatrics*, *25*, 41-47. doi: 10.1097/00004703-200402000-00007

World Health Organization (2018). *Constitution of the World Health Organization*: *Principles*. Retrieved from http://www.who.int/about/mission/en/

Youngstrom, E.A. (2013). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology*, *39*, 204-221. doi: 10.1093/jpepsy/jst062

Youngstrom, E.A., VanMeter, A., Frazier, T.W., Hunsley, J., Prinstein, M.J., Ong, M.L., & Youngstrom, J.K. (2017). Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice*, *24*, 331-363. doi: 10.1111/cpsp.12207

Youngstrom, E.A., Halverson, T.F., Youngstrom, J.K., Lindhiem, O., & Findling, R.L. (2018). Evidence-based assessment from simple clinical judgments to statistical learning: Evaluating a range of options using pediatric bipolar disorder as a diagnostic challenge. *Clinical Psychological Science*, *6*, 243-265. doi: 10.1177/21677026177418

Table 1

Multi-Informant Instruments with Documented Evidence of Low Cross-Informant Correspondence When Used in a School-Based Sample

Measure	Primary Measure Reference by Developers	Number of Citations for Primary Reference [a]	Informants [b]	Study of a School-Based Sample Documenting Low Cross-Informant Correspondence
Achenbach System of Empirically Based Assessment (ASEBA)	Achenbach & Rescorla (2001)	8,884	P, T, C	Lee, Elliott, & Barbour (1994)
Behavioral Assessment System for Children (BASC-2/BASC-3)	Reynolds & Kamphaus (2004)	4,467	P, T, C	Jarrett, Siddiqui, Lochman, & Qu (2014)
Social Skills Improvement System—Rating Scales (SSIS-RS)	Elliott & Gresham (2008)	857	P, T, C	Ratcliffe, Wong, Dossetor, & Hayes (2014)
Strength and Difficulties Questionnaire (SDQ)	Goodman (2001)	4,063	P, T, C	Ratcliffe et al., 2014
Vanderbilt ADHD Diagnostic Parent Rating Scale (VADPRS), Vanderbilt ADHD Diagnostic Teacher Rating Scale (VADTRS)	Wolraich et al. (1998, 2003)	653	P, T	Wolraich et al. (2004)
Vineland Adaptive Behavior Scale (VABS)	Sparrow, Balla, & Cicchetti (1984)	6,562	P,T	Voelker, Shore, Lee, & Szuszkiewicz (2000)

[a] Citation counts based on Google Scholar reference search conducted on September 1, 2018. [b] P = Parent; T = Teacher; C = Child.

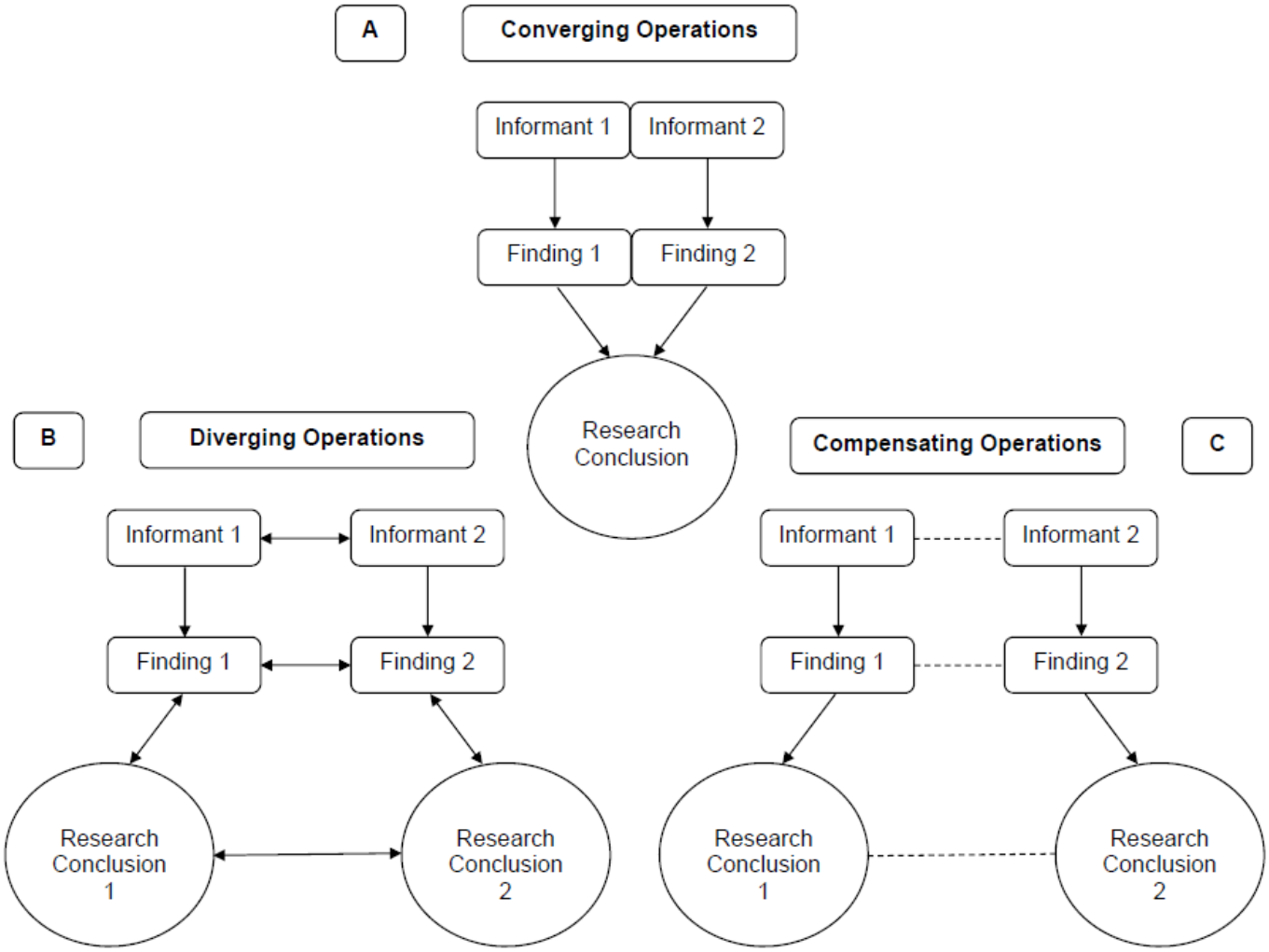

Figure 1. Graphical representation of the research concepts that comprise the Operations Triad Model. Originally published in De Los Reyes, Thomas, et al. (2013). © Annual Review of Clinical Psychology. Copyright 2012 Annual Reviews. All rights reserved. The Annual Reviews logo, and other Annual Reviews products referenced herein are either registered trademarks or trademarks of Annual Reviews. All other marks are the property of their respective owner and/or licensor.

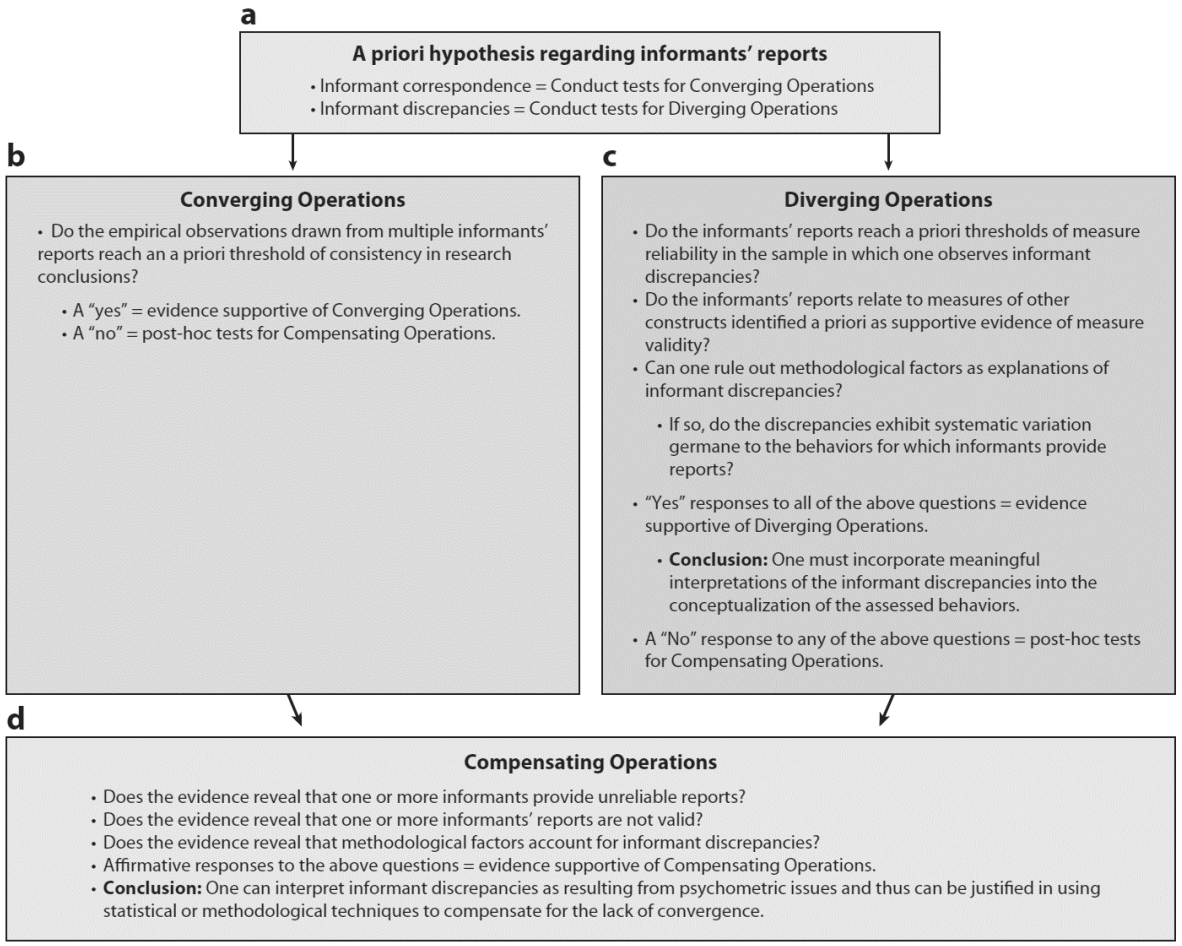

Figure 2. Graphical display of decision-making processes based on the Operations Triad Model. Originally published in De Los Reyes, Thomas, et al. (2013).

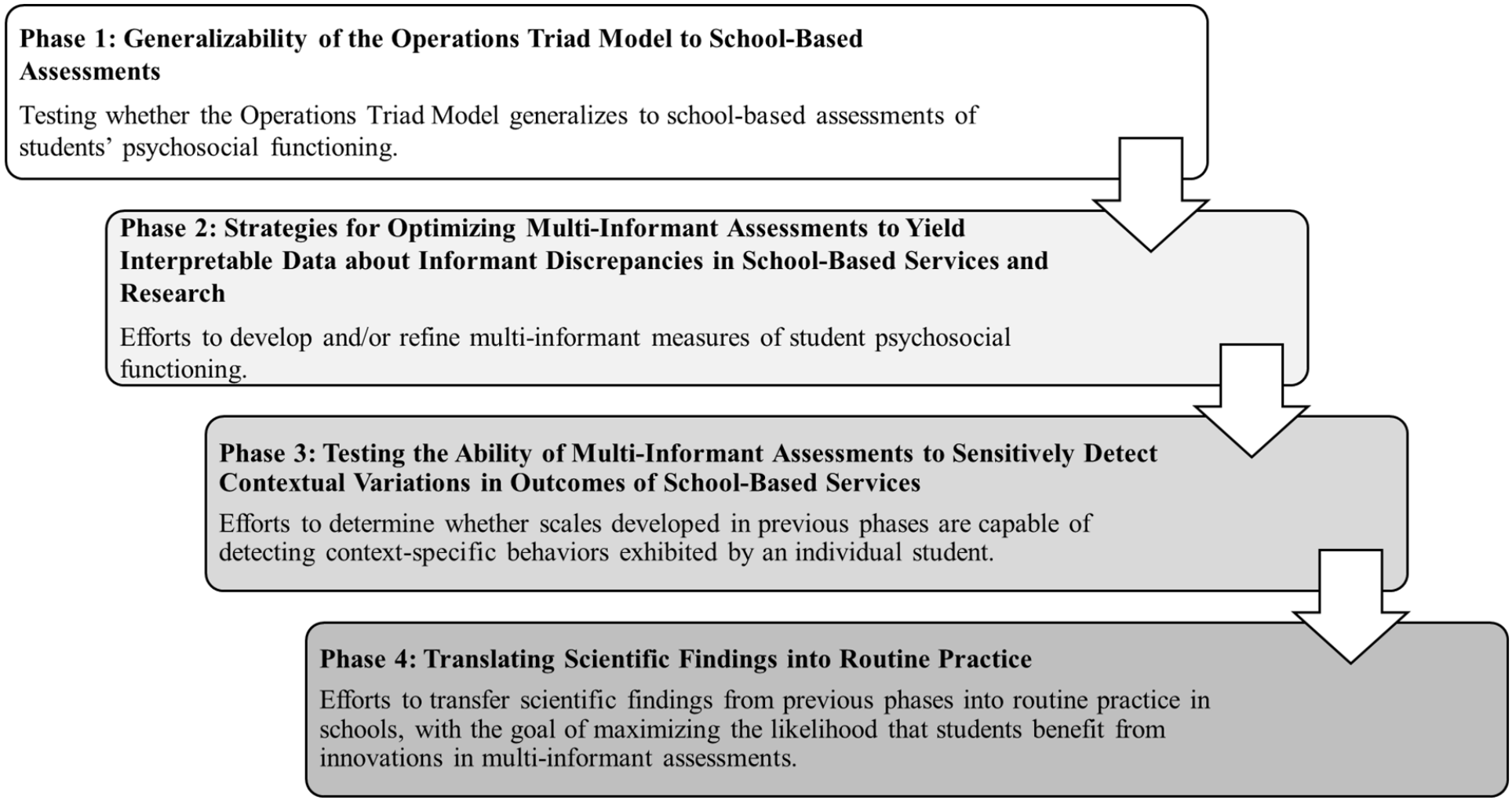

Figure 3. Graphical display of the four phases of proposed research agenda for informant discrepancies in school-based services and research.