



Published in final edited form as:

Annu Rev Clin Psychol. 2013 ; 9: 123–149. doi:10.1146/annurev-clinpsy-050212-185617.

Principles Underlying the Use of Multiple Informants' Reports

Andres De Los Reyes,

Comprehensive Assessment and Intervention Program, Department of Psychology, University of Maryland, College Park, MD

Sarah A. Thomas,

Comprehensive Assessment and Intervention Program, Department of Psychology, University of Maryland, College Park, MD

Kimberly L. Goodman, and

Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

Shannon M.A. Kunday

Department of Psychology, Hood College, Frederick, MD

Andres De Los Reyes: adlr@umd.edu; Sarah A. Thomas: thomas82@umd.edu; Kimberly L. Goodman: goodman.kimberly@gmail.com; Shannon M.A. Kunday: kunday@hood.edu

Abstract

Researchers use multiple informants' reports to assess and examine behavior. However, informants' reports commonly disagree. Informants' reports often disagree in their perceived levels of a behavior ("low" vs. "elevated" mood), and examining multiple reports in a single study often results in inconsistent findings. Although researchers often espouse taking a multi-informant assessment approach, they frequently address informant discrepancies using techniques that treat discrepancies as measurement error. Yet, recent work indicates that researchers in a variety of fields often may be unable to justify treating informant discrepancies as measurement error. In this paper, the authors advance a framework (Operations Triad Model) outlining general principles for using and interpreting informants' reports. Using the framework, researchers can test whether or not they can extract meaningful information about behavior from discrepancies among multiple informants' reports. The authors provide supportive evidence for this framework and discuss its implications for hypothesis testing, study design, and quantitative review.

Keywords

Converging Operations; Diverging Operations; Compensating Operations; Informant Discrepancies; Operations Triad Model

Does the average self and other agreement in children or adults constitute an epistemological crisis or a psychometric challenge? We think it is the latter. These estimates are the averages between two judges. Using two judges is like using a

two-item self-report scale. Few researchers would endorse the practice in the self-report domain and the observational domain is no different. Indeed, the modal level of interjudge agreement found in both the child and adult literature poses an easily surmountable psychometric challenge with a relatively straightforward prescription: Use more judges.

Roberts BW, Caspi A. 2001. Authors' response: Personality development and the person-situation debate: It's déjà vu all over again. *Psychol. Inq.* 12:104–109

As we indicated, in assessing youth, it has been a longstanding practice for clinicians to obtain assessment data from multiple informants such as parents, teachers, and the youths themselves. It is now commonly accepted that, because of differing perspectives, these informant ratings will not be interchangeable but can each provide potentially valuable assessment data...

Hunsley J, Mash EJ. 2007. Evidence-based assessment. *Annu. Rev. Clin. Psychol.* 3:29–51

These two statements refer to using and interpreting multiple informants' reports in psychological assessments. Researchers often use multiple informants' reports to assess children, adolescents, and adults (hereafter, we refer to children and adolescents collectively as "children" unless otherwise specified) (Achenbach, 2006; De Los Reyes & Kazdin, 2008; Oltmanns & Turkheimer, 2009). These informants include the person being assessed (e.g., patient), significant others (e.g., spouses in the case of adults; parents, teachers, and peers in the case of children), clinicians, and laboratory observers. In fact, use and interpretation of multiple informants' reports comprise key components of best practices in evidence-based assessment (Dirks, De Los Reyes, Briggs-Gowan, Cella, & Wakschlag, 2012; Hunsley & Mash, 2007). However, inconsistencies often arise among multiple informants' reports (hereafter referred to as "informant discrepancies") (Achenbach, 2006). Researchers observe informant discrepancies, even when informants complete parallel or identical measures (De Los Reyes, 2011).

Informants often disagree in their perceived levels of a behavior. For instance, one informant may report on a questionnaire that a patient's mood is "low" and another informant may report on a parallel questionnaire that the patient's mood is "elevated" (De Los Reyes, Youngstrom et al., 2011a, b). Consequently, informant discrepancies may profoundly impact how researchers interpret empirical findings, and how practitioners interpret assessment outcomes in clinical practice. Indeed, consider the implications informant discrepancies may have when observed within assessments of patients' functioning. For instance, informant discrepancies arise within assessments conducted for the purposes of: (a) screening (e.g., Does the patient evidence "clinically relevant" depressive symptoms?), (b) diagnosis (e.g., Does the patient meet diagnostic criteria for major depressive disorder?), and (c) treatment response (e.g., Does Treatment X effectively reduce the patient's depressive symptoms?) (De Los Reyes & Kazdin, 2005). Informant discrepancies often translate into inconsistent findings and thus raise dilemmas for interpreting the evidence regarding, for instance, the efficacy of interventions (De Los Reyes, Alfano, & Beidel, 2010, 2011). Thus, informant discrepancies introduce uncertainty into decision-making in research and practice settings.

The “Grand Discrepancy” in Multi-Informant Assessment

Taken together, the quoted passages at the beginning of this paper reflect disparate philosophies on how to interpret informant discrepancies. Roberts and Caspi (2001) interpret informant discrepancies as “nuisances” that require a methodological resolution. This interpretation can be traced to the concept of measurement error. Edgeworth (1888) translated the concept of measurement error from the physical sciences for use in the study of mental states, and it ultimately formed a key component of classical test theories of psychological measurement (Borsboom, 2005). Specifically, variations among multiple informants’ reports of the same behavior can be seen as “error” around a “true score mean” representation of the behavior being assessed (Edgeworth, 1888). Conversely, Hunsley and Mash (2007) interpret informant discrepancies as containing some kind of value. For example, informant discrepancies may represent variations in the expressions of assessed behaviors across contexts, as informants often vary in where they observe behavior (e.g., home vs. school; De Los Reyes, 2011). Achenbach, McConaughy, and Howell (1987) lucidly reflected this interpretation, in a metaanalysis of correspondence between informants’ reports of child mental health that was published nearly a century after Edgeworth (1888).

In this paper, we do not focus on differences between researchers’ views on multiple informants’ reports (i.e., Caspi & Roberts vs. Hunsley & Mash). Instead, we focus on differences within researchers—the differences between their views on multiple informants’ reports and the actions they often take to address informant discrepancies. We refer to this as the “Grand Discrepancy” in multi-informant assessment. Specifically, researchers often rationalize the importance of multi-informant assessment approaches using ideas that align with those of Hunsley and Mash (2007) (e.g., Klein, Dougherty, & Olino, 2005; McMahan & Frick, 2005; Pelham, Fabiano, & Massetti, 2005; Silverman & Ollendick, 2005). At the same time, researchers often manage the discrepancies that arise out of multi-informant assessments using techniques that more closely align with the ideas of Roberts and Caspi (2001).

Three techniques currently implemented in psychological assessment highlight the key issues that drive our discussion about principles underlying use of multiple informants’ reports. First, researchers often address discrepant reports by applying a combinational algorithm to the outcomes of the reports. For example, researchers may implement an “and” rule, or a requirement that at least two (or all) informants positively endorse the presence of a specific behavior in order for that behavior to be considered present (Offord et al., 1996). Conversely, researchers may apply an “or” rule, or a requirement of a positive endorsement from at least one informant in order for the behavior to be considered present (Piacentini, Cohen & Cohen, 1992; Youngstrom, Findling & Calabrese, 2003). Researchers often use these rules when they anticipate low correspondence among measures(s) completed by informants (Goodman et al., 1997; Lofthouse, Fristad, Splaingard & Kelleher, 2007). Second, researchers use statistical techniques to examine, in combination, multiple measures of a single behavior to arrive at an unobserved or “latent” representation of that behavior (e.g., structural equations models; Borsboom, 2005). Often, these techniques focus on examining the variance shared among multiple informants’ reports of the same behavior

(e.g., Arseneault et al., 2003; Holmbeck, Li, Schurman, Friedman, & Coakley, 2002; Zhou, Lengua, & Wang, 2009).¹ Third, in controlled trials of treatment efficacy, researchers often select, in advance of a trial, a single outcome indicator to represent treatment response—a “primary” outcome measure often completed by a single informant (e.g., DelBello et al., 2006; Klinberg et al., 2005; Papakostas, Mischoulon, Shyu, Alpert, & Fava, 2010). Within the same trial, other informants might complete outcome measures on the same outcome domain assessed by the primary outcome measure; researchers often label these “secondary outcome measures” (e.g., Pettinati et al., 2010; Thurstone, Riggs, Salomonsen-Sautel, Mikulich-Gilbertson, 2010).

A common assumption made when using combinational algorithms, structural equations models, or primary outcome measures is that informant discrepancies (i.e., unshared variance) should be treated as measurement error (see also De Los Reyes, Kundey & Wang, 2011; Dirks et al., 2012; Sood et al., in press). Thus, they highlight a Grand Discrepancy between ideas about the utility of multi-informant assessments, and the actions researchers take to address informant discrepancies. Importantly, researchers rarely conduct empirical tests to examine whether they can justifiably treat informant discrepancies as measurement error.

It is difficult to fault researchers for the Grand Discrepancy. Indeed, historically researchers have had one guide for interpreting multiple reports about the same behavior. Specifically, a core principle of current research training is that a single study should include different methods for quantifying the same phenomenon (e.g., multiple informants’ reports of the same behavior), with all findings derived from these methods converging on the same set of observations or conclusions. We train our students—as our professors trained us—to assume that convergence equals truth. The more our measurements agree, the greater our confidence should be in our interpretations of these measurements. Meaningful observations can be drawn from convergence among reports, and thus, there is a lack of meaning inherent in divergence among reports. Indeed, we even have a term for this interpretation: Converging Operations (Garner, Hake, & Erikson, 1956). Yet, researchers have no other guides for identifying alternative possibilities for interpreting multiple reports about the same behavior.

Should we assume that assessing convergence among findings is the only method for meaningfully interpreting patterns of findings across multiple informants’ reports? For instance, cannot a patient express a behavior to a greater degree in one setting (home) than she does in another (school, work)? If such behavioral expressions occur systematically, shouldn’t researchers be able to reliably and validly measure this divergence in behavioral

¹Recently, researchers in the behavior genetics literature have tested whether unshared variance reflects important information unique to informants, rather than measurement error (Bartels, Boomsma, Hudziak, van Beijsterveldt, & van den Oord, 2007). Specifically, researchers may construct structural models that partition variance reflecting genetic influences on the expression of the assessed behavior, shared environmental influences, and non-shared environmental influences, in addition to variance explained by error (Bartels et al., 2007). In fact, informant discrepancies have provided reliable information about how a person’s environment might affect his or her behavior differently, depending on where the informants providing reports tend to observe that person’s behavior (e.g., Derks, Hudziak, Van Beijsterveldt, Dolan, & Boomsma, 2006). Thus, within structural models seeking to explain variance in behavior, behavior genetics researchers may treat informant discrepancies as useful information (Baker, Jacobson, Raine, Lozano, & Bezdjian, 2007; Bartels et al., 2007; Derks et al., 2006; Tackett, Waldman & Lahey, 2009). At the same time, as alluded to previously not all literatures treat unshared variance in structural equations models in this way (cf. Holmbeck et al., 2002; Lambert, Salzer & Bickman, 1998; Valk, van den Oord, Verhulst & Boomsma, 2001).

expressions using the discrepancies among multiple informants' reports? Indeed, the settings within which a patient lives may vary as a function of the contingencies (e.g., corporeal punishment and praise) that reinforce expressions of her behavior (e.g., Skinner, 1953), and as mentioned previously, multiple informants may vary in where they observe a patient's behavior (De Los Reyes, 2011).

Researchers have developed a number of conceptual models for understanding informant discrepancies within specific areas (e.g., developmental psychopathology and controlled trials research; De Los Reyes, in press; De Los Reyes & Kazdin, 2005, 2006; Goodman, De Los Reyes, & Bradshaw, 2010; Kraemer et al., 2003). Yet, beyond Converging Operations, clinical science does not have a general framework for understanding and interpreting multiple informants' reports within a study. As a result, critical questions often go unaddressed within studies using multiple informants' reports. Specifically, what empirical questions would we pose to detect meaningful divergence among multiple informants' reports of the same behavior? How would we design studies to address questions of meaningful divergence? Do we even have a term for meaningful divergence? We do not. Without a framework for describing and testing the possibility of meaningfully interpreting informant discrepancies, researchers may default to erroneously interpreting such discrepancies using conceptual and measurement models that assume they reflect error.

The Present Review

This paper advances psychological assessment in three ways. First, we elaborate on the Grand Discrepancy. Specifically, we focus on Converging Operations: A set of measurement conditions for interpreting patterns of findings based on the consistency within which findings yield similar conclusions (Garner et al., 1956). Further, we review work in sub-disciplines of clinical science (e.g., developmental psychopathology, controlled trials research, and personality disorders assessments) that supports incorporating research concepts other than Converging Operations for interpreting multiple informants' reports.

Second, we review evidence indicating that, under some circumstances, informant discrepancies yield meaningful information about behavior. We introduce a concept to describe these instances (Diverging Operations). We also introduce a concept to describe instances within which researchers can interpret informant discrepancies as reflecting measurement error, and thus can justifiably treat informant discrepancies as measurement error (Compensating Operations). We integrate these concepts into a framework to guide interpretations of multi-informant assessments (Operations Triad Model [OTM]). The framework focuses on determining and characterizing the circumstances within which multiple informants' reports yield either converging or diverging findings. Lastly, we provide recommendations on how to address informant discrepancies and highlight directions for future research.²

²We focus on inconsistent findings as they relate to multi-informant assessments. Yet, inconsistent findings arise in many other circumstances of multi-method assessment. For instance, researchers commonly observe low correspondence between subjective reports of anxious arousal and objective measures of psychophysiology (e.g., measurements of heart rate or skin conductance; for a review, see Thomas, Aldao, & De Los Reyes, in press). We encourage further thought and inquiry on whether the issues we discuss this paper extend to comprehensive assessment occasions other than multi-informant assessments.

Understanding the Grand Discrepancy in Use and Interpretation of Multiple Informants' Reports

Converging Operations

Fundamental concepts in psychological measurement might provide insight into the disconnection between researchers' rationales for using multiple informants and how they interpret and manage informant discrepancies. Specifically, psychologists can rarely (if ever) argue that a single measure should be considered a definitive indicator to quantify a specific behavior (e.g., aggression, anxiety, attention, disruptive behavior, mood, racial bias, or self-esteem). Therefore, psychologists receive considerable training in constructing and administering various ways of quantifying behavior (e.g., multiple methods for assessing adulthood depressive symptoms). However, as mentioned previously, multi-informant, multi-method assessments robustly yield inconsistent outcomes and thus inconsistent empirical findings (Achenbach et al., 1987). Perhaps because inconsistent findings have historically been interpreted as a hindrance to advancing scientific knowledge, the typical approach to observing inconsistent findings is to stake the claim that science needs to progress in spite of these inconsistencies, not as a consequence of observing them (De Los Reyes, 2011).

As a result, psychological research training has historically relied on the concept of Converging Operations to interpret findings derived from multiple reports of the same behavior. Under Converging Operations, the strength of a single study's findings and thus the inferences one can draw from these findings rests on whether different ways of observing, quantifying, or examining the same behavior consistently point to the same research conclusion (Garner et al., 1956). Converging Operations differs from typical definitions of replication, or the ability to draw similar conclusions across multiple studies conducted under similar conditions (e.g., Pryor & Ostrom, 1981; Valentine et al., 2011). Relatedly, Campbell and Fiske (1959) drew distinctions between validity (i.e., agreement between two maximally different measures of the same trait) and reliability (i.e., agreement between two maximally similar measures of the same trait). One observes widespread use of Converging Operations across disparate fields including philosophy of science (e.g., Hempel, 1966), psychometrics (e.g., Nunnally & Bernstein, 1994), neuroscience (e.g., Gazzaniga, Ivry, & Mangun, 2008), and clinical science (e.g., Durston, 2008).

Applying Converging and Diverging Concepts to Sub-Disciplines in Clinical Science

The pervasiveness of the Converging Operations concept in multiple areas of science and the humanities appears sensible. Indeed, Converging Operations ought to lead researchers to base research conclusions on the preponderance of the evidence (e.g., Weisz, Jensen Doss, & Hawley, 2005; De Los Reyes & Kazdin, 2006). Yet, researchers should encounter difficulties with applying Converging Operations to every circumstance within which they seek to interpret patterns in research findings. For example, Converging Operations may insufficiently characterize instances within which (a) informants' reports yield diverging research findings and (b) this divergence reveals meaningful information germane to understanding the behavior being assessed. Thus, concepts for characterizing converging

and diverging research findings should play a role in interpreting the outcomes of multiple informants' reports.

In fact, research and theory within sub-disciplines of clinical science, namely developmental psychopathology, controlled trials research, and personality disorders, supports a focus on both converging and diverging research findings. First, in developmental psychopathology assessments, variations in people's behaviors may result from variations in people's environments (Skinner, 1953). For instance, one might infer meaningful divergence among parent and teacher reports of children's disruptive behavior by collecting data that meet two expectations. First, children considered as expressing clinically relevant disruptive behavior symptoms ought to vary in the settings within which they express such symptoms. Second, parent and teacher reports of children's behavior ought to reflect this expected variation in expression. Thus, if some children expressed disruptive behavior to a greater degree in home settings than they do school settings, parent reports of these children would evidence disruptive behavior symptoms to a greater degree than teacher reports of these same children. Conversely, observing meaningful convergence between reports might involve a study that supported researchers' expectations that (a) children express disruptive behavior symptoms consistently across the contexts within which parents and teachers observe behavior (e.g., home and school settings) and (b) these children evidence such behavior on reports completed by both parents and teachers.

Second, researchers conducting controlled trials typically hypothesize that any one of a trial's multiple outcomes should indicate the same or similar conclusions as to a treatment's efficacy. That is, one rarely (if ever) encounters a trial for which the investigative team hypothesizes: We expect one set of informants' outcome reports (e.g., self- and clinician reports) to support the treatment's efficacy to a greater degree than the support provided by another set (e.g., independent observations and teacher reports) (De Los Reyes & Kazdin, 2006). However, controlled trials often reveal inconsistent findings as to an intervention's effects (De Los Reyes & Kazdin, 2006, 2008, 2009; Koenig, De Los Reyes, Cicchetti, Scahill, & Klin, 2009). An important implication of relying on Converging Operations in controlled trials is that researchers typically refrain from advancing a priori hypotheses to address the possibility of inconsistent findings. Yet when prior studies do not support Converging Operations, it becomes important for researchers to understand inconsistent findings. Specifically, inconsistent findings might be explained by: (a) measurement error in one, some, or all of the informants' reports; or, in contrast, (b) meaningful variation across reports. Importantly, reliance on Converging Operations likely has led to researchers placing too little focus on whether inconsistent findings reveal information about variations in treatment response (De Los Reyes & Kazdin, 2008)

Third, as mentioned previously, low multi-informant correspondence is the norm rather than the exception in assessments of child, adolescent, and adult psychopathology (Achenbach et al., 1987; Achenbach, Krukowski, Dumenci, & Ivanova, 2005). These findings extend to observations of, at best, modest correspondence between self- and peer-reported personality pathology (Oltmanns & Turkheimer, 2009). Such findings are intriguing in light of the idea that personality disorder features are often conceptualized as stable across settings and time (American Psychiatric Association, 2001). Consistent with clinical child assessments (De

Los Reyes, 2011), the data indicate that discrepant reports of adult personality pathology may not be attributable to a lack of psychometric rigor on the part of self- and/or peer-reports (Thomas, Turkheimer, & Oltmanns, 2003). Yet, much of the focus on resolving the issue of discrepant reports involves identifying circumstances within which a given report (e.g., self, peer, and clinician) is “most valid” for assessing personality pathology (Klonsky, Oltmanns, & Turkheimer, 2002; Oltmanns & Turkheimer, 2009). An alternative method for interpreting these discrepant reports may involve distinguishing those circumstances within which the discrepant reports represent measurement error, or alternatively meaningful variation in the expression of personality pathology. Indeed, basic science in personality psychology indicates that a person’s expressions of behaviors indicative of personality traits can be dictated, in part, by the extent to which settings encountered by that person in daily life consistently elicit expressions of these behaviors (Mischel & Shoda, 1995). Alternatively, when discrepancies between reports reflect measurement error, this may provide justification to (a) identify a “most valid” informant for assessing personality pathology and/or (b) discard those informants’ reports that do not surpass acceptable levels of reliability and validity. Thus, even if a domain of personality pathology is conceptualized as stable across settings, it may be useful to examine whether discrepant outcomes in multi-informant assessments of such pathology yield meaningful information on how expressions of pathology vary across settings. In sum, research and theory in multiple areas of clinical science support the need to delineate research concepts for interpreting patterns of findings from multiple informants’ reports, beyond that of Converging Operations.

The Operations Triad Model (OTM): A Framework for Interpreting Multiple Informants’ Reports

Overview of Framework

In line with identifying multiple operations concepts apart from Converging Operations, it is important to conceptualize three decision-making processes. The first involves hypothesizing whether findings from multiple informants’ reports will converge on a common research conclusion, or diverge or point to different research conclusions. The second involves creating operational definitions of converging versus diverging research findings. The third involves characterizing or interpreting patterns of findings derived from multiple informants’ reports. We propose a framework by which to guide this decision making—the OTM.

A key tenet of the OTM is that researchers should anticipate whether their study will yield converging or diverging findings. Broadly, extensive evidence across multiple areas of clinical science indicates low-to-moderate levels of correspondence in multi-informant assessments (Achenbach, 2006; De Los Reyes & Kazdin, 2005; Klonsky et al., 2002; Oltmanns & Turkheimer, 2009). Thus, researchers can use this evidence to guide their predictions as to whether they will observe converging findings across multiple informants’ reports. In line with these predictions, researchers must use or develop rubrics for setting operational definitions for “converging” versus “diverging” research findings (De Los Reyes & Kazdin, 2006).

In turn, the OTM guides researchers toward proposing a priori hypotheses of whether findings based on informants' reports will converge, and if not, what they hypothesize the diverging findings will reflect (e.g., meaningful variation in behavior vs. measurement error). Specifically, for any one study, three questions are of interest: (a) Will the evidence support an interpretation of convergence among empirical findings across informants' reports of a behavior?, (b) If one expects divergence in empirical findings, will the observed divergence reflect meaningful information about variations in behavior?, and (c) In the absence of evidence indicating either converging research findings or meaningful reasons for observing diverging findings, will the divergence reflect measurement error or some other psychometric or methodological process? Thus, the OTM guides researchers toward posing a priori hypotheses as to whether informants' reports will converge or diverge, and then based on these reports, interpreting what patterns of converging or diverging empirical findings represent.

Considerations Underlying Development of the OTM

A key issue underlying development of the OTM involves delineating research concepts beyond Converging Operations. That is, what distinguishes new concepts for interpreting multiple informants' reports from Converging Operations? There are two key considerations.

Defining behavioral expressions within and across factors—The first consideration involves conceptualizing the consistency by which a behavior is expressed across a factor of interest. One factor might be context, or the settings within which an assessed behavior may be expressed. For instance, do children variably or invariably express disruptive behavior symptoms across home and school settings? Here, consider use of parent and teacher reports to assess childhood disruptive behavior, where such reports serve as indicators of a child's behavior as expressed in home (parent) versus school (teacher) settings. To define behavioral expressions across settings, a researcher decides both the possible settings within which children's disruptive behavior could be expressed, as well as whether children's disruptive behavior varies in its expression across these settings. Thus, if a researcher views "context" as reflecting variation in the expression of disruptive behavior within and/or across home and school settings, the key question is: Do I identify a child's behavior as disruptive if she or he expresses signs of disruptive behavior (a) across home and school settings, (b) in only one setting, or (c) either "a" or "b" (Kraemer et al., 2003)?

Ability of multiple informants to detect behavioral expressions across factors—A second consideration involves the expectation that multiple informants' reports about a behavior can reliably and validly "pick up" expressions of the behavior across a factor of interest (McFall & Townsend, 1998).³ Following from our previous example, a researcher might hypothesize that children's disruptive behavior varies as a function of context,

³These ideas share similarities with recent work on measurement validity (Borsboom, Mellenbergh, & van Heerden, 2004). Specifically, the two key factors to consider when gauging the validity of a measure are (a) whether the construct assessed by the measure exists and (b) whether variations in the construct causally produce variations in scores taken from the measure. Similarly, the features that may distinguish different ways of interpreting variations between multiple informants' reports should revolve around (1) how a researcher conceptualizes manifestations of the assessed construct and (2) whether a researcher considers convergence and/or divergence between informants' reports as valid reflections of the construct's manifestations.

whereby some children express disruptive behavior in only home or school settings. The key issue to address would be if parent and teacher reports reliably and validly detect variable expressions of behavior across settings. Stated another way, if one could independently assess children's disruptive behavior (e.g., laboratory observation task) so as to identify children who express such behavior within the home setting to a greater degree than they do the school setting, would this variation relate to parents reporting disruptive behavior to a greater degree than when based on teacher reports?

In sum, these two considerations allow researchers to develop operations concepts that can be (a) conceptually distinguished from Converging Operations and (b) used to interpret patterns of findings observed among multiple informants' reports. Below, we identify and describe two operations concepts apart from Converging Operations: Diverging Operations and Compensating Operations. We also discuss empirical work supporting the development of these research concepts. Further, we illustrate how researchers can interpret or characterize findings as instances of either of these two new concepts or Converging Operations. In Figure 1, we provide graphical depictions of Converging Operations (Figure 1a), as well as each of the new research concepts (Figures 1b, c).

Diverging Operations: When Informant Discrepancies Reflect Meaningful Information

Defining Diverging Operations—The Diverging Operations concept delineates circumstances within which a study yields evidence suggesting that (a) the behavior being assessed variably expresses itself across a factor of interest and (b) the discrepancies among multiple informants' reports reflect this meaningful variation in the behavior's expression (Figure 1b). On both conceptual and empirical grounds, extensive evidence supports developing this concept of Diverging Operations and distinguishing it from that of Converging Operations.

First, conceptual support for Diverging Operations comes from decades of basic psychological research indicating that systematic social and cognitive mental processes account for different people perceiving the same behaviors differently (Johnson, Hashtroudi, & Lindsay, 1993; Malle, 2006; Pasupathi, 2001). In fact, this work informed the development of a theoretical framework for why informant discrepancies exist in clinical assessments of child psychopathology: The Attribution Bias Context Model (ABC Model: De Los Reyes & Kazdin, 2005). The ABC Model posits that informant discrepancies exist because informants systematically differ on three central characteristics: (a) informants' own interpretations of why the child is expressing the assessed behaviors (i.e., child's disposition vs. environmental constraints on the child's behavior), (b) informants' perspectives or the decision thresholds that guide their identification of behaviors that warrant treatment, and (c) the contexts within which researchers and practitioners gather behavioral reports (e.g., clinical; community-based or epidemiological assessments), as well as the contexts within which informants observe the behavior (e.g., home and school). Importantly, the ABC Model delineates systematic differences among informants in how or within which contexts they observe behavior. Thus, theoretically these systematic differences among informants ought to translate into discrepancies among informants' reports that reflect meaningful variation in the behaviors about which informants provide reports.

Second, empirical support for Diverging Operations comes from a number of recent investigations of informant discrepancies conducted across diverse clinic and community samples, informants, and behavioral reports ****(De Los Reyes, in press, 2011). For the purposes of this paper, we will focus on those investigations that have examined whether discrepancies reflect variations in the contexts within which children express specific behaviors, and whether informants attend to and use contextual information when providing behavioral reports.

Informant discrepancies that reflect variations in behavior expressed in home versus school settings—Controlled laboratory observations support the idea that informant discrepancies reflect information about contextual variations in behavior. Specifically, in assessments of children’s disruptive behavior symptoms, researchers commonly encounter discrepancies between reports from parents and teachers about a child’s behavior. Much of the conceptual modeling of these discrepancies posits that they result from differences between informants’ reactions to the child’s behavior being assessed, and that these reactions reflect variations in children’s behavior across the contexts within which informants observe such behavior (e.g., home vs. school contexts; Achenbach et al., 1987; De Los Reyes & Kazdin, 2005; Dumenci, Achenbach, & Windle, 2011; Kraemer et al., 2003).

For example, a recent study examined whether informant discrepancies in reports of children’s disruptive behavior symptoms relate to assessments of these symptoms when measured across various laboratory controlled interactions between the child and multiple adults (De Los Reyes, Henry, Tolan, & Wakschlag, 2009). In this study, information on each child’s disruptive behavior was available via reports taken from parent and teacher. This study also included indices of disruptive behavior from an independent observational measure that assessed children’s disruptive behavior across multiple interactions between the child and her or his parent, or the child and an unfamiliar clinical examiner. Not surprisingly, De Los Reyes and colleagues (2009) identified substantial discrepancies in parent and teacher reports of children evidencing high levels of disruptive behavior symptoms, as well as substantial variation in disruptive behavior as observed in the laboratory. Interestingly, parent-teacher reporting discrepancies mapped onto variations in children’s disruptive behavior observed in the laboratory. For instance, laboratory observations of childhood disruptive behavior in the presence of the parent predicted disruptive behavior identified by parents only but not teachers only, relative to the children who did not evidence disruptive behavior based on either parent or teacher report. In contrast, laboratory observations of childhood disruptive behavior in the presence of an unfamiliar clinical examiner predicted disruptive behavior identified by teachers only but not parents only. Further, when disruptive behavior was identified by both parent and teacher, laboratory observations predicted disruptive behavior reports when the disruptive behavior was expressed within both parent-child and examiner-child interactions. However, observed disruptive behavior expressed exclusively within parent-child or examiner-child interactions (i.e., within one interaction and not the other) did not significantly predict disruptive behavior identified by both parent and teacher. Thus, De Los Reyes and colleagues (2009) identified variations between parent and teacher reports consistent with

Diverging Operations. Specifically, the discrepancies between reports reflected variations in children's disruptive behavior as a function of their laboratory interactions with parental versus non-parental adults.

As another example of informant discrepancies reflecting Diverging Operations, consider a recent study of parent and teacher reports of children's aggressive behavior (Hartley et al., 2011). In this study, parent-teacher discrepancies in reports of aggressive behavior related to informants' perceptions of the environmental cues that elicited aggressive behavior (e.g., negative peer interactions or demands placed on the child by adult authority figures). Specifically, Hartley and colleagues (2011) observed that increased informant correspondence related to increased similarities between the environments within which informants observed children's aggressive behavior. These findings dovetail with research and theory in personality psychology reviewed previously (Mischel & Shoda, 1995), indicating that whether a child expresses the same behavior (e.g., aggression) across contexts may be dictated, in part, by similarities among contexts in the contingencies that increase the likelihood of that behavior's expression.

Informants attend to contextual information when making reports—Importantly, the effects of contextual information on trained judges' (i.e., clinicians) and untrained informants' reports have recently received experimental support. For instance, De Los Reyes and Marsh (2011) presented clinicians with vignettes of children described as living in contexts that either posed risk for a conduct disorder diagnosis (e.g., exposure to parental psychopathology and poor peer relations) or contexts posing no such risk. Clinicians provided ratings of the likelihood that a comprehensive clinical evaluation of the child would suggest that the child meets criteria for a conduct disorder diagnosis. Based on these vignette descriptions, clinicians were more likely to view behaviors as indicative of a conduct disorder diagnosis if the children described in vignettes were living in contexts that pose risk for the disorder relative to contexts posing no such risk.

These findings from De Los Reyes and Marsh (2011) indicated that trained judges of patients' behavior attend to the contexts surrounding patients' clinical presentations when providing reports about patients' behavior. Thus, the findings raise the question as to whether discrepancies between untrained informants' reports might be explained, in part, by differences between informants in the contexts within which they observe behaviors about which they provide reports. Such an interpretation of informants' reports would fall within the scope of Diverging Operations. Interestingly, in another recent experimental study, researchers trained mothers and adolescents to systematically incorporate contextual information (e.g., settings such as "at parties" or "when shopping at the mall with friends") when providing reports about what parents know about their adolescents' whereabouts and activities (parental knowledge; De Los Reyes, Ehrlich et al., in press). In this study, informants' reports about parental knowledge completed after training became more discrepant from each other than when such reports were completed without training. Specifically, relative to reports completed without training, mothers reported significantly greater levels of parental knowledge than adolescents reported. Thus, these observations are consistent with Diverging Operations. Specifically, they support the idea that informants'

reports disagree, in part, because they meaningfully differ in the settings within which they observe the behaviors being assessed.

Recent studies using multiple research designs and samples demonstrate the ability of informant discrepancies to represent meaningful links with how children behave across and within contexts. An important implication of this work is that if informants observe behavior in different contexts, and these contextual differences account for discrepancies among informants' behavioral reports, then each informant's report may nonetheless be a valid report (e.g., De Los Reyes, Aldao et al., in press; De Los Reyes, Thomas et al., in press). In sum, recent research and theory indicates support for developing a Diverging Operations research concept for meaningfully interpreting discrepancies among multiple informants' reports.

Compensating Operations: When Informant Discrepancies Reflect Measurement Error

Defining Compensating Operations—To interpret informant discrepancies as reflecting Compensating Operations (Figure 1c), two conditions must be met. First, multiple informants' reports should yield discrepant empirical findings about the behavior being assessed. Second, these discrepancies should not reflect meaningful variation in the expression of the behavior being assessed. For example, the discrepancies may be due to methodological differences between two informant's otherwise reliable and valid reports. Alternatively, the discrepancies may be due to one of two informant's reports (or both informants' reports) yielding unreliable or invalid indices of the behavior being assessed. Thus, the Compensating Operations concept is similar to the Diverging Operations concept in that one observes informant discrepancies. Yet, Compensating Operations differs from Diverging Operations because it reflects instances within which evidence fails to indicate meaningful mechanisms to explain the informant discrepancies.

Observing patterns of diverging findings as reflections of Compensating Operations provides a researcher with the justification to approach informants' reports in one of two ways. First, if the informants provided reliable and valid data and the discrepancies were due to methodological differences between reports, a researcher may use the statistical and methodological approaches discussed previously (e.g., structural equations models and combinational algorithms) to examine the converging evidence among the informants' reports. Second, if the discrepancies were due to a subset of the informants providing unreliable or invalid reports, a researcher may discard these reports in favor of the informants' reports that yielded reliable and valid data. Importantly, findings that reflect Converging Operations may also justify management of multi-informant data using these methods. Indeed, convergence among findings may nonetheless result in non-trivial amounts of measurement error, prompting researchers to use methods to account for such error. As mentioned previously, current statistical and methodological approaches often assume that informant discrepancies reflect error. However, researchers ought to test these assumptions. Findings reflecting Compensating Operations (and Converging Operations) allow a researcher to treat informants' unique or unshared information as error. A key contribution of the OTM is that it guides researchers in testing whether informant discrepancies reflect measurement error.

Interpreting Multiple Informants' Reports Using Converging, Diverging, and Compensating Operations

As illustrated in Figure 1, Converging, Diverging, and Compensating Operations each yield distinct conceptualizations of empirical findings observed across multiple informants' reports. What empirically distinguishes these research concepts from each other? In Figure 2, we illustrate how to empirically test patterns of empirical findings among multiple informants' reports. Specifically, we organized Figure 2 into a "decision tree" consisting of a series of empirical questions. The outcomes of the empirical questions outlined in Figure 2 can assist in interpreting patterns of empirical findings, and ultimately in determining whether Converging, Diverging, or Compensating Operations best represents a given pattern of findings.

Specifically, interpreting multiple informants' reports using the OTM involves three steps. First, before data collection, one hypothesizes as to whether a study's findings from multiple informants' reports will yield converging research findings or diverging findings (Figure 2a). These hypotheses could be made based on conceptual models of convergence (e.g., De Los Reyes & Kazdin, 2006) and/or prior empirical estimates of multi-informant correspondence (e.g., Achenbach et al., 1987; Achenbach et al., 2005). Second, if the expectation is that the informants' reports will correspond or converge on a common conclusion, one tests for whether Converging Operations accounts for the study's findings (Figure 2b). Alternatively, if one expects that informants' reports will yield diverging findings and for meaningful reasons (e.g., contextual variability in behaviors being assessed), one tests whether Diverging Operations accounts for the study's findings (Figure 2c). Third, if neither Converging Operations nor Diverging Operations accounts for the study's findings, one conducts post-hoc tests of whether Compensating Operations explains the findings (Figure 2d).

In sum, the steps outlined in Figure 2 allow researchers to empirically delineate two sets of conditions. The first set characterizes conditions within which multiple informants' reports yield the same or similar research conclusions; Converging Operations accounts for these conditions. The second set characterizes conditions within which multiple informants' reports yield different research conclusions. For this second set of conditions, Diverging Operations and Compensating Operations can be used to characterize informant discrepancies as either reflecting meaningful (Diverging Operations) or mundane (Compensating Operations) information about the behaviors being assessed. Importantly, within any one study, a specific pattern of empirical findings among multiple informants' reports can be uniquely characterized as an instance of Converging, Diverging, or Compensating Operations.

Identifying Converging Operations—The key decision to be made regarding Converging Operations is to determine the a priori threshold that findings from multiple informants' reports must pass to be considered as "converging" on a common research conclusion (Figure 2a, b). Stated another way, a Converging Operations account of multiple informants' reports requires setting an operational definition or threshold for convergence. For instance, among a set of informants' outcome reports for a controlled trial of a treatment

for adulthood social phobia, how often must these reports indicate significant treatment improvement for the treatment condition, relative to a comparison condition (e.g., placebo control condition), for one to identify the informants' reports as collectively converging on a common conclusion (see also De Los Reyes & Kazdin, 2008)? Failing to meet this threshold for convergence would necessitate post-hoc tests to determine whether Compensating Operations accounts for the findings (Figure 2d). To be clear, one ought to avoid conducting post-hoc tests to probe for Diverging Operations because, similar to Converging Operations, hypotheses regarding Diverging Operations ought to be specified a priori.

A key issue regarding use of Converging Operations is identifying examples of operational definitions of convergence. Although rare, an example of an operational definition of Converging Operations can be found in recent conceptual work focused on interpreting the outcomes of randomized controlled trials. Specifically, as mentioned previously, controlled trials often yield inconsistent conclusions regarding the efficacy of treatments examined within these trials, particularly among multiple informants' outcome reports (De Los Reyes & Kazdin, 2006; Koenig et al., 2009). However, treatment studies might not all yield inconsistent findings (De Los Reyes & Kazdin, 2008). For example, a study might indicate positive effects of a treatment across multiple outcome measures, but another study of that same treatment might only indicate positive treatment benefits sporadically or across a few outcome measures (De Los Reyes & Kazdin, 2009). Thus, controlled trials researchers would benefit from a guide for distinguishing trials that yield converging versus diverging research conclusions across multiple informants' reports.

To this end, one operational definition of convergence applicable to Converging Operations can be found in the Range of Possible Changes (RPC) Model: A classification system for identifying consistencies and inconsistencies in research findings in controlled intervention studies (De Los Reyes & Kazdin, 2006). Specifically, the RPC Model includes categories that can serve as operational definitions of converging research findings within a single controlled trial. For instance, to classify a single study's findings, among the categories outlined in the RPC Model is one denoting those circumstances within which 80% of three or more outcome measures that vary by informant type, measurement method, and method of statistical evaluation reveal significant differences between treatment and control or comparison conditions. As an example, consider a study that assessed treatment response on 10 measures. Across these measures, parents, teachers, and clinicians provided outcome reports. These informants provided outcome reports within a multi-method assessment paradigm that included questionnaires, interviews, and laboratory observations. Further, researchers evaluated differences between treatment and control conditions using three methods of statistical evaluation (e.g., mean differences, diagnostic status, and clinically significant change). Using the threshold described previously, a researcher would operationally define converging findings as observing evidence of significant differences between treatment and control conditions on 8 of the 10 measures.

In sum, the RPC Model provides a categorical distinction of consistent research findings that can be used to delineate evidence supporting Converging Operations. Importantly, if one were to implement this categorical distinction as the threshold for classifying whether a study's findings reach convergence, any level of consistency which does not reach this

threshold would be identified as an incidence of informant discrepancies. If a researcher does not identify convergence, she or he would progress down the decision making process (Figure 2d) to testing whether the observed findings could be explained by Compensating Operations.

Identifying Diverging or Compensating Operations—As mentioned previously, the OTM includes decision-making rules for interpreting multiple informants' reports that indicate discrepant or inconsistent research conclusions (Figures 2a, c, d). Consistent with specifying a priori operational definitions of converging findings for tests of Converging Operations, for tests of Diverging Operations, one ought to specify a threshold by which one would identify diverging findings. Here, one sets a threshold with the expectation that the level of convergence in findings among multiple informants' reports will fall at or below the threshold. This can be contrasted with Converging Operations, in which the level of convergence among findings should meet or surpass the threshold.

For those circumstances within which a researcher hypothesizes and subsequently observes diverging research findings, it is crucial to address three questions (Figure 2c). First, will the evidence reveal that the informants provide reliable and valid reports? Here, it is important to set a priori thresholds for each of these psychometric properties of the informants' reports being examined. In the case of two informants' (e.g., parent and teacher) reports of a patient's anxiety, an example might include satisfying two criteria: (a) coefficient alpha estimates at or above 0.70 as evidence of internal consistency, and (b) significant correlations between the informants' anxiety reports and a third informant's report of a theoretically related construct as evidence of convergent validity (e.g., patient's self-reported depressive symptoms; see Nunnally & Bernstein, 1994). Here, if both informants' reports do not meet these thresholds, Compensating Operations likely account for the findings, and one would proceed to examining the questions regarding Compensating Operations outlined in Figure 2d.

If each informant's report does meet these thresholds, one addresses a second question: Will the evidence reveal that methodological factors cannot account for the discrepancies between informants' reports? This question should be addressed in advance of executing a study. Specifically, researchers should ensure that they will use parallel measures to gather multiple informants' reports about the same behavior. These measures should hold critical measurement factors constant (e.g., item content and response scaling; see De Los Reyes, in press). Importantly, for some clinical assessments (e.g., adult psychopathology), researchers and practitioners may encounter difficulties with holding measure methodology constant across informants' reports (e.g., patient and clinician reports). For example, a recent quantitative review of cross-informant correspondence in adult psychopathology assessments found less than 1% of studies contained sufficient information to assess correspondence (Achenbach et al., 2005). At minimum, researchers should ensure that the measures used to take informants' reports exhibit patterns of correspondence indicating that informant discrepancies arose as a result of different informants' perspectives. For instance, correlations between reports completed by the same informant should be significantly larger in magnitude than correlations between reports completed by different informants.

Provided that informants completed parallel measures, one addresses a third question to test whether informant discrepancies are best explained using Diverging Operations: Will the discrepancies empirically and conceptually relate to the assessed behavior? Importantly, when a researcher hypothesizes that informant discrepancies will yield meaningful information about the assessed behaviors, it is important that the researcher designs the study to test this hypothesis. That is, the researcher ought to administer measures in the study that would yield data to test whether informant discrepancies reveal meaningful information. If the data revealed that the discrepancies meaningfully informed their understanding of the assessed behavior, then the researcher identified evidence consistent with a Diverging Operations interpretation of the discrepancies. The researcher could subsequently use this evidence and the Diverging Operations concept to inform interpretations of their research findings.

In sum, affirmative (i.e., “yes”) responses to all three of the questions outlined in Figure 2c can be used to identify informant discrepancies that can be explained by Diverging Operations. Under the OTM, a “no” response to any of the three questions would prompt a researcher to test whether the informant discrepancies may be explained by Compensating Operations (Figure 2d). That is, any circumstance within which one observes (a) one or more informants providing unreliable or invalid reports or (b) methodological differences between the measures used by informants to provide behavioral reports would result in a researcher being justified in testing whether a Compensating Operations interpretation best explains their observed findings.

Illustration of the OTM

In advancing a framework to guide understanding patterns of findings from multiple informants’ reports, it is important to illustrate how to apply the framework. We will assume that our illustration supports a Diverging Operations interpretation of the research findings (Figure 2a, c). Specifically, consider a controlled trial in which an investigative team uses parent- and teacher-reported outcomes to test the effects of a parent training intervention developed to decrease oppositional behavior in young children. Prior work suggests discrepancies often arise between parents’ and teachers’ reports of young children’s oppositional behavior (e.g., Hinshaw, Han, Erhardt, & Huber, 1992). Further, a recent quantitative review of studies of this intervention indicates that the measures that most consistently support its efficacy are parent-reported outcomes (De Los Reyes & Kazdin, 2009). Thus, in addressing the question posed in Figure 2a, based on prior research, the investigative team might expect informant discrepancies that can be characterized by observing the intervention reduce oppositional behavior to a greater degree when based on parent- relative to teacher-reported outcomes.

The expectation by the investigative team that they will observe informant discrepancies in their research findings ultimately results in the team planning their study so as to adequately test the questions posed in Figure 2c. First, the researchers will seek to examine whether psychometric tests reveal that the discrepant informants’ reports nonetheless support the reliability and validity of the individual parent and teacher reports. Tests of these

psychometric properties might be informed by prior psychometric work on evidence-based assessments for childhood oppositional behavior (McMahon & Frick, 2005).

Second, consistent with the questions posed in Figure 2c, the investigative team would examine whether the discrepancies between parent and teacher reports meaningfully inform their understanding of the oppositional behavior observed in their sample. Here, the investigative team might first conceptualize how parent-teacher discrepancies in reports of children's oppositional behavior meaningfully reflect an important component of oppositional behavior. Along these lines, perhaps the investigative team would draw from theory suggesting that parents and teachers provide discrepant reports because they observe children in different contexts or settings (parents at home vs. teachers at school; Achenbach et al., 1987; Kraemer et al., 2003). As a result, the team might hypothesize that parent-teacher reporting discrepancies reflect that children behave differently in home and school settings, and thus administer a measure to test this hypothesis. An example would be an observational measure of young children's disruptive behavior that indicates the extent to which children express disruptive behavior when interacting with parental and non-parental adults (Wakschlag et al., 2008).

If at the end of the study the investigative team did observe parent-teacher discrepancies that reflected contextual variation in children's oppositional behavior, then the team would use this information to interpret their findings. For example, if the inconsistent findings revealed that the parent reports suggested the intervention was effective whereas the teacher reports did not, the team would conclude that their findings suggest that the intervention effectively targeted contingencies influencing behavior in home settings (e.g., dysfunctional parent-child interactions) and not school settings (e.g., child's reaction to teacher's classroom behavior management strategies). In sum, the OTM's value lies in the guidance provided to researchers on interpreting patterns of empirical findings derived from multiple informants' reports.

Recommendations for Future

Research Research and Theoretical Implications

Primary and secondary outcome measure methodology—The OTM has a number of implications for future research and theory. First, the OTM guides researchers toward advancing a priori hypotheses as to whether they expect findings based on multiple informants' reports to result in discrepant research conclusions. In posing hypotheses about discrepant research conclusions, the OTM encourages researchers to also conceptualize what these discrepancies represent. In contrast, current practices in controlled trials research involve designating multiple outcomes within a controlled trial in terms of their use as a "primary" versus "secondary" indicator of treatment response (De Los Reyes, Kundey et al., 2011). Importantly, we previously reviewed reasons why this strategy appears misaligned with the impetus behind taking multi-informant approaches to clinical assessment (Hunsley & Mash, 2007).

In our previous illustrations of the utility in adopting the OTM in research using multiple informants (Figure 2), we discussed circumstances within which the OTM improves use and

interpretation of multi-informant assessments. These improvements may occur, regardless of whether assessment outcomes represent Converging Operations, or in the case of informant discrepancies, meaningful variation in behavior (Diverging Operations) or not (Compensating Operations). In turn, the OTM may increase the cost-effectiveness of multi-informant assessments, and any clinical services informed by these assessments. For example, consider a study of a sample of adult patients in which patients' and spouses' reports taken during structured clinical interviews comprise the information used to diagnose social phobia. Here, the discrepant reports may signify that the informants vary on which contexts they base their reports of social phobia symptoms. That is, perhaps patients base their self-reports on behaviors expressed within work settings, whereas spouses base their reports about patients on behaviors expressed within social gatherings outside of the work setting? Using Diverging Operations, a research team has a guide for using multiple informants in diagnosis to distinguish between a diagnosis of social phobia (i.e., social phobia expressed across social interactions) and a subtype diagnosis typified by expression of symptoms within a specific social interaction (e.g., performance situations such as public speaking; see Bögels et al., 2010). Conversely, using a primary outcome strategy might lead to diagnoses that inform treatment plans that fail to represent, and thus optimally address, patients' presenting concerns. For example, a single report might suggest administering a comprehensive treatment for social phobia, when multiple informants' reports might indicate that a more focused treatment on symptoms expressed within performance settings would sufficiently reduce the patients' concerns. Notably, a key strength of the OTM is that it yields a multi-informant strategy that eliminates the need for researchers to address informant discrepancies by designating a primary measure in advance of the study.

Constructing multi-informant measurement batteries—The OTM also has implications for informing the construction of measurement protocols within studies that include multiple informants' reports. As mentioned previously, it is a well-accepted notion that upholding best practices in clinical assessment involves collecting reports from multiple informants (Hunsley & Mash, 2007; Mash & Hunsley, 2005). However, little guidance exists for researchers to decide on the number of informants on which to rely. Often the implicit assumption appears to be: Somewhere between two and as many as possible.

The OTM guides researchers toward hypothesizing whether they should expect informant discrepancies in research conclusions. Thus, in research areas within which investigative teams frequently expect to observe informant discrepancies, OTM-guided research might result in researchers constructing measurement batteries that include informants for whom reports disagree for meaningful reasons. For instance, use of the OTM to interpret commonly discrepant parent and teacher reports of children's disruptive behavior in developmental psychopathology research might result in the construction of a standardized battery of parent, teacher, and independent observer measures. Such a battery of measures might be similar to that used in work reviewed previously to assess disruptive behavior in young children (e.g., De Los Reyes et al., 2009; Wakschlag et al., 2008). Indeed, the purpose of this battery might involve developing a consensus assessment paradigm for systematically identifying cross-contextual versus contextspecific expressions of children's disruptive behavior. Therefore, the OTM might have the positive "side effect" of improving

researchers' precision in selecting, using, and interpreting informants' reports. We encourage researchers to use the OTM to both interpret informant discrepancies, and develop and refine comprehensive multi-informant assessment approaches.

Interpreting patterns of findings within investigations and meta-analyses—

Lastly, the OTM may encourage the development of operational definitions of converging findings among multiple informants' reports. Indeed, we previously illustrated the RPC Model, as used in controlled trials research, as one operational definition of convergence (De Los Reyes & Kazdin, 2006). However, researchers might seek to operationally define convergence at rates higher or lower than the definitions advanced by the RPC Model (De Los Reyes, Kundery et al., 2011). For example, a research area for which precision in outcome measurement is underdeveloped might be justified in setting a lower threshold for converging findings. In any event, we encourage future research to quantify the hypothesized levels of converging findings within specific literatures, and develop operational definitions of converging findings to match the particular methodological parameters of studies within these literatures.

In a similar vein, a long line of research across the social and general medical sciences attests to how often studies of the same or similar phenomena yield diverging findings (e.g., treatment efficacy, genetic and environmental risk factors; Achenbach, 2006, 2011; De Los Reyes & Kazdin, 2005, 2006, 2009; Goodman et al., 2010; Ioannidis, 2005; Ioannidis & Trikalinos, 2005; Monroe & Reid, 2008). It is often the case that this variation between studies is examined in relation to such factors as basic study characteristics (sample size and demographic characteristics) and methodological characteristics (informant and measure type). These factors are of particular interest in meta-analytic research (Rosenthal & DiMatteo, 2001).

Yet, meta-analytic researchers vary widely in how they manage multiple outcome measures within each of the studies in a sample. For example, researchers may select a single indicator to represent a study's effects or average the observed effects across outcome measures within the study; these practices often result in different conclusions (Matt, 1989). Importantly, each of these methods also treats any variability across measures as error. Use of the OTM may address this issue. That is, systematic reviews can improve upon their ability to synthesize research evidence if studies classify their findings according to whether they reflect Converging, Diverging, or Compensating Operations. Stated another way, if the OTM leads to more reliable and valid conclusions drawn from within multiple studies, this may, in turn, improve the reliability and validity of meta-analyses, and in particular conclusions drawn from between-study comparisons (i.e., comparisons of studies that used the OTM to interpret research conclusions).

Limitations of the OTM

There are limitations to the OTM. First, as mentioned previously, using and interpreting multiple informants' reports of a behavior may involve conducting multiple statistical tests on these reports. Under the general linear model (GLM), researchers may be hesitant to conduct multiple tests, as statistical corrections for conducting multiple tests may reduce statistical power to detect significant effects (Tabachnick & Fidell, 2001). Of course, these

statistical corrections are typically made to correct for multiple independent statistical tests, or tests that assume that multiple tests yield multiple independent observations. However, although multiple informants' reports typically correspond no higher than low-to-moderate in magnitude (e.g., r 's ranging from 0.20 to 0.40; see Achenbach et al., 1987; Achenbach et al., 2005), they nonetheless often correspond with each other at statistically significant magnitudes. Thus, use of the OTM may result in researchers examining multi-informant assessment outcomes using methods such as (a) extensions of the GLM that account for correlated observations among dependent variables (e.g., generalized estimating equations; see Hanley, Negassa, Edwardes, & Forrester, 2003) or (b) Bayesian modeling approaches (e.g., person-centered latent class models; see Bartholomew, Steele, Moustaki, & Galbraith, 2002). Indeed, previous work indicates that these approaches can be applied to testing questions about multi-informant assessments (e.g., De Los Reyes, Thomas et al., in press; Chilcoat & Breslau, 1997; Goodman et al., 2010; Horton & Fitzmaurice, 2004; Mick, Santangelo, Wypij, & Biederman, 2000; Nomura & Chemtob, 2009). In fact, two of the studies reviewed previously that provided supportive evidence for the Diverging Operations concept used these approaches (De Los Reyes, Ehrlich et al., in press; De Los Reyes et al., 2009). Therefore, we encourage future research on the OTM that applies quantitative methods for understanding variation among multiple informants' reports.

Second, use of the OTM within a given research area requires a guide within that area for operationally defining converging research findings. With the possible exception of the RPC Model (De Los Reyes & Kazdin, 2006); the extant literature lacks guidelines for operationally defining convergence. This is an important limitation in that, as mentioned previously, convergence and the thresholds used to define it may vary greatly depending on the research area. New operational definitions of convergence might come from the development of conceptual models that define convergence within specific research areas. Further, empirically based estimates of convergence might be constructed based on quantitative reviews of multiinformant correspondence (e.g., Achenbach et al., 1987; Achenbach et al., 2005). In any event, we encourage research and theory that seeks to identify alternative definitions of convergence.

Third, much of the evidence reviewed previously that supports Diverging Operations comes from research on multi-informant assessments of children's externalizing behaviors. This stems from two aspects of assessments of externalizing behavior. The first is that the informants used in these assessments are typically adult informants who vary in the contexts within which they observe children's externalizing behaviors (e.g., parents vs. teachers; Dirks et al., 2012). The second is that a key factor that differentiates these informants (i.e., the context within which they observe behavior) can be readily tested with existing observational paradigms (Wakschlag et al., 2008). Further, assessments in other patient populations may only infrequently incorporate multi-informant assessments (e.g., adult patients; see Achenbach et al., 2005).

Psychological assessment research is ripe with opportunities to extend the evidence supporting the ability to use Diverging Operations in multi-informant assessments. For instance, adult patient self-reports of psychopathology and clinician reports about adult patients correspond in the low-to-moderate range (Achenbach et al., 2005). Further, recent

longitudinal work indicates that discrepancies between informants' reports exist across the lifespan of patients into middle age (van der Ende, Verhulst, & Tiemier, 2012). Yet, research on multi-informant correspondence in adult psychopathology assessments has largely focused on the implications of low levels of correspondence for identifying evidence-based interventions and defining treatment response (e.g., Frank et al., 1991; Lambert, Hatch, Kingston, & Edwards, 1986; Ogles, Lambert, Weight, & Payne, 1990; Zimmerman et al., 2006). In fact, this work likely played a key role in researchers transitioning away from identifying treatment response using multi-informant assessments in favor of primary outcome measures (De Los Reyes, Kundery et al., 2011).

However, there may be circumstances within which informant discrepancies usefully inform diagnosis and treatment response methods in adult psychopathology assessments. As an example, consider diagnostic assessments of adulthood social phobia. Informant discrepancies may inform interpretations of assessments of adulthood social phobia in two ways.

First, discrepancies between patient and clinician reports might reflect subtype diagnoses in adulthood social phobia assessments. Specifically, as mentioned previously, patients can be empirically distinguished based on symptom expression and impairment evident across multiple settings versus symptom expression and impairment evident only within performance settings (e.g., public speaking; Bögels et al., 2010). Further, patients' self-reports of internalizing difficulties often yield low-to-moderate levels of correspondence with clinicians' reports about patients' internalizing difficulties (Achenbach et al., 2005). Patients and clinicians likely vary in the contexts on which they base their social phobia reports. For instance, patient self-reports may reflect social phobia symptoms expressed in home and work settings. In contrast, clinician reports about patients may reflect social phobia based on patient self-reports, as well as their clinical observations of the patient. Clinician reports may be meaningfully distinguished from patient reports in that they incorporate symptom expressions using information from a setting (i.e., the clinical setting) that patients likely do not incorporate into their own self-reports. At the same time, informant pairs may vary in how much their reports correspond with other informants' reports (e.g., De Los Reyes, Alfano et al., 2011). That is, some patients' reports do not correspond with other informants' reports, and some patients' reports correspond to a considerable extent with other informants' reports (De Los Reyes, Youngstrom et al., 2011a, b). Thus, circumstances within which patient and clinician reports correspond on questionnaire measures of social phobia may indicate a strong likelihood of a social phobia diagnosis endorsed on a structured interview (i.e., symptom expression and impairment evidence across multiple settings). Conversely, low correspondence between reports may indicate a strong likelihood of a diagnosis specific to symptom expression and impairment within performance settings. Therefore, the lack of correspondence between patient and clinician reports could serve as an external indicator for corroborating subtype diagnoses of social phobia.

Second, a number of standardized behavioral tasks are available to assess associated impairments germane to social phobia, such as tasks that assess social skills deficits (e.g., Beidel, Rao, Scharfstein, Wong, & Alfano, 2010; Beidel, Turner, Stanley, & Dancu, 1989;

Heimberg, Hope, Dodge, & Becker, 1990; Curran, 1982; Turner, Beidel, Cooley, Woody, & Messer, 1994; Turner, Beidel, Dancu, & Keys, 1986). Further, similar to cross-contextual assessments of children's disruptive behavior (Wakschlag et al., 2008), behavioral assessments of adulthood social skills deficits vary in the contexts within which they assess behaviors indicative of social skills deficits (e.g., public speaking and one-on-one social interactions). Therefore, patterns of consistency and inconsistency in social skills deficits exhibited across these tasks could also be used to corroborate the utility of discrepancies between patient and clinician reports as indicators of subtype diagnoses in adulthood social phobia assessments. In sum, we encourage (a) future research on factors other than context that meaningfully explain informant discrepancies and (b) hypothesis-driven research seeking to understand whether meaningful variation exists in discrepant reports of behaviors other than childhood externalizing behaviors.

Conclusions

One of the greatest challenges facing clinical science involves drawing proper research conclusions when multiple informants' reports about clinically meaningful behaviors disagree. Exponentially increasing these challenges is that the current paradigm for interpreting research findings leads researchers to drawing indications of the strength of evidence contingent upon the convergence of research findings. Yet, few studies observe consistent findings among informants' reports. Thus, the current paradigm (i.e., Converging Operations) does not allow researchers to draw strong inferences from inconsistent findings.

At the beginning of this paper, we argued that these realities of assessment in clinical science have resulted in what we termed the Grand Discrepancy in multi-informant assessment. Specifically, researchers have an abstract notion that multi-informant assessments capitalize on the unique perspectives of informants and their observations of clinically relevant behaviors. However, observations of discrepant assessment outcomes within multi-informant assessments often prompt researchers to use statistical and methodological techniques that treat these informant discrepancies as measurement error. Importantly, we reviewed a variety of circumstances within which informant discrepancies reveal meaningful divergence or variability in clinically relevant behaviors (e.g., across contexts within which such behaviors may be expressed). Under these circumstances, using techniques that treat discrepancies as error undermines the principles underlying multi-informant approaches to clinical assessment. Therefore, researchers require guides for interpreting patterns of both converging and diverging findings from multiple informants' reports.

To address the Grand Discrepancy in multi-informant assessment, we advanced a guiding framework that researchers can use to systematically investigate the convergence or divergence of research findings derived from multi-informant assessments. This framework may result in improvements in the veracity of research conclusions drawn from multi-informant assessments. For instance, this framework may encourage judicious applications of techniques that treat informant discrepancies as measurement error, and guide researchers toward meaningful interpretations of research findings when informants' reports meaningfully diverge. However, these improvements will only come if researchers draw

from the informant discrepancies literature—a literature spanning multiple decades—to form hypotheses about whether findings based on multiple informants' reports will converge or diverge. If researchers hypothesize divergence, they can subsequently form and test hypotheses about the reasons for this divergence. We encourage clinical researchers to incorporate expectations of informant discrepancies and the reasons why they might occur when making decisions about hypothesis testing, study design, and data interpretation.

Acknowledgments

We thank Susan Nolen-Hoeksema, the Editorial Board of the *Annual Review of Clinical Psychology*, Carl W. Lejuez, Matthew K. Nock, Thomas S. Wallsten, and Eric A. Youngstrom for their helpful and insightful comments on and discussions about previous versions of this paper. We also thank Michael Van Wie and Erika Roberge for their assistance with literature reviews conducted in preparation of this manuscript.

Definitions Used in Text

Informant Discrepancies	Inconsistencies among two or more informants' reports of a behavior as provided on parallel measures of that behavior.
Grand Discrepancy	The inconsistency between one's rationale for assessing behavior using multiple informants' reports and the methods used to manage informant discrepancies.
Converging Operations	Set of measurement conditions for interpreting patterns of findings based on the consistency within which findings yield similar conclusions.
Diverging Operations	Set of measurement conditions for interpreting patterns of inconsistent findings based on hypotheses about variations in the behavior(s) assessed.
Compensating Operations	Set of measurement conditions for interpreting patterns of inconsistent findings based on methodological features of the study's measures or informants.
Operations Triad Model	Theoretical framework that guides hypothesis testing and study design in empirical work taking multi-informant approaches to assessing behavior.
Attribution Bias Context Model	Theoretical framework that explains why informant discrepancies occur in clinical assessments of children and adolescents.
Range of Possible Changes Model	Theoretical framework that provides operational definitions of converging and diverging research findings.

Acronyms Used in Text

ABC	Attribution Bias Context
OTM	Operations Triad Model

RPC Range of Possible Changes

Literature Cited

- Achenbach TM. As others see us: Clinical and research implications of cross-informant correlations for psychopathology. *Curr. Dir. Psychol. Sci.* 2006; 15:94–98.
- Achenbach TM. Definitely more than measurement error: But how should we understand and deal with informant discrepancies? *J. Clin. Child Adolesc. Psychol.* 2011; 40:80–86. [PubMed: 21229445]
- Achenbach TM, Krukowski RA, Dumenci L, Ivanova MY. Assessment of adult psychopathology: Meta-analyses and implications of cross-informant correlations. *Psychol. Bull.* 2005; 131:361–382. [PubMed: 15869333] ■ Informants' adult psychopathology reports correspond at magnitudes similar to informants' child and adolescent psychopathology reports.
- Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychol. Bull.* 1987; 101:213–232. [PubMed: 3562706] ■ Seminal quantitative review documenting low-to-moderate correspondence among multiple informants' reports of child and adolescent psychopathology.
- Am. Psychiatr. Assoc. Diagnostic and Statistical Manual of Mental Disorders-Text Revision. 4th ed.. Washington, DC: Am. Psychiatr. Assoc.; 2001. (DSM-IV-TR).
- Arseneault L, Moffitt T, Caspi A, Taylor A, Rijdsdijk F, Jaffee S, et al. Strong genetic effects on cross-situational antisocial behaviour among 5-year-old children according to mothers, teachers, examiner-observers, and twins' self-reports. *J. Child Psychol. Psychiatry.* 2003; 44:832–848. [PubMed: 12959492]
- Baker L, Jacobson K, Raine A, Lozano D, Bezdjian S. Genetic and environmental bases of childhood antisocial behavior: A multi-informant twin study. *J. Abnorm. Psychol.* 2007; 116:219–235. [PubMed: 17516756]
- Bartels M, Boomsma D, Hudziak J, van Beijsterveldt T, van den Oord E. Twins and the study of rater (dis)agreement. *Psychol. Methods.* 2007; 12:451–466. [PubMed: 18179355]
- Bartholomew, DJ.; Steele, F.; Moustaki, I.; Galbraith, JI. *The Analysis and Interpretation of Multivariate Data for Social Scientists.* Philadelphia: Chapman & Hall/CRC; 2002.
- Beidel DC, Rao PA, Scharfstein L, Wong N, Alfano CA. Social skills and social phobia: An investigation of DSM-IV subtypes. *Behav. Res. Ther.* 2010; 48:992–1001. [PubMed: 20637452]
- Beidel DC, Turner SM, Stanley MA, Dancu CV. The Social Phobia and Anxiety Inventory: Concurrent and external validity. *Behav. Ther.* 1989; 20:417–427.
- Bögels SM, Alden L, Beidel DC, Clark LA, Pine DS, Stein MB, et al. Social anxiety disorder: Questions and answers for the DSM-V. *Depress. Anxiety.* 2010; 27:168–189. [PubMed: 20143427]
- Borsboom, D. *Measuring the Mind.* New York: Cambridge University Press; 2005.
- Borsboom D, Mellenbergh GJ, van Heerden J. The concept of validity. *Psychol. Rev.* 2004; 111:1061–1071. [PubMed: 15482073]
- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* 1959; 56:81–105. [PubMed: 13634291]
- Chilcoat HD, Breslau N. Does psychiatric history bias mothers' reports? An application of a new analytic approach. *J. Am. Acad. Child Adolesc. Psychiatry.* 1997; 36:971–979. [PubMed: 9204676]
- Curran, JP. A procedure for the assessment of social skills: The simulated social interaction test. In: Curran, JP.; Monti, PM., editors. *Social Skills Training: A Practical Handbook for Assessment and Treatment.* New York, NY: Guilford; 1982. p. 348-373.
- DeBello MP, Kowatch RA, Adler CM, Stanford KE, Welge JA, Barzman DH, et al. A double-blind randomized pilot study comparing quetiapine and divalproex for adolescent mania. *J. Am. Acad. Child Adolesc. Psychiatry.* 2006; 45:305–313. [PubMed: 16540815]

- De Los Reyes A. Strategic objectives for improving understanding of informant discrepancies in developmental psychopathology research. *Dev. Psychopathol.* in press.
- De Los Reyes A. Introduction to the special section. More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *J. Clin. Child Adolesc. Psychol.* 2011; 40:1–9. [PubMed: 21229439] ■ Outlines a basis for meaningfully interpreting informant discrepancies in clinical child and adolescent assessments.
- De Los Reyes A, Aldao A, Thomas SA, Daruwala SE, Swan AJ, Van Wie M, et al. Adolescent self-reports of social anxiety: Can they disagree with objective psychophysiological measures and still be valid? *J. Psychopathol. Behav. Assess.* in press.
- De Los Reyes A, Alfano CA, Beidel DC. The relations among measurements of informant discrepancies within a multisite trial of treatments for childhood social phobia. *J. Abnorm. Child Psychol.* 2010; 38:395–404. [PubMed: 20013046]
- De Los Reyes A, Alfano CA, Beidel DC. Are clinicians' assessments of improvements in children's functioning "global"? *J. Clin. Child Adolesc. Psychol.* 2011; 40:281–294. [PubMed: 21391024]
- De Los Reyes A, Ehrlich KB, Swan AJ, Luo T, Van Wie M, Pabón SC. An experimental test of whether informants can report about child and family behavior based on settings of behavioral expression. *J. Child Fam. Stud.* in press. ■ Experiment documenting that informants can provide behavioral reports based on settings within which behaviors occur.
- De Los Reyes A, Henry DB, Tolan PH, Wakschlag LS. Linking informant discrepancies to observed variations in young children's disruptive behavior. *J. Abnorm. Child Psychol.* 2009; 37:637–652. [PubMed: 19247829] ■ Informant discrepancies in behavioral reports correspond to contextual variations in controlled laboratory observations of behavior.
- De Los Reyes A, Kazdin AE. Measuring informant discrepancies in clinical child research. *Psychol. Assess.* 2004; 16:330–334. [PubMed: 15456389]
- De Los Reyes A, Kazdin AE. Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychol. Bull.* 2005; 131:483–509. [PubMed: 16060799] ■ Presents framework to explain why informant discrepancies occur in clinical assessments of children and adolescents.
- De Los Reyes A, Kazdin AE. Conceptualizing changes in behavior in intervention research: The range of possible changes model. *Psychol. Rev.* 2006; 113:554–583. [PubMed: 16802881] ■ Provides operational definitions of converging and diverging findings from controlled studies of treatment efficacy.
- De Los Reyes A, Kazdin AE. When the evidence says, "Yes, no, and maybe so": Attending to and interpreting inconsistent findings among evidence-based interventions. *Curr. Dir. Psychol. Sci.* 2008; 17:47–51. [PubMed: 21243087]
- De Los Reyes A, Kazdin AE. Identifying evidence-based interventions for children and adolescents using the range of possible changes model: A meta-analytic illustration. *Behav. Modif.* 2009; 33:583–617. [PubMed: 19684328]
- De Los Reyes A, Kundey SMA, Wang M. The end of the primary outcome measure: A research agenda for constructing its replacement. *Clin. Psychol. Rev.* 2011; 31:829–838. [PubMed: 21545781]
- De Los Reyes A, Marsh JK. Patients' contexts and their effects on clinicians' impressions of conduct disorder symptoms. *J. Clin. Child Adolesc. Psychol.* 2011; 40:479–485. [PubMed: 21534058]
- De Los Reyes A, Thomas SA, Swan AJ, Ehrlich KB, Reynolds EK, Suarez L, et al. "It depends on what you mean by 'disagree'": Differences between parent and child perceptions of parent-child conflict. *J. Psychopathol. Behav. Assess.* in press.
- De Los Reyes A, Youngstrom EA, Pabón SC, Youngstrom JK, Feeny NC, Findling RL. Internal consistency and associated characteristics of informant discrepancies in clinic referred youths age 11 to 17 years. *J. Clin. Child Adolesc. Psychol.* 2011a; 40:36–53. [PubMed: 21229442]
- De Los Reyes A, Youngstrom EA, Swan AJ, Youngstrom JK, Feeny NC, Findling RL. Informant discrepancies in clinical reports of youths and interviewers' impressions of the reliability of informants. *J. Child Adolesc. Psychopharmacol.* 2011b; 21:417–424. [PubMed: 22040187]

- Derks E, Hudziak J, Van Beijsterveldt C, Dolan C, Boomsma D. Genetic analyses of maternal and teacher ratings on attention problems in 7-year-old Dutch twins. *Behav. Genet.* 2006; 36:833–844. [PubMed: 16773450]
- Dirks MA, De Los Reyes A, Briggs-Gowan MJ, Cella D, Wakschlag LS. Embracing not erasing contextual variability in children's behavior: Theory and utility in the selection and use of measures and informants in the assessment of developmental psychopathology. *J. Child Psychol. Psychiatry.* 2012; 53:558–574. [PubMed: 22360546]
- Dumenci L, Achenbach TM, Windle M. Measuring context-specific and cross-contextual components of hierarchical constructs. *J. Psychopathol. Behavi. Assess.* 2011; 33:3–10.
- Durston S. Converging methods in studying attention-deficit/hyperactivity disorder: What can we learn from neuroimaging and genetics? *Dev. and Psychopathol.* 2008; 20:1133–1143.
- Edgeworth FY. The statistics of examinations. *J. R. Stat. Soc.* 1888; 51:598–635.
- Frank E, Prien RF, Jarrett RB, Keller MB, Kupfer DJ, Lavori PW, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder. *Arch. Gen. Psychiatry.* 1991; 48:851–855. [PubMed: 1929776]
- Garner WR, Hake HW, Eriksen CW. Operationism and the concept of perception. *Psychol. Rev.* 1956; 63:149–159. [PubMed: 13323170] ■ Foundational paper in psychological science of the concept of Converging Operations.
- Gazzaniga, MS.; Ivry, RB.; Mangun, GR. *Cognitive Neuroscience: The Biology of the Mind.* 3rd ed.. New York: W. W. Norton & Company; 2008.
- Gizer I, Waldman I, Abramowitz A, Barr C, Feng Y, Wigg K, et al. Relations between multi-informant assessments of ADHD symptoms, DAT1, and DRD4. *J. Abnorm. Psychol.* 2008; 117:869–880. [PubMed: 19025233]
- Goodman KL, De Los Reyes A, Bradshaw CP. Understanding and using informants' reporting discrepancies of youth victimization: A conceptual model and recommendations for research. *Clin. Child Fam. Psychol. Rev.* 2010; 13:366–383. [PubMed: 20799062]
- Goodman S, Lahey B, Fielding B, Dulcan M, Narrow W, Regier D. Representativeness of clinical samples of youths with mental disorders: A preliminary population-based study. *J. Abnorm. Psychol.* 1997; 106:3–14. [PubMed: 9103713]
- Hanley JA, Negassa A, Edwardes DB, Forrester JE. Statistical analysis of correlated data using generalized estimating equations: An orientation. *Am. J. Epidemiol.* 2003; 157:364–375. [PubMed: 12578807]
- Hartley AG, Zakriski AL, Wright JC. Probing the depths of informant discrepancies: Contextual influences on divergence and convergence. *J. Clin. Child Adolesc. Psychol.* 2011; 40:54–66. [PubMed: 21229443]
- Heimberg RG, Hope DA, Dodge CS, Becker RE. DSM-III-R subtypes of social phobia: Comparison of generalized social phobics and public speaking phobics. *J. Nerv. Ment. Dis.* 1990; 178:172–179. [PubMed: 2307969]
- Hempel, CG. *Philosophy of Natural Science.* Englewood Cliffs, NJ: Prentice-Hall; 1966.
- Hinshaw SP, Han SS, Erhardt D, Huber A. Internalizing and externalizing behavior problems in preschool children: Correspondence among parent and teacher ratings and behavior observations. *J. Clin. Child Psychol.* 1992; 21:143–150.
- Holmbeck G, Li S, Schurman J, Friedman D, Coakley R. Collecting and managing multisource and multimethod data in studies of pediatric populations. *J. Pediatr. Psychol.* 2002; 27:5–18. [PubMed: 11726675]
- Horton JP, Fitzmaurice GM. Regression analysis of multiple source and multiple informant data from complex survey samples. *Stat. Med.* 2004; 23:2911–2933. [PubMed: 15344194]
- Hunsley, J.; Lee, CM. *Introduction to Clinical Psychology.* Hoboken: Wiley & Sons; 2010.
- Hunsley J, Mash EJ. Evidence-based assessment. *Annu. Rev. Clin. Psychol.* 2007; 3:29–51. [PubMed: 17716047] ■ Authoritative paper on the rationale and concepts underlying evidence-based assessment in clinical psychology.
- Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *J. Amer. Med. Assoc.* 2005; 294:218–228.

- Ioannidis JPA, Trikalinos TA. Early extreme contradictory estimates may appear in published research: The proteus phenomenon in molecular genetics research and randomized trials. *J. Clin. Epidemiol.* 2005; 58:543–549. [PubMed: 15878467]
- Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. *Psychol. Bull.* 1993; 114:3–28. [PubMed: 8346328]
- Klein DN, Dougherty LR, Olino TM. Toward guidelines for evidence-based assessment of depression in children and adolescent. *J. Clin. Child Adolesc. Psychol.* 2005; 34:412–432. [PubMed: 16026212]
- Klinberg T, Fernell E, Olesen PJ, Johnson M, Gustafsson P, Dahlström K, et al. Computerized training of working memory in children with ADHD: A randomized, controlled trial. *J. Am. Acad. Child Adolesc. Psychiatry.* 2005; 44:177–186. [PubMed: 15689731]
- Klonsky ED, Oltmanns TF, Turkheimer E. Informant reports of personality disorder: Relation to self-reports and future research directions. *Clin. Psychol. Sci. Pract.* 2002; 9:300–311.
- Koenig K, De Los Reyes A, Cicchetti D, Scahill L, Klin A. Group intervention to promote social skills in school-age children with pervasive developmental disorders: Reconsidering efficacy. *J. Autism Dev. Disord.* 2009; 39:1163–1172. [PubMed: 19326199]
- Kraemer HC, Measelle JR, Ablow JC, Essex MJ, Boyce WT, Kupfer DJ. A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *Am. J. Psychiatry.* 2003; 160:1566–1577. [PubMed: 12944328]
- Lambert MJ, Hatch DR, Kingston MD, Edwards BC. Zung, Beck, and Hamilton Rating Scales as measures of treatment outcome: A meta-analytic comparison. *J. Consult. Clin. Psychol.* 1986; 54:54–59. [PubMed: 3958302]
- Lofthouse N, Fristad M, Splaingard M, Kelleher K. Parent and child reports of sleep problems associated with early-onset bipolar spectrum disorders. *J. Fam. Psychol.* 2007; 21:114–123. [PubMed: 17371116]
- McFall RM, Townsend JT. Foundations of psychological assessment: Implications for cognitive assessment in clinical science. *Psychol. Assess.* 1998; 10:316–330.
- McMahon RJ, Frick PJ. Evidence-based assessment of conduct problems in children and adolescents. *J. Clin. Child Adolesc. Psychol.* 2005; 34:477–505. [PubMed: 16026215]
- Malle BF. The actor-observer asymmetry in attribution: A (surprising) meta-analysis. *Psychol. Bull.* 2006; 132:895–919. [PubMed: 17073526]
- Mash, EJ.; Hunsley, J., editors. *J. Clin. Child Adolesc. Psychol.* 2005. Developing guidelines for the evidence-based assessment of child and adolescent disorders [Special issue]; p. 34
- Matt GE. Decision rules for selecting effect sizes in meta-analysis: A review and reanalysis of psychotherapy outcome studies. *Psychol. Bull.* 1989; 105:106–115. [PubMed: 2648436]
- Mick E, Santangelo SL, Wypij D, Biederman J. Impact of maternal depression on ratings of comorbid depression in adolescents with attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry.* 2000; 39:314–319. [PubMed: 10714051]
- Mischel W, Shoda Y. A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychol. Rev.* 1995; 102:246–268. [PubMed: 7740090]
- Monroe SM, Reid MW. Gene-environment interactions in depression research. *Psychol. Sci.* 2008; 19:947–956. [PubMed: 19000200]
- Nomura Y, Chemtob CM. Effect of maternal psychopathology on behavioral problems in preschool children exposed to terrorism: Use of generalized estimating equations to integrate multiple informant reports. *Arch. Pediatr. Adolesc. Med.* 2009; 163:531–539. [PubMed: 19487609]
- Nunnally, JC.; Bernstein, IH. *Psychometric Theory*. 3rd ed.. New York: McGraw-Hill; 1994.
- Offord DR, Boyle MH, Racine Y, Szatmari P, Fleming JE, Sanford M, et al. Integrating assessment data from multiple informants. *J. Am. Acad. Child Adolesc. Psychiatry.* 1996; 35:1078–1085. [PubMed: 8755805]
- Ogles BM, Lambert MJ, Weight DG, Payne IR. Agoraphobia outcome measurement: A review and meta-analysis. *Psychol. Assess.* 1990; 2:317–325.

- Oltmanns TF, Turkheimer E. Person perception and personality pathology. *Curr. Dir. Psychol. Sci.* 2009; 18:32–36. [PubMed: 20539833] ■ Reviews research supporting potential uses of informant discrepancies when interpreting reports of adult personality pathology.
- Papakostas GI, Mischoulon D, Shyu I, Alpert JE, Fava M. S-Adenosyl methionine (SAME) augmentation of serotonin reuptake inhibitors for antidepressant nonresponders with major depressive disorder: A double-blind, randomized clinical trial. *Am. J. Psychiatry.* 2010; 167:942–948. [PubMed: 20595412]
- Pasupathi M. The social construction of the personal past and its implications for adult development. *Psychol. Bull.* 2001; 127:651–672. [PubMed: 11548972]
- Pelham WE, Fabiano GA, Massetti GM. Evidence-based assessment of attention deficit hyperactivity disorder in children and adolescents. *J. Clin. Child Adolesc. Psychol.* 2005; 34:449–476. [PubMed: 16026214]
- Pettinati HM, Oslin DW, Kampman KM, Dundon WD, Xie H, Gallis TL, et al. A double-blind, placebo-controlled trial combining sertraline and naltrexone for treating co-occurring depression and alcohol dependence. *Arch. Gen. Psychiatry.* 2010; 167:668–675.
- Piacentini J, Cohen P, Cohen J. Combining discrepant diagnostic information from multiple sources: Are complex algorithms better than simple ones? *J. Abnorm. Child Psychol.* 1992; 20:51–63. [PubMed: 1548394]
- Pryor JB, Ostrom TM. The cognitive organization of social information: A converging-operations approach. *J. Person. Soc. Psychol.* 1981; 41:628–641.
- Roberts BW, Caspi A. Authors' response: Personality development and the person-situation debate: It's déjà vu all over again. *Psychol. Inq.* 2001; 12:104–109.
- Rosenthal R, DiMatteo MR. Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annu. Rev. Psychol.* 2001; 52:59–82. [PubMed: 11148299]
- Silverman WK, Ollendick TH. Evidence-based assessment of anxiety and its disorders in children and adolescents. *J. Clin. Child Adolesc. Psychol.* 2005; 34:380–411. [PubMed: 16026211]
- Skinner, BF. *Science and Human Behavior*. New York: MacMillan; 1953.
- Sood ED, Pendley JS, Delamater AM, Rohan JM, Pulgaron ER, Drotar D. Mother-father informant discrepancies regarding diabetes management: Associations with diabetes-specific family conflict and glycemic control. *Health Psychol.* in press.
- Tabachnick, B.; Fidell, LS. *Using Multivariate Statistics*. 4th ed.. Boston: Allyn and Bacon; 2001.
- Tackett J, Waldman I, Lahey B. Etiology and measurement of relational aggression: A multi-informant behavior genetic investigation. *J. Abnorm. Psychol.* 2009; 118:722–733. [PubMed: 19899842]
- Thomas C, Turkheimer E, Oltmanns TF. Factorial structure of pathological personality traits as evaluated by peers. *J. Abnorm. Psychol.* 2003; 112:1–12.
- Thomas SA, Aldao A, De Los Reyes A. Implementing clinically feasible psychophysiological measures in evidence-based assessments of adolescent social anxiety. *Prof. Psychol. Res. Pract.* in press.
- Thurstone C, Riggs PD, Salomonsen-Sautel S, Mikulich-Gilbertson SK. Randomized, controlled trial of atomoxetine for attention-deficit/hyperactivity disorder in adolescents with substance use disorder. *J. Am. Acad. Child Adolesc. Psychiatry.* 2010; 49:573–582. [PubMed: 20494267]
- Turner SM, Beidel DC, Cooley MR, Woody S, Messer SC. A multicomponent behavioral treatment for social phobia: Social Effectiveness Therapy. *Behav. Res. Ther.* 1994; 32:381–390. [PubMed: 8192637]
- Turner SM, Beidel DC, Dancu CV, Keys D. Psychopathology of social phobia and comparison to avoidant personality disorder. *J. Abnorm. Psychol.* 1986; 95:389–394. [PubMed: 3805504]
- Valentine JC, Biglan A, Boruch RF, Castro FG, Collins LM, Flay BR, et al. Replication in prevention science. *Prev. Sci.* 2011; 12:103–117. [PubMed: 21541692]
- Valk J, van den Oord E, Verhulst F, Boomsma D. Using parental ratings to study the etiology of 3-year-old twins' problem behaviors: Different views or rater bias? *J. Child Psychol. Psychiatry.* 2001; 42:921–931. [PubMed: 11693587]
- van der Ende J, Verhulst FC, Tiemeier H. Agreement of informants on emotional and behavioral problems from childhood to adulthood. *Psychol. Assess.* 2012; 24:293–300. [PubMed: 21928910]

- Wakschlag LS, Hill C, Carter AS, Danis B, Egger HL, Keenan K, et al. Observational assessment of preschool disruptive behavior. Part I. Reliability of the Disruptive Behavior Diagnostic Observation Schedule (DB-DOS). *J. Am. Acad. Child Adolesc. Psychiatry.* 2008; 47:622–631. [PubMed: 18434926]
- Weisz JR, Jensen Doss A, Hawley KM. Youth psychotherapy outcome research: A review and critique of the evidence base. *Annu. Rev. Psychol.* 2005; 56:337–363. [PubMed: 15709939]
- Youngstrom E, Findling R, Calabrese J. Who are the comorbid adolescents? Agreement between psychiatric diagnosis, youth, parent, and teacher report. *J. Abnorm. Child Psychol.* 2003; 31:231–245. [PubMed: 12774858]
- Zhou Q, Lengua L, Wang Y. The relations of temperament reactivity and effortful control to children's adjustment problems in China and the United States. *Dev. Psychol.* 2009; 45:724–739. [PubMed: 19413428]
- Zimmerman M, McGlinchey JB, Posternak MA, Friedman M, Attiullah N, Boerescu D. How should remission from depression be defined? The depressed patient's perspective. *Am. J. Psychiatry.* 2006; 163:148–150. [PubMed: 16390903]

Summary of Central Points

- Clinical assessments of patients across the lifespan (e.g., children, adolescents, adults, and elderly) often incorporate reports from multiple informants.
- Researchers often justify taking multi-informant approaches to clinical assessment based on the idea that informants each have unique and valid perspectives when observing the behaviors about which they provide reports.
- When researchers observe informant discrepancies, they often address them using methodological and statistical techniques that assume discrepancies reflect measurement error.
- Recent research and theory in clinical assessment indicates that informant discrepancies often cannot be interpreted as reflections of measurement error.
- Researchers' interpretations of informant discrepancies as measurement error stem from Converging Operations: A set of measurement conditions for interpreting patterns of findings based on the consistency within which findings yield similar conclusions.
- Difficulties with interpreting informant discrepancies stem from a lack of options for interpreting such discrepancies other than Converging Operations.
- This paper advances a framework (Operations Triad Model) that provides researchers with guides for interpreting multi-informant assessment outcomes, depending on whether the outcomes converge on the same conclusion or diverge toward distinct conclusions.

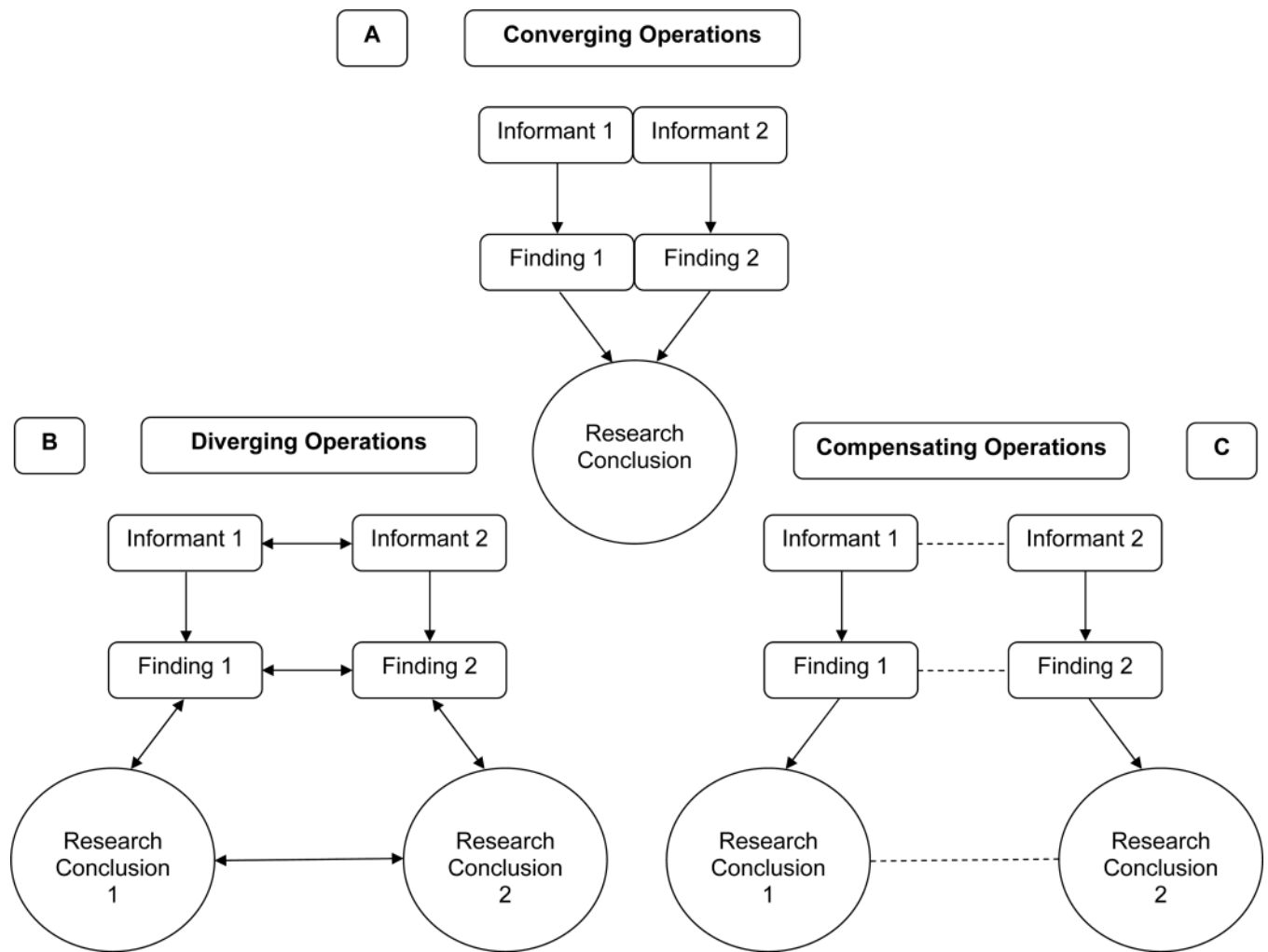


Figure 1.

Graphical representation of the research concepts that comprise the Operations Triad Model. The top half (Figure 1a) represents Converging Operations: A set of measurement conditions for interpreting patterns of findings based on the consistency within which findings yield similar conclusions. The bottom half denotes two circumstances within which researchers identify discrepancies across empirical findings derived from multiple informants' reports, and thus discrepancies in the research conclusions drawn from these reports. On the left (Figure 1b) is a graphical representation of Diverging Operations: A set of measurement conditions for interpreting patterns of inconsistent findings based on hypotheses about variations in the behavior(s) assessed. The solid lines linking informants' reports, empirical findings derived from these reports, and conclusions based on empirical findings denote the systematic relations among these three study components. Further, the presence of "dual arrowheads" in the figure representing Diverging Operations conveys the idea that one ties meaning to the discrepancies among empirical findings and research conclusions and thus how one interprets informants' reports to vary as a function of variation in the behaviors being assessed. Lastly, on the right (Figure 1c) is a graphical representation of Compensating Operations: A set of measurement conditions for

interpreting patterns of inconsistent findings based on methodological features of the study's measures or informants. The dashed lines denote the lack of systematic relations among informants' reports, empirical findings, and research conclusions.

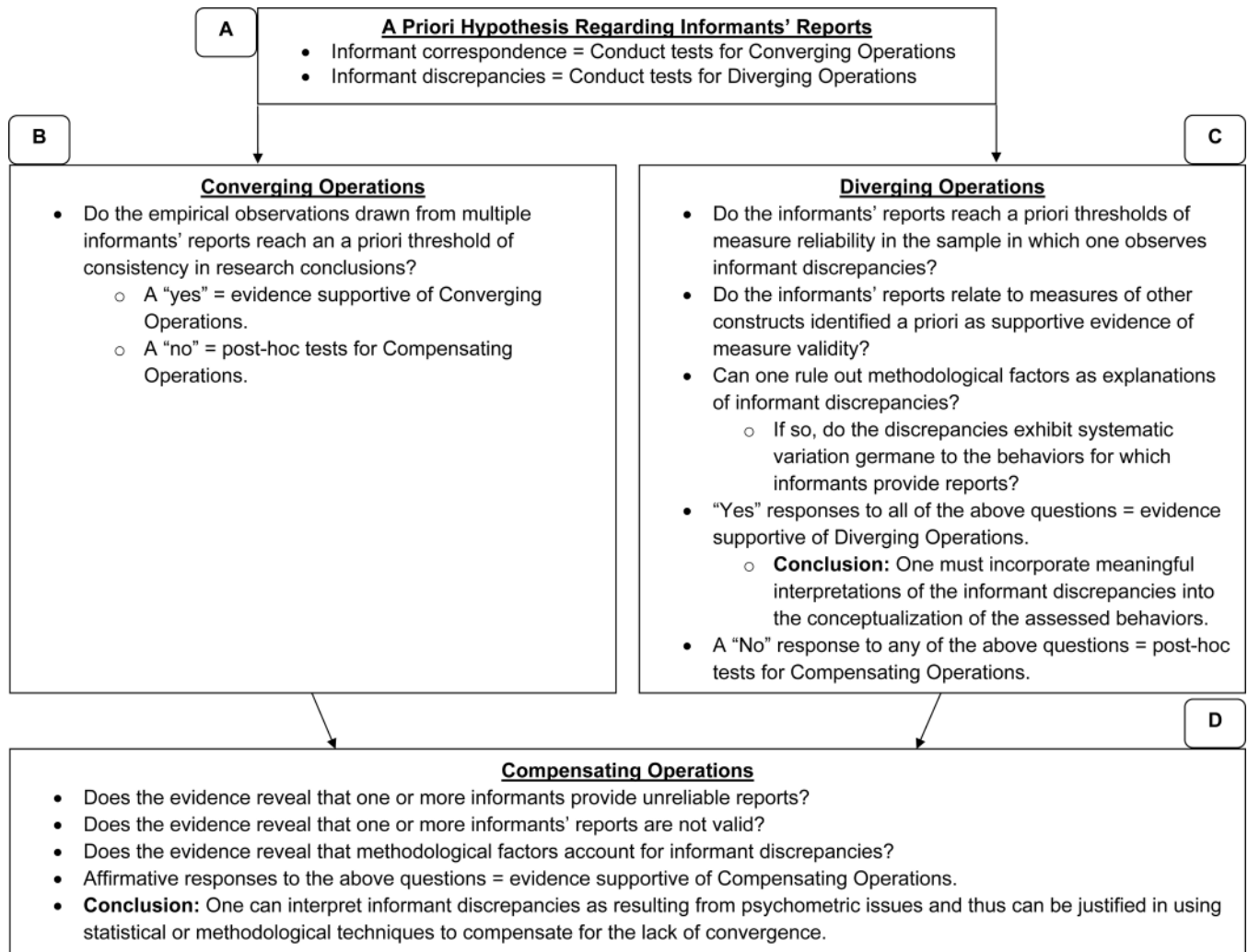


Figure 2.

Graphical display of decision making processes based on the Operations Triad Model. To use this process, researchers must pose a priori hypotheses as to whether they expect converging findings or diverging findings (Figure 2a). Empirical questions outlined in the figure can then guide researchers' tests of their 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, researchers can test whether the observations are best explained by measurement error (i.e., Compensating Operations; Figure 2d).