Improving Interpretability of Subjective Assessments about Psychological Phenomena: A Review and Cross-Cultural Meta-Analysis

Andres De Los Reyes[1] Matthew D. Lerner[2] Lauren M. Keeley[1]

Rebecca J. Weber[2] Deborah A.G. Drabick[3] Jill Rabinowitz[4]

Kimberly L. Goodman[5]

[1]University of Maryland at College Park

[2]Stony Brook University

[3]Temple University

[4]Johns Hopkins University

[5]Washington, DC

CITATION

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

Author Notes

Andres De Los Reyes and Lauren Keeley, Comprehensive Assessment and Intervention Program, Department of Psychology, University of Maryland, College Park, MD. Matthew D. Lerner and Rebecca Weber, Department of Psychology, Stony Brook University, Stony Brook, NY. Deborah A.G. Drabick, Department of Psychology, Temple University. Jill Rabinowitz, Department of Public Health, Johns Hopkins University. Kimberly L. Goodman, Washington, DC.

The first and second authors agreed to share primary authorship of this manuscript.

Effort by Andres De Los Reyes was supported, in part, by a grant from the Institute of Education Sciences (R324A180032) and a grant from the National Science Foundation (SES-1461392). Effort by Matthew D. Lerner was supported, in part, by the National Institute of Mental Health (R01MH110585).

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

Abstract

Attempts to understand *subjectivity* have historically involved distinguishing the strengths of subjective methods (e.g., survey ratings from informants) from those of alternative methods (e.g., observational/performance-based tasks). Yet, a movement is underway in Psychology that considers the merits of *intersubjectivity:* Understanding the space between two or more informant's subjective impressions of a common person or phenomenon. In mental health research, understanding differences between subjective impressions has less to do with informants' characteristics and more to do with the social environments or contexts germane to the people or phenomena examined. Our paper focuses on one relatively understudied social environment: the cultural context. We draw from seminal work on *psychological universals*, as well as emerging work on cultural norms (i.e., *cultural tightness*) to understand intersubjectivity effects through a cross-cultural lens. We report a meta-analysis of 314 studies of intersubjectivity effects in mental health, revealing that (a) this work involves independent research teams in over 30 countries, (b) informants rating a target person's mental health (e.g., parent and teacher ratings of a child's behavior) commonly provide diverging estimates of that person's mental health, and (c) greater convergence between subjective reports relates to a "tighter" or more norms-bound culture. Our paper illustrates strategies for understanding divergence between subjective reports. In particular, we highlight theoretical and methodological frameworks for examining patterns of divergence between subjective reports in relation to data from non-subjective methods. We also describe how research on intersubjectivity informs efforts to improve the interpretability of subjective assessments in multiple sub-disciplines in Psychology.

Keywords: converging operations; informant discrepancies; intersubjectivity; operations triad model; multiple informants

Psychology benefits from a strong tradition of using and interpreting multiple methods to measure psychological phenomena. These methods span subjective measures (e.g., surveys and interviews), laboratory observations (i.e., by trained raters), and objective measures (e.g., performance-based behavioral tasks). In this paper, we describe an emerging body of work that seeks to integrate knowledge gained from multi-method approaches to inquiry. Specifically, this paper reviews research on *intersubjectivity*—the space between two or more people's subjective judgments of a common person or psychological phenomenon (Gillespie & Cornish, 2010). In this paper, we examine intersubjectivity effects as they manifest in mental health research, an area that examines the space between subjective measures in relation to patterns of data taken from non-subjective measures (De Los Reyes, Thomas, Goodman, & Kundey, 2013a). Research in mental health finds that understanding the space between subjective measures requires understanding the social environments or contexts germane to the people or phenomena examined (De Los Reyes et al., 2015). We focus on better understanding one relatively understudied social environment: the cultural context. To do so, we draw from seminal work on *psychological universals* (Norenzayan & Heine, 2005), as well as emerging work on cultural norms (i.e., *cultural tightness*; Gelfand et al., 2011). We apply this work to a meta-analysis of 314 studies of intersubjectivity effects in mental health research. In turn, we highlight links between knowledge gained from this meta-analysis and research and theory on use and interpretation of subjective assessments in mental health. In doing so, we describe strategies for understanding the divergence between subjective reports of psychological phenomena. We conclude our review by describing how research on intersubjectivity informs efforts to improve the interpretability of subjective assessments in multiple sub-disciplines in Psychology.

Foundational concepts in Psychology converge on a philosophical notion of multi-method approaches to empiricism: Progress in acquiring knowledge of psychological phenomena advances most with the *combined* use of multiple measurement methods, rather than the *exclusive* use of any one method. Two foundational concepts are particularly illustrative. First, *converging operations* is the key paradigm through which scientists interpret data from multiple investigations of a given phenomenon (Garner, Hake, & Eriksen, 1956). Under converging operations, one draws firmer conclusions as multiple *methodologically distinct* investigative methods yield similar conclusions (see also De Los Reyes, Thomas, Goodman, & Kundey, 2013a). Second, the concepts of *reliability* (for a review, see Borsboom, 2005) and *validity* (Borsboom, Mellenbergh, & van Heerden, 2004; Campbell & Fiske, 1959; Cronbach & Meehl, 1955) of psychological measurement both hinge on taking multiple measurements of the constructs under investigation. In the case of validity, one might incorporate multiple methods to rule out threats to drawing sound inferences about the consistency and accuracy of measurement (e.g., shared method variance; see Kazdin, 2003).

The concepts of converging operations, reliability, and validity necessitate use of multi-method measurement. That is, all of these concepts assume that there exists a common ontological truth from which each method pulls some bit of meaningful variance, with the consequence that their use also yields less meaningful variance (i.e., error). Yet, use of multiple methods carries risks. In particular, these concepts often beget new traditions that arise from difficulties posed by the following question: What happens when estimates from multiple measurement methods do not yield the same result? For many, the answer involves embracing the merits of some methods and shunning the use of others. We characterize these traditions as *methodological boundary disputes*. Within these disputes, proponents of a particular method

(e.g., subjective vs. objective) argue for the use of that method (for a review, see Gough & Madill, 2012). They might craft arguments designed to pit methods against one another (for reviews, see Erwin, 2000; Greiffenhagen & Sharrock, 2008). They might even argue that one method (e.g., subjectivity) clouds the ability of scientists to draw sound conclusions from research (see also Hager, 1982; Shapin, 2012). For example, methodological boundary disputes often arise among studies of psychological functioning as it relates to health, education, and the workplace. In each of these areas, estimates from self-report subjective measures often yield conclusions that substantially diverge from those drawn via estimates from performance-based and observational measures, and these discrepant conclusions often result in questioning the veracity of self-reports (e.g., Dunning, Heath, & Suls, 2004).

Methodological boundary disputes do little good for advancing knowledge, because these disputes most commonly result in different research teams building justification for exclusive use of a single measurement method (e.g., eschewing subjective methods to rely on laboratory observation methods; see also De Los Reyes, Kundey, & Wang, 2011). We surmise that these disputes arise because in most research areas, converging operations is still the dominant paradigm for interpreting findings. As such, researchers tend to draw negative impressions about instances in which multi-method data yield diverging findings. Our paper demonstrates that diverging findings do not always signal "noisy data," and at times might index meaningful psychological phenomena. Further, we review research that demonstrates that for some psychological phenomena, focusing on converging findings when interpreting multi-modal assessments not only applies unrealistic standards, but also might obscure important information about the phenomenon under investigation.

In this paper, we illustrate a paradigm that may unify the various measurement traditions in Psychology and provide an important lens through which to interpret psychological research findings derived from multi-modal data. Within this paradigm, each method (e.g., subjective, laboratory observation, objective) yields useful information about the experiences and/or contexts in which one observes displays of the phenomenon under study. This paper focuses on a particular facet of this paradigm termed *intersubjectivity*, or the process of examining data drawn from the space between people's subjective impressions of a common person or phenomenon (e.g., Gillespie & Cornish, 2010). The study of intersubjectivity most commonly focuses on *participants*' and *researchers*' impressions. In our paper we call for a novel application and expansion of intersubjectivity: Studying not only the space between multiple people's subjective impressions of the same psychological phenomena, but also the links between this intersubjective space and estimates of psychological phenomena taken using non-subjective measures. Here, intersubjectivity could involve examining relations between two *informants'* subjective impressions, such as the reports of two adults (e.g., parent and teacher) about a target child's behavior. One could also examine relations between informants' impressions and those of people trained by researchers to make ratings about participants, such as clinical interviewers or researchers themselves. Studies of intersubjectivity might also integrate or use methods other than subjective reports to test theories about why differences among subjective experiences arise and how they reflect meaningful variations in psychological phenomena.

Our review focuses on intersubjectivity effects as they manifest in mental health research. We focus on work in mental health for a few reasons. Specifically, the last few decades of mental health research reveal that informant differences in subjective impressions generally are not associated with characteristics of the informants providing the subjective reports. In fact,

several meta-analyses and comprehensive reviews conducted since the 1980s find either inconsistent or no support for the idea that informant differences are related to informant characteristics such as gender, age, ethnic or racial background, social desirability, or mood-congruent rater biases (e.g., depression-related distortions; Achenbach, McConaughy, & Howell, 1987; Achenbach, Krukowski, Dumenci, & Ivanova, 2005; De Los Reyes & Kazdin, 2005; De Los Reyes et al., 2015; Duhig, Renk, Epstein, & Phares, 2000; Renk & Phares, 2004; Richters, 1992). This is an important observation, because effects specific to rater characteristics may attenuate the ability of subjective measures to account for the assessed construct(s), and thus may compromise the established psychometric properties and utility of these measures (see also Groth-Marnat & Wright, 2016). Instead, an emerging body of work finds that differences between subjective impressions reflect variations in the social environments or contexts relevant to the people or phenomena examined, such as the home, school, or work contexts where people might display mental health concerns (e.g., De Los Reyes et al., 2013a). Thus, the study of intersubjectivity in mental health provides fertile ground for developing paradigms that improve the interpretability of subjective impressions of psychological phenomena.

As we describe below, much of the research linking intersubjectivity effects to social context focuses on how differences in subjective impressions of what is being rated (e.g., child's behavior) reflect the idea that (a) the rated phenomenon displays differently depending on the context (e.g., home vs. school), and (b) informants providing subjective reports often vary in terms of the context in which they observe the phenomenon being rated (e.g., parent at home vs. teacher at school) (De Los Reyes, Ohannessian, & Racz, 2019). Nevertheless, this is only one potential approach to social context and other elements of context warrant consideration. In intersubjectivity research in mental health, one relatively understudied element is the cultural

context. Several recent large-scale studies point to the idea that, across countries, informants commonly provide discrepant estimates of child and adolescent mental health (Rescorla et al., 2012, 2013, 2014). This finding points to a core idea: People subjectively perceive a common person or phenomenon in diverging ways and the degree of this divergence is quite stable. Yet, epidemiological research clearly demonstrates that although mental health concerns manifest all over the world, variability also exists across countries in estimates of mental health concerns (e.g., Hunsley & Lee, 2014). Similarly, it is unknown whether divergence among subjective reports of mental health concerns displays between-country variations. If between-country variations exist, to what extent do they reflect meaningful variations in cultural phenomena?

To address these questions, we consider two elements of cross-cultural research and theory. In particular, Norenzayan and Heine (2005) provide a *psychological universals* framework from which to test theoretically relevant questions that provides a foundation for testing cross-cultural consistency in displays of psychological phenomena. In doing so, Norenzayan and Heine (2005) posit several questions of particular importance. Does the phenomenon manifest similarly across cultures? If cross-cultural differences in the phenomenon exist, do they stem from measurement error or some methodological process (e.g., poor translation of subjective measures across cultures)? If a phenomenon manifests similarly across cultures and variations reflect meaningful psychological phenomena, is the phenomenon accessible to the same degree across cultures (i.e., accessed with the same frequency across cultures)? To address these questions, in this paper we report findings of a cross-cultural meta-analysis of over 300 studies of intersubjectivity effects in mental health research.

Norenzayan and Heine (2005) provide us with a framework to test for the cross-cultural generalizability of intersubjectivity effects in mental health research. However, we require a

conceptual frame to facilitate tests of whether any variations across cultures reflect "noise" or instead meaningful cultural phenomena. Thus, in our review we also incorporate the work of Gelfand and colleagues (2011) on key facets of cultural norms, namely, their *cultural tightness* framework. Specifically, recent work focuses on characterizing countries by the extent of the (a) strength of their cultural norms and (b) tolerance for behavioral deviations from these norms (Gelfand et al., 2011). In this approach, countries display a measurable *cultural tightness* that exists on a continuum. The low end of the range (i.e., *looseness*) characterizes countries with cultures that have weak norms and a high tolerance for deviation from norms. The high end of the range (i.e., *tightness*) characterizes countries with cultures that have strong norms and a low tolerance for deviation from norms. The value in studying cross-cultural variations in intersubjectivity effects within this *tightness-looseness framework* is that it results in a parsimonious, quantifiable characteristic that relates to a host of political and economic indices (e.g., presence of autocratic political systems and/or resource scarcity), as well as the presence of territorial and environmental threats (e.g., population density, exposure to natural disasters, territorial conflicts between neighboring countries; Gelfand, 2012; Gelfand & Jackson, 2016).

Informed by cultural tightness, we can also make specific predictions regarding culturally meaningful variations in intersubjectivity effects. Specifically, if cultural tightness relates to intersubjectivity effects, then one should observe increased cultural tightness relating to increased correspondence between informants' subjective impressions. Indeed, a country with relatively tight cultural norms should result in not only constraints on public displays of deviant behavior but also heightened awareness among its citizens as to when such deviations occur. The implication is that if deviations from culturally normed behavior manifest (e.g., children's displays of mental health concerns), then informants providing reports about these concerns (e.g.,

parents and teachers) should be culturally attuned to recognizing and reporting these behaviors when prompted to do so (e.g., on mental health surveys). This should result in more informants rating mental health concerns when displayed but also fewer informants rating mental health concerns when absent. Thus, inspired by the work by Norenzayan and Heine (2005) and Gelfand and colleagues (2011), we tested two hypotheses. First, although we expected low overall convergence between informants' mental health reports across countries, we also expected to observe some country-level variation. Second, given this country-level variation, we expected to observe that increased cultural tightness would relate to increased magnitudes of correspondence between subjective reports, such that the "tighter" a country's rated culture, the greater the correspondence would be between informants' mental health reports.

There are also various methodological factors that make mental health research an ideal area for examining intersubjectivity effects. For instance, many independent teams of mental health researchers leverage this approach to scientific inquiry, allowing for a conservative test of our claims about intersubjectivity effects. Further, intersubjectivity in mental health research traverses multiple fields (e.g., Clinical, Developmental, and Counseling Psychology; Education; Human Development; Nursing; Pediatrics; Psychiatry; Social Work). Thus, by studying intersubjectivity in mental health, we examine effects that likely generalize through not just the social sciences, but any field that uses subjective methods. Within mental health research, we focus on work with children and adolescents. Indeed, over 50 years of work within these developmental periods has incorporated multiple informants' subjective reports, and only recently has work involving adults begun taking this multi-informant approach to subjective measurement (e.g., Achenbach, 2017; Achenbach et al., 2005; De Los Reyes et al., 2015; Hunsley & Mash, 2007). Therefore, to maximize precision in estimating intersubjectivity effects,

we focus on those developmental periods with the greatest available multi-informant data. When possible, we discuss findings of relevance to work with adults.

The Present Review

In sum, the goal of the current review is to advance knowledge on intersubjectivity in Psychology. We do so in three ways. First, we review key findings and theoretical frameworks germane to the study of intersubjectivity in mental health research. Second, we use this review as a backdrop for reporting findings of the cross-cultural meta-analysis described previously. Third, we describe the research and theoretical implications of our work, with particular emphasis on describing how intersubjectivity research might improve the interpretability of subjective reports as used in multiple sub-disciplines in Psychology.

Intersubjectivity in Mental Health Research

Mental health researchers seek to understand the psychological, social, and biological etiologies of mental health concerns, and apply this knowledge to developing effective techniques for preventing or ameliorating the negative effects of such concerns (e.g., American Psychiatric Association, 2013; Cicchetti, 1984). As a consequence of this multi-faceted approach to mental health services and in conjunction with other measurement methods (e.g., biological, laboratory observations, official records), subjective behavioral reports have enjoyed a long history of use and interpretation in mental health research, particularly in research involving children, adolescents, and their families (Achenbach, 2017; De Los Reyes, 2011; De Los Reyes & Ohannessian, 2016; Kraemer et al., 2003). Specifically, a key component of best practices in child and adolescent mental health research involves collecting subjective reports about child and adolescent behavior from the perspectives of multiple informants (Hunsley & Mash, 2007). Using this multi-informant approach, researchers collect self-reports from children and

adolescents, who capably provide reliable and valid self-reports about a variety of mental health domains (e.g., aggression, anxiety, mood; see Dirks, De Los Reyes, Briggs-Gowan, Cella, & Wakschlag, 2012). Additionally, clinicians and researchers rely on the reports of significant others who spend a great deal of time observing the children and adolescents being assessed, with a focus on informants who vary in their observational contexts (for a review, see De Los Reyes, Thomas, Goodman, & Kundey, 2013a). For instance, researchers conceptualize parents and teachers as informants of children's behavior primarily in home and school contexts, respectively (e.g., Kraemer et al., 2003).

A key goal of the multi-informant approach is to collect subjective reports of child and adolescent behavior to capitalize on individual differences in youths' displays of mental health concerns within and across contexts (De Los Reyes, 2013). Specifically, if informants vary in where they observe child/adolescent behavior, then researchers can use variations in informants' subjective impressions of this behavior to understand whether and to what extent children and adolescents vary in displays of mental health concerns, depending on the social context (Dirks et al., 2012). This contextually sensitive information has implications for not only diagnosing child and adolescent mental health concerns, but also accurately estimating treatment response and identifying evidence-based treatments (De Los Reyes & Kazdin, 2008). Not surprisingly, the evidence supporting the efficacy of child and adolescent mental health treatments largely rests on informants' subjective reports of treatment outcomes (De Los Reyes & Aldao, 2015; De Los Reyes, Kundey, & Wang, 2011; Hunsley & Mash, 2007; Weisz, Jensen Doss, & Hawley, 2005).

Ubiquity of Divergence in Outcomes from Multi-Informant Assessments

Multi-informant assessment approaches capitalize on the unique perspectives of informants. In line with informants' unique perspectives, one of the most robust observations in

mental health research is that multiple informants of child and adolescent mental health provide reports that yield diverging estimates of such mental health (De Los Reyes & Kazdin, 2005). Stated another way, in any one study sample and for any one child or adolescent, estimates taken from one informant's mental health report (e.g., parents) rarely correspond with the estimates taken from another informant's mental health report (e.g., teachers). Importantly, low correspondence between informants' reports occurs even when researchers hold key aspects of the psychometric properties of measures constant across informants' reports (e.g., identical item content and response options; same or similar estimates of reliability and validity; for reviews, see Achenbach & Rescorla, 2001; De Los Reyes, 2011, 2013; De Los Reyes et al., 2015).

In Table 1, we briefly outline key findings from quantitative reviews, cross-cultural studies, and longitudinal studies of cross-informant correspondence. Briefly, multi-informant assessments yield informants' reports that diverge from each other on symptom levels of multiple mental health domains (e.g., disruptive behavior, inattention, hyperactivity, mood, social anxiety; for reviews, see Achenbach, McConaughy, & Howell, 1987; De Los Reyes et al., 2015; Duhig et al., 2000). In fact, quantitative reviews summarized in Table 1 indicate that not only do any two informants' reports diverge from one another, but that the level of divergence, in part, depends on such factors as the informant pair (e.g., mother-father vs. parent-child vs. parent-teacher) and the mental health domain assessed (e.g., aggression vs. anxiety). Between-informant divergence also characterizes assessments of risk and protective factors of mental health concerns (e.g., exposure to violence, parenting, sleep behavior, social competence; Alfano, Patriquin, & De Los Reyes, 2015; Goodman, De Los Reyes, & Bradshaw, 2010; Renk & Phares, 2004; Taber, 2010). Importantly, between-informant divergence translates into discrepant research findings and thus diverging conclusions about important research topics,

including the prevalence and treatment of child and adolescent mental health (e.g., De Los Reyes & Kazdin, 2005). For example, quantitative reviews of psychosocial treatment studies indicate that the effects of child and adolescent mental health treatments range from small-to-large in magnitude, depending on the informant completing the outcome measures (e.g., Casey & Berman, 1985; De Los Reyes & Kazdin, 2006, 2009; Weisz, Weiss, Alicke, & Klotz, 1987).

Theory on Informants' Subjective Reports in Mental Health Research

Some researchers have theorized that between-informant divergence reflects low informant reliability, invalidity, or bias (for reviews, see De Los Reyes, 2013; De Los Reyes & Kazdin, 2005; Richters, 1992). Indeed, measurement models suggesting that informants' reports emerge from common latent variables assume this to be so (see De Los Reyes et al., 2011; Holmbeck, Li, Schurman, Friedman, & Coakley, 2002). If true, the implication for interpreting diverging findings appears straightforward: except for these hindrances to the psychometric soundness of subjective data, one would observe converging findings among informants' reports.

Yet, the last 10 to 15 years of theory and research in this area has focused on identifying when divergence between subjective reports captures meaningful contextual variation in behaviors about which informants provide reports (De Los Reyes et al., 2013a; Kraemer et al., 2003). Below, we review studies that indicate the level and stability of correspondence between informants' reports can yield tools for understanding how children and adolescents behave and view their world differently, depending on the context. In fact, the two major quantitative reviews to date of between-informant divergence collectively reviewed nearly 500 studies conducted over 50 years, and although both reviews were separated by over 25 years they obtained nearly identical mean estimates of between-informant divergence (e.g., Achenbach et al., 1987; De Los Reyes et al., 2015). Under these circumstances, it would be unrealistic to

expect converging findings among informants' reports. Further, simply assuming that the lack of convergence stems from issues surrounding the veracity of informants' reports might result in a loss of important information about the behaviors being assessed.

In line with recent empirical work, researchers have sought to conceptualize the nature of between-informant divergence and its reflections of individual differences in psychological phenomena. In Figure 1, we graphically depict this conceptualization. Briefly, the Operations Triad Model holds that reports from informants who vary in the contexts in which they observe behavior (e.g., parents at home vs. teachers at school) may predictably converge if one hypothesizes that the behavior being rated displays invariability across contexts (i.e., *converging operations*; Figure 1a), or diverge if one expects the behavior to vary across contexts (i.e., *diverging operations*; Figure 1b) (De Los Reyes et al., 2013a). An example of converging operations might involve an assessment of a child who behaves hyperactively across home and school contexts, for which both the parent and teacher report that the child displays relatively high levels of hyperactive behavior. Conversely, an example of diverging operations might involve an assessment of another child who behaves hyperactively at school but not at home, in which the teacher reports that the child displays relatively high levels of hyperactive behavior, whereas the parent reports that the child displays relatively low levels of this behavior. Stated another way, the Operations Triad Model holds that one should expect informants to provide converging estimates of the behaviors being assessed insofar as these behaviors manifest consistently across the contexts in which informants observe the behaviors (Figure 1a). However, if the behaviors being assessed might meaningfully vary between informants' observational contexts, it may be more reasonable to expect that informants' reports will provide diverging estimates of the assessed behaviors (Figure 1b).

Alternatively, informants' subjective reports may diverge because of methodological artifacts (i.e., *compensating operations*; Figure 1c) inherent in the measurement process (e.g., measurement error), rather than true variation in the behavior being rated *per se*. An example of compensating operations might involve an assessment of a child's hyperactivity in which the parent and teacher provide diverging estimates of the child's levels of hyperactive behavior. However, these diverging estimates of hyperactive behavior are parsimoniously explained by the fact that the measures completed by the informants differed on key characteristics such as differences in item content or response options (see also Schwarz, 1999). As illustrated in Figure 2, the methodological artifacts that reflect compensating operations can take many forms beyond this example, including but not limited to differences in score reliability or validity. In the following two sections, we report several lines of research that support the main tenets of the Operations Triad Model in research on child and adolescent mental health.

The Interpretive Value of Divergence between Subjective Reports

Mental health researchers examine the *interpretive* value of diverging reports, as well as the *predictive* value of such reports (i.e., divergence as predictor or outcome). In fact, the Operations Triad Model capitalizes on the value of subjective reports for guiding research on the assessment, development, prevention, and treatment of child and adolescent mental health concerns. To this end, in Table 2, we summarize a number of studies that support the interpretive and predictive value of diverging reports in mental health research and by extension, the Operations Triad Model. These studies include examinations of divergence between participants' reports (e.g., parent and teacher; parent and adolescent), as well as divergence between participant and researcher reports (e.g., patient and clinician). Broadly, empirical work supports the idea that divergence between participants' subjective reports contains useful information

about child and adolescent mental health and risk factors for these concerns. This divergence functions as both a useful interpretive tool for understanding contextual variation in child and adolescent behavior, as well as a key predictor of maladaptive mental health outcomes among children and adolescents (Table 2). Multiple lines of research provide such support.

Multi-informant assessments of child and adolescent behavior. As mentioned previously, researchers often conceptualize parents' and teachers' subjective reports of child and adolescent behavior as reflecting displays of such behavior in home versus school contexts. Initial support for this idea came from the Achenbach et al. (1987) quantitative review, which observed greater levels of correspondence among pairs of informants observing behavior within the same context (e.g., pairs of parents and pairs of teachers), relative to informants observing behavior within different contexts (e.g., parents and teachers).

In Table 2, we summarize findings of prior work that directly tests these ideas about contextual information gleaned from studying divergence between informants' reports. One line of work evaluates parent and teacher reports of children's behavior in relation to laboratory observations of such behavior (De Los Reyes, Henry, Tolan, & Wakschlag, 2009). In this work, researchers constructed assessments that included laboratory observations that quantified the extent to which preschool children exhibited disruptive behavior in interactions with parental versus non-parental adults (i.e., clinical examiner), or in both or neither one of these interactions (for a review, see Wakschlag, Tolan, & Leventhal, 2010). This study thus demonstrated that laboratory observations could be used to take an independent measure of cross-contextual variability in displays of disruptive behavior. Outcomes from these observations could then be compared against instances in which parent and teacher reports diverged or converged with one another on whether a given child displayed disruptive behavior. In line with the laboratory

observations, parent and teacher reports of disruptive behavior converged when children displayed disruptive behavior across laboratory interactions (i.e., converging operations; Figure 1a), and diverged from each other when children displayed disruptive behavior in specific laboratory interactions (i.e., diverging operations; Figure 1b) (De Los Reyes et al., 2009). Moreover, these findings are not specific to interpreting parents' and teachers' reports about preschoolers' behavior. Parent and teacher reports of child and adolescent aggressive behavior converge to a greater extent when parents and teachers report similar as opposed to disparate environmental circumstances preceding displays of such behavior (e.g., teasing by peers; Hartley, Zakriski, & Wright, 2011).

Recent research has extended this work to examining parental caregivers' reports of adolescent mental health, or informants who both observe adolescents' behavior in the home context. Specifically, relative to caregiver dyads that diverge in their reports of adolescent mental health, caregiver dyads that converge in reporting relatively high adolescent mental health concerns have adolescents who evidence greater levels of hostility within structured interactions with caregivers (De Los Reyes, Alfano, Lau, Augenstein, & Borelli, 2016). Consistent with converging operations (Figure 1a), when informants who observe behavior in the same context converge in reporting relatively high levels of adolescent mental health concerns, this may point to increased risk for such concerns. In fact, recent work suggests that for some clinical conditions (e.g., autism spectrum symptoms), when informants who observe behavior in different contexts (i.e., parents and teachers) converge in reports of relatively high clinical concerns, this may point to children who evidence increased clinical impairments according to external validator metrics (e.g., gold-standard diagnostic tools, service receipt) relative to children for whom informants' reports diverge (e.g., parent's report reveals concerns that the teacher's report does not

corroborate; Lerner, De Los Reyes, Drabick, Gerber, & Gadow, 2017). In sum, divergence between parents' and teachers' reports (and convergence between caregivers' reports) of child and adolescent behavior may reflect meaningful variations in displays of such behavior.

Informants attend to contextual information when providing subjective reports. If a key assumption underlying use and interpretation of informants' subjective reports is that informants observe behavior in different contexts, then empirical support of such an assumption might come from evaluating whether informants attend to contextual information when providing subjective reports. In recent experimental work, trained judges of children's behavior (i.e., clinicians experienced in the assessment and treatment of child and adolescent mental health) were more likely to rate children described in vignettes as evidencing conduct problems if the environments in which these children lived contained risk factors for conduct problems (e.g., presence of parental mental health concerns) versus environments containing no such risk factors (De Los Reyes & Marsh, 2011). More recent work indicates that effects of contextual information generalize to clinicians' judgments of other mental health domains (i.e., attention and hyperactivity; panic disorder), as well as laypeople's judgments (Marsh, Burke, & De Los Reyes, 2016; Marsh & De Los Reyes, 2018; Marsh, De Los Reyes, & Wallerstein, 2014).

Researchers have extended work on understanding trained clinicians' subjective reports to interpreting divergence between untrained participants' reports. In prior experimental work (De Los Reyes et al., 2013b), researchers examined mother and adolescent reports of parental knowledge of adolescent whereabouts and activities, a key risk factor for adolescent conduct problems (for a review, see Smetana, 2008). By construction, parental knowledge is intimately connected to social contexts outside of the home environment, or environments in which adolescents can engage in activities outside of direct parental observation. In line with this, De

Los Reyes and colleagues (2013b) trained mothers and adolescents to make subjective reports about parental knowledge based on the number of contexts in which they observed behaviors indicative of parental knowledge. In this study, trained research staff provided mothers and adolescents with lists of environmental contexts, such as parties that the adolescent attends or where the adolescent spends time on the weekend. From these lists, mothers and adolescents identified contexts that were "great examples" of where the behaviors described in survey items of parental knowledge "happen" (example item: "'Does your mother know what you do during your free time?'"). Participants were then instructed to make ratings on parental knowledge surveys based on the number of contexts that came to mind as they read the items (i.e., greater contexts → greater ratings). For example, on the 1-5 parental knowledge scale (range of "Not at all" [1] to "A lot" [5]), if a parent identified 13 or more "great example" contexts for which they knew where adolescents spent their free time, then the parent would rate the "free time" item a "5". This training produced *greater differences* between mothers' and adolescents' reports about parental knowledge, relative to the differences observed when mothers and adolescents completed their parental knowledge surveys without such training. This finding supports the idea that mothers' and adolescents' subjective reports about parental knowledge diverge, in part, because they each think about different kinds of contextual information when forming their subjective reports about behaviors indicative of parental knowledge.

Using intersubjectivity to predict behavioral outcomes. Much of our discussion thus far has focused on research examining divergence between participants' reports as a tool for characterizing contextual variability in displays of child and adolescent behavior. What this work indicates is that intersubjectivity effects yield incremental, context-sensitive knowledge about the *current* mental health of children and adolescents being assessed, over-and-above what one can

learn from one person's subjective impressions about mental health. Yet, another line of research in intersubjectivity focuses on using divergence between reports as a tool for *longitudinal* prediction of child and adolescent outcomes (see De Los Reyes & Ohannessian, 2016; Goodman et al., 2010). In this view, informants, namely parents and their children/adolescents, are often asked to provide reports of behaviors that hold significant saliency to them, such as aspects of their relationship quality, the level of conflict in the home, and the presence of specific parenting behaviors (e.g., corporal punishment). Thus, when parents and their children hold converging or diverging views about the home or family environment, the convergence/divergence itself may reflect how they interact with each other and how they perceive and/or react to their environment (De Los Reyes, Lerner, Thomas, Daruwala, & Goepel, 2013c). In this view, the degree of divergence between subjective reports may serve as a marker or predictor of long-term functioning (e.g., De Los Reyes et al., 2019).

In Figure 3, we graphically present our theoretical model for divergence between subjective perspectives, with a focus on parents' and adolescents' subjective judgments about family functioning. The concepts depicted in Figure 3 are modified versions of the Operations Triad Model concepts depicted in Figure 1 (see also De Los Reyes & Ohannessian, 2016). For example, when parents and adolescents converge in reporting that the family environment contains relatively high levels of a positive family domain (e.g., parental acceptance of the adolescent), this convergence may signal the presence of protective factors in the family environment that buffer against adolescent mental health concerns (i.e., *converging operations*; Figure 3a). Conversely, parents and adolescents might diverge in reporting about levels of family risk factors (i.e., *diverging operations*; Figure 3b). This form of diverging reports based on levels reported by parents and adolescents may signal a lack of parental awareness about the

adolescent's whereabouts and activities, a risk factor for the development of adolescent mental health concerns (see also Goodman et al., 2010).

In Table 2, we summarize two studies that support the framework described in Figure 3, with detailed reviews of this literature available elsewhere (see De Los Reyes & Ohannessian, 2016; De Los Reyes et al., 2019). In one study, researchers took parents' and adolescents' reports of parental monitoring (i.e., parental knowledge and adolescents' disclosure about their whereabouts and activities), as well as structured interview assessments of how much parents and adolescents diverged in their views about everyday life events (e.g., completing chores and homework; De Los Reyes, Salas, Menzer, & Daruwala, 2013d). Consistent with converging operations (Figure 3a), parents and adolescents who converged on survey reports of high parental monitoring evidenced the lowest levels of perceived diverging views about topics that arise in their daily lives (e.g., relative to instances in which parents and adolescents diverged on parental monitoring reports). That is, indirect assessments of divergence between informants' reports (i.e., measured convergence between reports of parental monitoring) relate strongly with direct assessments of informants' perceived divergence (i.e., interviews of perceived disagreements about family life).

Another study provided support for the ability of convergence between participants' reports to predict adolescent mental health outcomes (Laird & De Los Reyes, 2013). In a sample of 218 adolescents and their parents, researchers took parent and adolescent survey reports of various aspects of the parent-adolescent relationship, and examined the utility of the convergence between reports in predicting adolescent depressive symptoms. Consistent with converging operations (Figure 3a) and relative to other reporting patterns (e.g., divergence between reports), convergence between parent and adolescent reports on both low levels of parent-adolescent

conflict and high levels of parental acceptance of adolescent behavior predicted relatively low levels of adolescent depressive symptoms.

A series of studies from Lippold and colleagues (2013, 2014) yielded findings supporting diverging operations (Figure 3b). In this work, adolescents and parents completed reports about parental knowledge and adolescents provided reports about substance use. The authors observed fairly stable patterns of convergence and divergence between reports about parental knowledge. Consistent with diverging operations (Figure 3b), the combination of parents' reporting relatively high knowledge about an adolescent's whereabouts and activities and adolescents' reporting relatively low parental knowledge of such activities placed adolescents at increased risk for substance use concerns, relative to other parent-adolescent reporting patterns. In sum, researchers have found value in examining convergence and divergence between participants' reports as reflections of contextual variability in child and adolescent behavior, qualitative features of participants' subjective perceptions, and longitudinal predictors of child and adolescent behavioral outcomes. This work reflects both the descriptive and predictive value of intersubjectivity in mental health research.

Studying Divergence between Participants' and Researchers' Reports

As evidenced in Table 2, mental health researchers focus much attention on understanding divergence between *participants'* reports. Yet, much can be learned by studying the relations between reports completed by *participants and researchers*. Interestingly, we can also point to recent mental health research addressing intersubjectivity in reports completed by participants and information sources trained by researchers to provide subjective impressions about participants, such as trained clinical interviewers and laboratory confederates. In Table 3, we outline the findings of some of this work. Likewise, we identify work on the divergence

between participants' reports and putatively objective measures.

Recent work leveraged reports from laboratory personnel to clarify the outcomes of multi-informant assessments of adolescent social anxiety. These assessments often include adolescent and parent reports, and the low reporting correspondence levels typically observed in these assessments often results in uncertainties in clinical decision-making (De Los Reyes et al., 2015). An important feature of adolescent social anxiety concerns is that these concerns often manifest within non-home contexts such as peer interactions (Alfano & Beidel, 2011). Yet, current practices in evidence-based assessments of adolescent social anxiety rarely involve collecting peer reports (Silverman & Ollendick, 2005), presumably because collecting such reports raises confidentiality concerns (Card & Hodges, 2008). In particular, adolescent social anxiety assessments would benefit from context-specific reports relevant to adolescent social anxiety, namely, interactions with unfamiliar peers (American Psychiatric Association, 2013). In fact, reports about social anxiety displayed in peer interactions may reveal data that cannot be reliably gleaned from reports by other observers of adolescent social anxiety. Thus, in a sample of 89 adolescents (30 clinic-referred; 59 community control), Deros and colleagues (2018) collected social anxiety reports from adolescents, their parents, and unfamiliar peer confederates who interacted with adolescents during 20-minute mock social interactions. Adolescents' social anxiety self-reports correlated with reports on parallel measures from parents in the .30s and with peer confederates in the .40s to .50s. However, reports from parent-confederate dyads correlated in the .07 to .22 range. These reporting patterns indicate that the lack of correspondence between adolescent and parent reports cannot be attributed to faulty self-reports on the part of adolescents (e.g., social desirability) given their strong correspondence with reports from peer confederates. Further, all of these informants' social anxiety reports distinguished clinic-referred from

community control adolescents. However, adolescents' and peer confederates' (but not parents') social anxiety reports predicted adolescents' state arousal within interactions with peer confederates. These findings indicate that adolescents' self-reports can diverge from parent reports and still yield valid data in their own right. Further, these findings indicate that the divergence between parent and peer confederate reports provides important information about how adolescents' concerns manifest in home versus non-home (e.g., peer interaction) contexts.

In other work, researchers have tested ideas about informants' subjective reports reflecting contextual information by comparing them to "objective" measures. Specifically, adolescents provide internally consistent and valid self-reports of social anxiety, even when these reports (a) evidence fewer symptoms than parent reports about adolescents and (b) yield estimates that diverge with estimates taken from wireless heart rate monitors that directly assess adolescents' psychophysiological regulation during a baseline resting period (De Los Reyes et al., 2012). That is, even when adolescents' self-reports of social anxiety diverge with objective measures of adolescent psychophysiology, these self-reports nevertheless can validly distinguish those adolescents who completed assessments as part of a clinical evaluation for social anxiety from those who completed assessments as part of a community-based (i.e., non-clinical) study. These findings are in line with prior work indicating that parents and children who converge in their anxiety symptom reports do so to a greater extent when the reports are about anxiety symptoms displayed in non-school versus school contexts (Comer & Kendall, 2004). Stated another way, adolescents' self-reports may diverge from objective indices of adolescents' psychophysiology, and each may still be valid indicators of social anxiety because both objective and subjective measures assess meaningful processes relevant to understanding and treating social anxiety. In prior work, using direct measurements of physiological arousal, adolescents

diagnosed with social anxiety disorder responded physiologically to anxiety-provoking, laboratory controlled social interaction tasks much like adolescents who did not meet criteria for any mental disorder, in that adolescents all displayed gradual habituation to these situations over a 20-minute period (Anderson & Hope, 2009). However, in this study adolescents diagnosed with social anxiety disorder displayed a stable and high level of perceived physiological distress throughout this period. Interestingly, closing this "gap" between objective and subjective physiological reactivity forms a key component of exposure-based treatments for social anxiety (Thomas, Aldao, & De Los Reyes, 2012). Indeed, these treatments essentially provide adolescents with training on how to become better "detectives" of their emotional states, or learn to identify when they habituate to the very social situations that prompted treatment (Alfano & Beidel, 2011). Thus, understanding patterns of convergence and divergence between adolescents' self-reports and objective psychophysiological measures can be useful metrics in the assessment and treatment of adolescent social anxiety (see also De Los Reyes & Aldao, 2015). More broadly, this suggests circumstances in which diverging findings support the use of subjective perspectives, even when the nature of the divergence involves data taken from subjective reports and data taken from objective measurements.

Cross-Cultural Meta-Analysis of Intersubjectivity Effects

Thus far, we have reviewed work in mental health research, namely, research examining the interpretive value of effects observed in research on intersubjectivity. This work demonstrates the value of multi-informant approaches to subjective measurement. More importantly, it illustrates paradigms for not only integrating subjective reports with other methods, but also in fact *using* data taken from other methods to increase the interpretability of subjective reports. Much of this work highlights the relevance of understanding the social

context in which mental health concerns manifest, but as mentioned previously, relatively little is known about links between intersubjectivity effects in mental health and the cultural context. To address this gap, we conducted a cross-cultural meta-analysis of prior work on cross-informant correspondence in mental health reports.

Psychological Universals

We started with the questions stemming from the psychological universals framework by Norenzayan and Heine (2005) described previously to test the cross-cultural consistency of cross-informant correspondence. We modified the question structure slightly to match the research questions and methods examined in prior work on intersubjectivity in mental health research (see Table 1). First, does the phenomenon manifest similarly across cultures? For the purposes of our review, does cross-informant correspondence in reports about mental health remain low-to-moderate across cultures? Alternatively, might cultures exist in which cross-informant correspondence rises to a magnitude or level in which one can assert that informants' reports provide redundant estimates of mental health?

Second, even if the phenomenon manifests similarly across cultures, cross-cultural differences in the phenomenon might nonetheless exist. That is, though low-to-moderate magnitudes of cross-informant correspondence might be normative, could correspondence estimates for some cultures significantly vary from others? If so, do these cross-cultural differences manifest because of measurement error or some methodological process (e.g., poor translation of subjective measures across cultures)? Alternatively, might variations in the magnitudes of cross-informant correspondence meaningfully reflect variations in cultural norms?

Third, even if the phenomenon manifests similarly across cultures and variations in the phenomenon meaningfully relate to variations in cultural norms, is the phenomenon accessible to

the same degree as well? For the purposes of our review, we were curious as to whether links among magnitudes of cross-informant correspondence and cross-cultural variations in cultural norms operate similarly for informants' reports of behaviors that are relatively easier to observe or accessible (i.e., externalizing problems such as hyperactivity) and relatively difficult to observe or less accessible (i.e., internalizing problems such as mood). Addressing these three questions allowed us to examine the degree to which intersubjectivity effects are universal or broadly relevant to the study of cross-informant reports about child and adolescent mental health.

Cultural Tightness

To test cross-cultural effects on cross-informant correspondence, we considered examining country-level variations in cross-informant correspondence (i.e., correspondence levels for Country X vs. Country Y). This would have been both atheoretical and a considerably under-powered approach to addressing our aims (i.e., for between-country comparisons for most countries with the exception of comparisons involving the United States, which contributed the most studies to the meta-analysis). Further, the number of tests would have resulted in inflated Type I error rates (i.e., individual countries compared to each other). Consequently, our dimensional approach to addressing our aims was informed by the cultural tightness work by Gelfand and colleagues described previously. Specifically, we drew from recent cross-cultural work that quantified the countries represented in our sample in terms of cultural tightness. Using their tightness-looseness scale (i.e., Likert-typed scaled self-report instrument administered to participants from 33 countries), Gelfand and colleagues (2011) reported country-level cultural tightness scores drawn from large samples of participants from each country. This scale included six items consisting of statements regarding the country's cultural tightness. Respondents rated these statements on a six-point scale of agreement with the statement (i.e., 1 = *strongly disagree*

to 6 = *strongly agree*; example item: "In this country, if someone acts in an inappropriate way, others will strongly disapprove"). Greater scores reflected greater cultural tightness. In Table 1 of Gelfand and colleagues (2011), the authors reported mean cultural tightness scores from this scale for 33 countries, with ranges including a low score of 1.6 (Ukraine), to a cross-cultural mean score of 6.5 (Germany [former West]), to a high score of 12.3 (Pakistan). We inputted cultural tightness scores for each study in our sample that was based in a country that had a cultural tightness score reported in Table 1 of Gelfand and colleagues (2011).

Method

Study database. We examined a sample of 314 articles (396 samples, 1560 data points). Two-hundred and sixty-eight of these articles came from a previous meta-analysis of cross-informant correspondence in child and adolescent mental health reports (De Los Reyes et al., 2015). (As an aside, the original sample overlapped with the cross-cultural studies examined by Rescorla and colleagues [2012, 2013, 2014]; to avoid redundancies with this work, we excluded these publications from our own cross-cultural review). In De Los Reyes and colleagues (2015), two doctoral students blind to hypotheses reliably coded effect size metrics from these studies, along with the study characteristics described below. Additional coding procedures and effect size calculations are described elsewhere (De Los Reyes et al., 2015). Importantly, the studies examined by De Los Reyes and colleagues (2015) covered research published up until 2014, and so we performed an additional search for studies in 2015-2016, using the same search procedures as De Los Reyes and colleagues (2015). This search resulted in an additional 46 studies included in the review, for a grand total of 314 studies coded in our meta-analysis. In total, coders recorded 1,560 effect sizes from these studies, along with country of origin and mental health domain assessed (i.e., externalizing, internalizing).

Countries coded in the current review. For the current review, two coders masked to study hypotheses coded the country of origin for all studies meta-analyzed in the previous reviews reported in Table 4. These include all studies from seven meta-analyses in Table 1 (i.e., Achenbach et al., 1987, 2005; Duhig et al., 2000; Renk & Phares, 1994; Rescorla et al., 2012, 2013, 2014), 338 studies from De Los Reyes and colleagues (2015), and 46 studies on cross-informant correspondence in child and adolescent mental health assessments published between 2015 and 2016 (see Table 4). The coders exhibited "almost perfect" reliability between their judgments ($\kappa = .92$; $p < .001$), based on benchmarks for the Kappa statistic by Landis and Koch (1977). Following tests of inter-rater reliability, we resolved disagreements on country coding judgments via group discussions among the two coders and two of the study authors.

Data-analytic plan. We carried out a data-analytic plan consistent with recent meta-analytic work on related issues (De Los Reyes et al., 2015). Specifically, for research aims involving cross-informant correspondence, we examined published or calculated *r*s to estimate the precision of the mean for all included studies, using Comprehensive Meta-analysis Version 2 (Biostat, Englewood, NJ, n.d.) software. Studies included in the meta-analyses varied in methodology and design. Thus, we calculated a random-effects model. In addition, for some studies and samples, we observed multiple effect sizes (e.g., multiple effect sizes reported for internalizing vs. externalizing reports). We accounted for this nesting in the data by calculating (a) effect sizes and variances for each cohort or sample, and then (b) an overall effect size using a weighted mean for each study cohort (i.e., effect sizes drawn from the same sample). We based these weights on the inverse of the total variance associated with each of the data points. Lastly, we computed a weighted mean of the effect sizes for each of the study cohorts, which were based on both within-cohort error and between-cohort variance, to produce an overall summary effect

(Borenstein, Hedges, Higgins, & Rothstein, 2009). This method allowed us to capitalize on the multiple sources of variance present both within and across studies, rather than alternative methods that would have resulted in lost sources of variance (e.g., taking a simple average of correspondence estimates for a study that included multiple mental health domains).

In our data-analytic plan, we also addressed the issue of statistical stability. Specifically, given the possibility of publication bias (i.e., significant findings are more likely to be published), we calculated Orwins' Fail-Safe *N* (Borenstein et al., 2009), which provides an index of the number of data points necessary to make the overall effect size trivial (i.e., defined here as an *r* of .10).

After determining the overall effect size across studies, we considered heterogeneity of this effect size across studies. We then examined the *Q* statistic, which is based on the weighted sum of squares for each "level" or country and thus provides an index of dispersion (Borenstein et al., 2009; Higgins, Thompson, Deeks, & Altman, 2003), to determine whether the magnitude of correspondence differed across countries. We also took a random-effects modeling approach to examine both the relation between magnitude of cross-informant correspondence and cultural tightness. Further, to examine whether the significance of this relation was consistent across cross-informant reports of child and adolescent externalizing and internalizing problems, we conducted ancillary, subgroup analyses for studies coded for cross-informant correspondence estimates of child and adolescent externalizing and internalizing problems. For these analyses, we calculated unstandardized betas (*b*), *SE* of *b*, and associated *p* values (Borenstein et al., 2009).

Results and Discussion

Frequencies of countries in prior meta-analytic research. In Table 4 we report the country representation of studies examined in previous meta-analyses. A key observation is that

the first 25 years of research on cross-informant correspondence in child and adolescent mental health assessments meta-analyzed by Achenbach and colleagues (1987) consisted of research based on samples taken from participants in seven countries. In the next 25 years of research on cross-informant correspondence in reports taken from child and adolescent mental health assessments (De Los Reyes et al., 2015), we observed an over four-fold increase in the countries from which participants were sampled, relative to data from Achenbach and colleagues (1987).

Descriptive statistics for cultural tightness and magnitudes of cross-informant correspondence across countries. In Table 5, we report descriptive statistics for cultural tightness and magnitudes of cross-informant correspondence across studies in our sample. Consistent with the countries coded from previous meta-analyses (Table 4), we observed the largest number of studies from North America. That being said, our meta-analysis included studies from countries that traversed six continents, indicating that intersubjectivity research in assessments of mental health has received considerable international attention. Thus, the heterogeneity in cultural representation suggests that we were well-positioned to address our research aims. Further, cultural tightness also varied considerably in our sample, ranging from a low (i.e., relatively culturally loose) score of 3.1 (Israel) to a high (i.e., relatively culturally tight) score of 10.4 (Singapore) (Table 5). Interestingly, we observed significant heterogeneity in magnitudes of cross-informant correspondence, Q (36) = 325.80, $p < .001$, supporting tests of relations between cultural tightness and cross-informant correspondence reported below. Further, the r for cross-informant correspondence was in line with prior work (i.e., .28; Achenbach et al., 1987; De Los Reyes et al., 2015), the Fail-Safe N was 608, and the high-point of the 95% confidence interval was in the low-to-moderate range (i.e., .31; Table 6). These data indicate that even with significant heterogeneity across countries, the magnitude of correspondence never rose

to a level indicating that informants' subjective reports were redundant with one another. Taken together, these cross-cultural data indicate that studies of intersubjectivity in mental health consistently reveal low correspondence among multiple informants' mental health reports.

Relation between cultural tightness and cross-informant correspondence. In Table 6, we also report findings from our total study sample on the relation between cultural tightness and cross-informant correspondence. Consistent with our expectations, increased cultural tightness was related to increased cross-informant correspondence. Stated another way, countries with relatively strong cultural norms and relatively low tolerance of deviations from these norms tended to exhibit relatively higher levels of correspondence among informants' subjective reports about mental health, relative to countries with relatively weak cultural norms and relatively high tolerance of deviation from these norms. These findings indicate that the significant heterogeneity we observed in magnitudes of cross-informant correspondence reflected, in part, meaningful international variation in cultural norms. Further, we continued to observe this same significant relation in separate tests for cross-informant reports of externalizing and internalizing problems (i.e., greater cultural tightness relates to greater cross-informant correspondence). Thus, the relation between cultural tightness and cross-informant correspondence tended to manifest similarly when accounting for the type of mental health domain about which informants provided reports (i.e., accessibility of intersubjectivity effects).

Research and Theoretical Implications

Our meta-analytic findings have important implications for research on intersubjectivity effects, which could greatly inform use of subjective reports in each of the sub-disciplines in Psychology that we describe below.

Clinical and Developmental Psychology

In our meta-analysis, multiple informants' mental health reports yielded consistently low-to-moderate levels of between-informant correspondence across cultures (Tables 5 and 6). Further, not only did we observe links between cross-informant correspondence and meaningful cultural variations across countries (i.e., cultural tightness), this relation generalized across tests for different mental health domains (i.e., externalizing and internalizing; Table 6). In light of these findings, a key issue for future research in Clinical and Developmental Psychology is whether the mechanisms underlying low-to-moderate between-informant correspondence manifests across cultures. Indeed, as mentioned previously, current theoretical work posits that multi-informant assessments display low-to-moderate between-informant correspondence because (a) people being assessed vary in their behaviors depending on the specific contexts in which they display mental health concerns (e.g., home vs. school/work vs. peer interactions) and (b) informants vary in the contexts in which they have opportunities for observing displays of mental health concerns (e.g., parent at home vs. teacher at school) (De Los Reyes et al., 2013a). Further, multiple recent studies conducted in the United States support this theoretical work (Table 2), and thus provide examples for how to conduct theory-driven tests of intersubjectivity.

Yet, it remains unclear if the same processes or factors account for low-to-moderate correspondence in mental health assessments conducted in other cultures. Carrying out this line of work will require not only translation and administration of multi-informant mental health assessments, as done with some widely used mental health checklists (e.g., Achenbach System of Empirically Based Assessments; Rescorla et al., 2012, 2013, 2014), but also translations of controlled behavioral observations that focus on identifying contextual variations in displays of clinically relevant behaviors (e.g., Beidel, Rao, Scharfstein, Wong, & Alfano, 2010; Wakschlag et al., 2010). We encourage future cross-cultural research on the processes underlying

intersubjectivity effects.

Work on intersubjectivity effects may also inform interpretations of subjective reports within specialty areas of Clinical Psychology, such as Forensic Psychology. In particular, researchers and practitioners often encounter great difficulty with interpreting divergence among reports about child mental health taken for the purposes of child maltreatment evaluations (for a review, see Romano, Weegar, Babchishin, & Saini, 2018). Indeed, in these evaluations, divergence between reports could arise, for instance, because of extraneous factors such as a motivation on the part of an informant to report socially desirable behaviors in an effort to influence a key outcome (e.g., custody placement). Alternatively, divergence between reports could arise because of meaningful differences in displays of mental health concerns across key contexts. For example, two parents might provide diverging reports of their child's mental health, in part, because risk for mental health concerns manifest differently based on the child's specific care environment. Paradigms informed by the Operations Triad Model (Figure 1) may assist in teasing apart these two kinds of possibilities (see De Los Reyes et al., 2009, 2016). That is, by comparing patterns of subjective reports with data taken from non-subjective sources (e.g., observations of parenting, official records), future research might leverage methods informed by the Operations Triad Model for the purposes of informing assessments conducted within forensic evaluations for child maltreatment or custody. We encourage future work that applies the Operations Triad Model to other specialty areas in Clinical Psychology.

Personality Psychology

Research on intersubjectivity in mental health research may greatly inform use and interpretation of subjective measures in sub-disciplines other than Clinical and Developmental Psychology. For instance, research in personality often involves understanding the form and

function of adulthood traits (e.g., agreeableness, aggression, neuroticism, narcissism) and their links to such domains as interpersonal or intergroup relations and mental health (Hunsley & Lee, 2014). Historically, this work has largely relied on subjective self-reports to assess personality, perhaps because researchers previously have questioned the value of collateral informants (e.g., spouse, co-worker) in terms of cost, psychometric properties, incremental value (i.e., relative to self-reports), and compliance with completing reports that assess others' personality traits (for a review, see Vazire, 2006). Importantly, a great deal of research in these areas now indicates that collateral informants' personality reports (a) can provide cost-efficient, reliable, and valid information about others' personality traits; (b) can be administered via multiple formats (e.g., in-person clinical assessments, online assessments); and (c) predict variance in external criterion variables over-and-above self-report personality measures (e.g., Balsis, Cooper, & Oltmanns, 2015; Galione & Oltmanns, 2013; Vazire & Mehl, 2008). Yet, with some exceptions (cf. Oltmanns & Turkheimer, 2009), current research in this area remains focused on how particular subjective assessments of personality (e.g., collateral informants) "outperform" subjective assessments from other sources (i.e., self-report), or in cases when discrepancies exist between subjective reports, identifying which report is "most valid" (e.g., Vazire, 2010; Vazire & Carlson, 2011). This work largely mirrors work discussed at the beginning of this paper on distinguishing the relative value of information sources, rather than examining sources collectively via the lens of intersubjectivity.

In many respects, the state of multi-informant assessments in adult personality research shares commonalities with the progression of multi-informant assessments in mental health research, particularly with regard to work with children and adolescents. Indeed, in early studies of child and adolescent mental health, information sources deemed impractical for use in this

work became commonplace over time once the empirical data supported their use (e.g., children's self-reports; Achenbach et al., 1987; De Los Reyes et al., 2013a, 2015; Hunsley & Mash, 2007; Lapouse & Monk, 1958). Further, as mentioned previously, much of the research on multi-informant assessments in mental health tests whether the links between informants' subjective reports (or lack thereof) reflect meaningful phenomena (e.g., variations in behavior within and across contexts), over-and-above individual subjective reports (Achenbach, 2011; De Los Reyes et al., 2013a, 2015; Kraemer et al., 2003). In this respect, mental health research has embraced key concepts underlying intersubjectivity (Gillespie & Cornish, 2010).

An interesting direction for future work on adult personality involves testing whether examining personality reports through the lens of intersubjectivity informs an understanding of the intersection between personality traits and the social contexts that elicit these traits. For instance, multi-informant assessments of personality vary in that low-to-moderate degrees of correspondence tend to characterize comparisons between self-reports and reports from collateral informants (e.g., Clifton, Turkheimer, & Oltmanns, 2004, 2005; Cooper, Balsis, & Oltmanns, 2012; South, Oltmanns, Johnson, & Turkheimer, 2011), whereas studies tend to yield relatively high correspondence between collateral informants' reports about the same person (for a review, see Klonsky, Oltmanns, & Turkheimer, 2002). As with subjective reports in mental health research and the Operations Triad Model, these variations in correspondence rates could manifest for multiple reasons, such as meaningful variations in displays of personality traits or that one or more of the reports evidences relatively poor reliability and/or validity (see Figure 1).

In support of instances in which low correspondence signals meaningful variation in displays of personality (i.e., diverging operations), consider theory and research that posits that the correspondence among trait assessments depends, in part, on the "match" among the traits

displayed by the individual, the likelihood that exposure to a given social context elicits displays of the trait, and the capacity of a given assessment approach (e.g., self- or collateral-report) to yield scores reflecting displays of the trait in this context (e.g., Hartley et al., 2011; Mischel & Shoda, 1995). Through this lens of personality assessments, consider a married, adult male who is employed in a large business and who is undergoing a personality assessment. The assessment involves taking multiple subjective reports of his aggressive disposition. What if he carries an aggressive disposition, but only manifests signs of aggression within social contexts typified by competition among peers for monetary resources (e.g., upward mobility in the organization)? Further, what if signs of aggression were unlikely to manifest in non-work contexts, even in the presence of competitive contexts that did not involve money (e.g., game of basketball with relatives at a social gathering)? Here, such a circumstance might result in a high degree of correspondence between self-reports and work-based collateral reports of the employee's aggressive disposition, but low correspondence between self-reports and family-based collateral reports of this disposition. These circumstances may be greatly informed by the Operations Triad Model and in particular from empirical efforts to identify circumstances in multi-informant assessments where low correspondence between reports meaningfully reflects context-based changes in behavior (e.g., De Los Reyes et al., 2009, 2013e, 2016; Deros et al., 2018; Glenn et al., 2019; Lerner et al., 2017). A particularly fruitful area of research may involve examining whether variations in correspondence levels among self-reports and collateral reports of personality traits relate to variations in displays of these traits on independent assessments (e.g., home and work observations of employee made by trained raters).

Importantly, as with intersubjectivity effects in mental health research, not all instances of low correspondence among reports on personality assessment will reflect meaningful

variations in displays of personality traits. Indeed, prior work indicates that low correspondence sometimes reflects differences among the psychometric properties of reports (i.e., compensating operations). For instance, when comparing the psychometric properties of self- and informant-reported personality assessments, informant-reported assessments tend to display greater internal consistency and mean inter-item correlation estimates, relative to self-reports (Balsis et al., 2015). Importantly, these differences in internal consistency estimates may play a role in informant-reported personality explaining more variance in external criterion measures, relative to self-reports. This work indicates that sometimes the relatively low correspondence often observed between self- and informant-reports on these personality assessments might be parsimoniously explained by compensating operations. Overall, these two examples highlight the potential for theory and research on intersubjectivity effects in mental health research to inform interpretations of multi-informant assessments of personality, and we encourage future work on these issues.

Social Psychology

Our review also has important implications for research on the replicability of research in other sub-disciplines of Psychology. Indeed, recent work highlights the challenges that researchers in Social Psychology may face to replicate studies published in top-tier journals. The Open Science Collaboration (2015) examined 100 studies seeking to replicate experimental or correlational studies published in *Psychological Science*, *Journal of Personality and Social Psychology*, and *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Whereas 97% of the original studies reported statistically significant effects, 36% of the replication studies yielded statistically significant effects in the direction hypothesized in the original studies, and 47% yielded effect sizes within the 95% confidence interval of effect sizes

observed in the original studies.

A key finding of the Open Science Collaboration (2015) was that increased magnitude of effects observed in the original study (i.e., magnitude of effect size and statistical significance results) predicted a greater likelihood that the effects of subsequent studies would replicate the original study. This is not a surprising result. Indeed, the scientific research enterprise tends to both incentivize the publication of innovative, novel findings and not only discourages replication studies but also the publication of studies yielding null findings (e.g., Borenstein, Hedges, Higgins, & Rothstein, 2009; Ioannidis, 2005). Thus, among a distribution of studies already likely biased toward reporting statistically significant, strong effects, the strongest effects ought to be more likely to replicate in subsequent research.

Examining the characteristics of studies about intersubjectivity in the mental health field may inform our broader understanding of factors that give rise to replicable findings in research more generally. That is, an interesting direction for future research might involve studying the mechanisms that underlie the replicability of intersubjectivity effects. We can think of at least two candidate mechanisms. First, studies of cross-informant correspondence in mental health tend to be psychometric investigations that use non-experimental designs. Psychometric studies often involve recruiting relatively large samples, and thus potentially evincing high statistical power, relative to samples recruited for experimental research in Social Psychology.

Second, intersubjectivity research may be less susceptible to publication biases. As mentioned previously, publication biases tend to favor studies reporting novel, statistically significant effects. In contrast, much of the research on intersubjectivity comes from secondary analyses of existing datasets; in fact, meta-analyses of intersubjectivity effects often involve coding studies whose aims had little to do with intersubjectivity (see Achenbach et al., 1987; De

Los Reyes et al., 2015). Further, the modal finding in intersubjectivity research is low-to-moderate correspondence between informants' subjective reports, which by definition involves examining relatively weak, not strong, effects. Thus, in this work, researchers are not penalized for seeking to publish results that yielded relatively low-magnitude effects (i.e., low-to-moderate between-informant correspondence): These effects are considered normative in work on intersubjectivity. Perhaps the most striking aspect of intersubjectivity research is that the counterfactual finding (i.e., observing large correspondence between informants' subjective reports) carries with it all the characteristics of effects that would be interpreted with intrigue, not the skepticism typically associated with null effects. Overall, studying the replicability of intersubjectivity effects in mental health research may be an important resource for research on understanding and improving replicability in psychological research more generally.

Quantitative Psychology

Intersubjectivity research may inform future work aiming to understand factors that contribute to low convergence of psychological research findings. A key component of intersubjectivity research deals with the implications that researchers ascribe to low convergence of research findings. The theoretical and empirical work reviewed previously clearly highlights that researchers who study intersubjectivity effects do not hold negative views about low convergence in research findings. In fact, these researchers make active attempts to understand *why* informants' subjective reports display low levels of convergence (Figures 1 and 3) – and consider that, paradoxically, this in fact represents a robust, interpretable finding unto itself (Table 6). To accomplish these goals, intersubjectivity studies should not only incorporate multiple informants and methods but also include informants and methods that will *likely* yield low levels of convergence (e.g., Kraemer et al., 2003).

We suspect that study design practices in intersubjectivity research result in researchers holding a relatively more open stance to not only data collection approaches but also practices in open reporting of scientific findings. We suspect that such an approach to interpreting scientific findings may reduce problems stemming from such practices as selective reporting of favorable outcomes (for reviews, see De Los Reyes et al., 2011; Ioannidis, 2005) or interpreting non-significant outcomes in a favorable light (e.g., Boutron, Dutton, Ravaud, & Altman, 2010). That is, if investigators actively study factors that give rise to low convergence of research findings in an area, they may be more likely to create a climate of exhaustive reporting of research findings.

An exciting implication of intersubjectivity research is that methodological strategies developed to understand and interpret subjective reports may inform the development of general methods for identifying factors that influence levels of replicability. In fact, the theoretical work discussed in this paper (Figures 1 and 3), coupled with practices developed to strategically select information sources for providing subjective reports (Kraemer et al., 2003), may yield insights on new ways for developing lines of research on replicability of research findings. For instance, as mentioned previously, informants' subjective reports can systematically vary as to the contexts in which they observe those being rated (e.g., parent observes child at home vs. teachers observe the same child at school). An insightful observation made by Kraemer and colleagues (2003) is that if low convergence among informants' reports is typical, then researchers ought to select informants whose reports systematically diverge with each other. Indeed, strategically "mixing and matching" informants in an assessment allows one to statistically decompose scores from the informants' reports into separable, predictable components. For example, a principal components analytic approach developed by Kraemer and colleagues (2003) decomposes multi-informant assessments into separable components that reflect rated behaviors that manifest (a)

consistently across informants (i.e., trait scores), (b) specifically within a particular context (e.g., home) and outside of that context (e.g., non-home such as school; i.e., context scores), and (c) specifically within the perspective of a particular kind of informant (e.g., child self-report vs. observer report provided by parent or teacher; i.e., perspective scores). In essence, the mixing and matching approach advanced by Kraemer and colleagues (2003) uses low convergence as a tool for understanding informants' reports, and in a way that facilitates understanding and interpreting child and adolescent mental health concerns.

This mixing and matching approach (Kraemer et al., 2003) readily informs future research on the factors that give rise to low replicability of research findings. Consider the dominant publication model in psychological research. This model tends to follow a "replication and extension" approach to empirical inquiry that places principal value on innovation in empirical work (e.g., Nosek, Spies, & Motyl, 2012). Within this approach, researchers seek to make incremental contributions to scientific knowledge by replicating effects observed in prior work, but with a "twist." In a prototypical case, a research team observes findings from a previous study or set of studies, and conducts a follow-up study that addresses the same or similar research aims as in prior work, but with some variation (e.g., study design, measurement, sampling characteristics) in the approach to scientific inquiry. A key goal of this model is to produce publishable work that contributes findings informed by scientific precedent (i.e., replication), that at the same time are not wholly redundant with results from a previous study or studies (i.e., extension).

Yet, a key implication of the findings from the Open Science Collaboration (2015) is that even "pure replications" of prior published work yield relatively low rates of replicability. A logical extension of this observation is that in all likelihood, replication and extension studies

may yield even lower replicability rates given variations in how different investigators can approach similar research questions (see also Silberzahn & Uhlmann, 2015). Further, some have advanced tools for addressing low replicability of research findings by modifying publication incentives and motivations for accurately and openly reporting study findings (see Nosek et al., 2012). Yet, these ideas often flow from the assumption that low rates of replicability arise largely from of a combination of measurement error, publication biases, and motivations to place a "positive spin" on research findings (see also Boutron et al., 2010; Ioannidis, 2005).

Importantly, key findings from intersubjectivity research indicate that other systematic, meaningful factors may account for at least some forms of diverging findings (Figures 1 and 3). Similarly, we cannot assume that all failed attempts at replication are created equal. For some, we should take concerted efforts to systematically reduce their occurrence; for others, we should not eliminate or discard them. Instead, we should manipulate, measure, and identify factors that give rise to them (see also Lewin, 1947). We argue that tools designed to promote openness in scientific inquiry should be complemented by active attempts to empirically examine factors that influence rates of replicability of research findings.

To draw from the approach described by Kraemer and colleagues (2003), we envision attempts to strategically implement replication and extension models of scientific inquiry. Within this approach, an investigative team conducts a study, and then other investigative teams follow up with replication and extension studies that systematically "mix and match" measurement methods, study designs, and/or informants. The key goal would be to systematically vary study features hypothesized to contribute variance in levels of replicability. For instance, researchers might test study features that increase the likelihood of yielding findings that replicate those of prior work. Other investigative teams might test study features that increase the likelihood of

failing to replicate findings of prior work. Such an approach may fruitfully inform our understanding of how decisions made by research teams impact the accumulation of scientific knowledge, and we encourage future dialogue on these issues.

Concluding Comments

In this paper, we sought to expand on prior work on intersubjectivity in Psychology by reviewing areas of research that have used a multi-informant, multi-method approach to understanding relations among multiple subjective impressions of psychological phenomena. Using this approach, researchers seek to understand the information that can be gleaned from examining where subjective impressions diverge and converge (De Los Reyes, et al., 2013a). To this end, we reviewed research and theory in the mental health field that views subjectivity from two lenses (Tables 1 and 2). First, points of convergence and divergence in subjective impressions might reflect contextual variations in behavior. This is because behavior may vary across contexts, and informants often vary in the contexts in which they observe behavior (Achenbach et al., 1987; De Los Reyes et al., 2015; Kraemer et al., 2003).

Second, mental health researchers often collect subjective data from informants, namely parents and children and adolescents, who are asked to report about their relationship and home environment. Thus, points of convergence and divergence between subjective reports might reflect crucial, covert information about participants' subjective perspectives (De Los Reyes, 2011). In turn, these subjective impressions may reveal important information about how parents and children and adolescents interact with one another, and thus predict psychological outcomes in children, adolescents, and broader family units (Laird & De Los Reyes, 2013). Further, mental health researchers have found value in studying relations between participants' subjective impressions of their own behavior, and trained evaluators' impressions of these same

participants, as these relations often also indicate contextual variations in subjective experience and behavior (De Los Reyes et al., 2013e; Deros et al., 2018; Glenn et al., 2019).

Intersubjectivity effects reflect aspects of the social environment beyond that of social context. Indeed, across all major meta-analytic reviews conducted to date (Table 4), researchers in mental health have studied issues surrounding intersubjectivity in over 30 countries. This element of prior work allowed for a novel application of cross-cultural research and theory to the study of intersubjectivity effects (Gelfand et al., 2011; Norenzayan & Heine, 2005). In fact, levels of correspondence between informants' subjective reports are remarkably consistent across reports taken in a diverse array of cultures worldwide (Tables 5 and 6). Further, levels of correspondence meaningfully relate to variations among countries and their cultural norms (i.e., cultural tightness), and this relation holds regardless of the observability of the mental health domain about which informants provide reports (i.e., externalizing and internalizing; Table 6).

In highlighting how researchers measure and interpret intersubjectivity in child, adolescent, and adult mental health research, we illustrated how incorporating subjectivity in research leads to a greater understanding of subjective reports across multiple sub-disciplines in Psychology beyond that of Clinical and Developmental, including Personality, Social, and Quantitative. Overall, we wish to promote an increased focus on the theoretical, methodological, and empirical value of intersubjectivity, and to do so in a way that informs the interpretability of psychological research findings.

References

*The articles coded for our meta-analyses (see Tables 4-6) can be found here: https://drive.google.com/file/d/0B5vTWq7f7jtdV0M3NmVXQzZROFU/view?usp=sharing

Achenbach, T. M. (2017). Future directions for clinical research, services, and training: Evidence-based assessment across informants, cultures, and dimensional hierarchies. *Journal of Clinical Child and Adolescent Psychology, 46*, 159-169. doi: 10.1080/15374416.2016.1220315

Achenbach, T.M. (2011). Commentary: Definitely more than measurement error: But how should we understand and deal with informant discrepancies? *Journal of Clinical Child and Adolescent Psychology, 40,* 80-86. doi: 10.1080/15374416.2011.533416

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

Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101,* 213-232. doi: 10.1037/0033-2909.101.2.213

Achenbach, T.M., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms and profiles*. Burlington, VT: University of Vermont Research Center for Children, Youth, and Families.

Alfano, C.A., & Beidel, D.C. (2011). *Social anxiety in adolescents and young adults: Translating developmental science into practice*. Washington, DC: American Psychological Association.doi:10.1037/12315-000

Alfano, C.A., Patriquin, M.A., & De Los Reyes, A. (2015). Subjective-objective sleep comparisons and discrepancies among clinically-anxious and healthy children. *Journal of*

Abnormal Child Psychology, 43, 1343-1353. doi: 10.1007/s10802-015-0018-7

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

Anderson, E.R., & Hope, D.A. (2009). The relationship among social phobia, objective and perceived physiological reactivity, and anxiety sensitivity in an adolescent population. *Journal of Anxiety Disorders, 23,* 18-26. doi: 10.1016/j.janxdis.2008.03.011

Balsis, S., Cooper, L.D., & Oltmanns, T.F. (2015). Are informant reports of personality more internally consistent than self reports of personality. *Assessment, 22*, 399-404. doi: 10.1177/1073191114556100

Beidel, D.C., Rao, P.A., Scharfstein, L., Wong, N., & Alfano, C.A. (2010). Social skills and social phobia: An investigation of DSM-IV subtypes. *Behaviour Research and Therapy, 48,* 992–1001. doi:10.1016/j.brat.2010.06.005

Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.

Borsboom, D. (2005). *Measuring the mind*. New York: Cambridge University Press.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111,* 1061-1071. doi: 10.1037/0033-295X.111.4.1061

Boutron, I., Dutton, S., Ravaud, P., & Altman, D.G. (2010). Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *Journal of the American Medical Association, 303*, 2058-2064. doi: 10.1001/jama.2010.651

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105. doi: 10.1037/h0046016

Card, N.A., & Hodges, E.V.E. (2008). Peer victimization among schoolchildren: Correlations, causes, consequences, and considerations in assessment and intervention. *School Psychology Quarterly, 23*, 451-461. doi: 10.1037/a0012769

Casey, R.J., & Berman, J.S. (1985). The outcomes of psychotherapy with children. *Psychological Bulletin, 98,* 388-400. doi: 10.1037/0033-2909.98.2.388

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

Clifton, A., Turkheimer, E., & Oltmanns, T.F. (2004). Contrasting perspectives on personality problems: Descriptions from the self and others. *Personality and Individual Differences, 36*, 1499-1514. doi: 10.1016/j.paid.2003.06.002

Clifton, A., Turkheimer, E., & Oltmanns, T.F. (2005). Self and peer-perspectives on pathological personality traits and interpersonal problems. *Psychological Assessment, 17*, 123-131. doi: 10.1037/1040-3590.17.2.123

Cole, D.A., Tram, J.M., Martin, J.M., Hoffman, K.B., Ruiz, M.D., Jacquez, F.M., & Maschman, T.L. (2002). Individual differences in the emergence of depressive symptoms in children and adolescents: A longitudinal investigation of parent and child reports. *Journal of Abnormal Psychology, 111,* 156-165. doi: 10.1037/0021-843X.111.1.156

Comer, J.S., & Kendall, P.C. (2004). A symptom-level examination of parent-child agreement in the diagnosis of anxious youth. *Journal of the American Academy of Child & Adolescent Psychiatry, 43,* 878-886. doi: 10.1097/01.chi.0000125092.35109.c5

Cooper, L.D., Balsis, S., & Oltmanns, T.F. (2012). Self- and informant-reported perspectives on symptoms of narcissistic personality disorder. *Personality Disorders: Theory, Research, and Treatment, 3*, 140-154. doi: 10.1037/a0026576

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302. doi: 10.1037/h0040957

De Los Reyes, A. (2011). Introduction to the special section. More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 40,* 1-9. doi: 10.1080/15374416.2011.533405

De Los Reyes, A. (2013). Strategic objectives for improving understanding of informant discrepancies in developmental psychopathology research. *Development and Psychopathology, 25,* 669-682. doi: 10.1017/S0954579413000096

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

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

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

De Los Reyes, A., Aldao, A., Thomas, S.A., Daruwala, S.E., Swan, A.J., Van Wie, M.,. . . De Los Reyes (2012). Adolescent self-reports of social anxiety: Can they disagree with

objective psychophysiological measures and still be valid? *Journal of Psychopathology and Behavioral Assessment, 34,* 308-322. doi: 10.1007/s10862-012-9289-2

De Los Reyes, A., Thomas, S.A., Goodman, K.L., & Kundey, S.M.A. (2013a). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology, 9,* 123-149. doi: 10.1146/annurev-clinpsy-050212-185617

De Los Reyes, A., Ehrlich, K.B., Swan, A.J., Luo, T., Van Wie, M., & Pabón, S.C. (2013b). An experimental test of whether informants can report about child and family behavior based on settings of behavioral expression. *Journal of Child and Family Studies, 22,* 177-191. doi: 10.1007/s10826-012-9567-3

De Los Reyes, A., Lerner, M.D., Thomas, S.A., Daruwala, S.E., & Goepel, K.A. (2013c). Discrepancies between parent and adolescent beliefs about daily life topics and performance on an emotion recognition task. *Journal of Abnormal Child Psychology, 41,* 971-982. doi: 10.1007/s10802-013-9733-0

De Los Reyes, A., Salas, S., Menzer, M.M., & Daruwala, S.E. (2013d). Criterion validity of interpreting scores from multi-informant statistical interactions as measures of informant discrepancies in psychological assessments of children and adolescents. *Psychological Assessment, 25,* 509-519. doi: 10.1037/a0032081

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

De Los Reyes, A., Henry, D.B., Tolan, P.H., & Wakschlag, L.S. (2009). Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology, 37,* 637-652. doi: 10.1007/s10802-009-9307-3

De Los Reyes, A., & Kazdin, A.E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131,* 483-509. doi: 10.1037/0033-2909.131.4.483

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

De Los Reyes, A., & Kazdin, A.E. (2008). When the evidence says, "Yes, no, and maybe so": Attending to and interpreting inconsistent findings among evidence-based interventions. *Current Directions in Psychological Science, 17,* 47-51. doi: 10.1111/j.1467-8721.2008.00546.x

De Los Reyes, A., & Kazdin, A.E. (2009). Identifying evidence-based interventions for children and adolescents using the range of possible changes model: A meta-analytic illustration. *Behavior Modification, 33,* 583-617. doi: 10.1177/0145445509343203

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

De Los Reyes, A., & Marsh, J.K. (2011). Patients' contexts and their effects on clinicians' impressions of conduct disorder symptoms. *Journal of Clinical Child and Adolescent Psychology, 40,* 479-485. doi: 10.1080/15374416.2011.563471

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

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

Deros, D.E., Racz, S.J., Lipton, M.F., Augenstein, T.M., Karp, J.N., Keeley, L.M., . . . De Los Reyes (2018). Multi-informant assessments of adolescent social anxiety: Adding clarity by leveraging reports from unfamiliar peer confederates. *Behavior Therapy, 49*, 84-98. doi: 10.1016/j.beth.2017.05.001

Dirks, M.A., De Los Reyes, A., Briggs-Gowan, M.J., Cella, D., & Wakschlag, L.S. (2012). Embracing not erasing contextual variability in children's behavior: Theory and utility in the selection and use of measures and informants in the assessment of developmental psychopathology. *Journal of Child Psychology and Psychiatry, 53,* 558-574. doi: 10.1111/j.1469-7610.2012.02537.x

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

Dunning, D., Heath, C., & Suls, J.M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5,* 69-106. doi: 10.1111/j.1529-1006.2004.00018.x

Erwin, E. (2000). Is a science of psychotherapy possible? Subjectivity problems. *American Psychologist, 55,* 1133-1138. doi: 10.1037//0003-066X.55.10.1133

Fjermestad, K.W., Lerner, M.D., McLeod, B.D., Wergeland, G.J.H., Heiervang, E.R., Silverman, W.K., Öst, L.-G., De Los Reyes, A., Havik, O.E., & Haugland, B.S.M. (2016). Therapist-youth agreement on alliance change predicts long-term outcome in CBT for anxiety

disorders. *Journal of Child Psychology and Psychiatry, 57*, 625-632. doi: 10.1111/jcpp.12485

Galione, J.N., & Oltmanns, T.F. (2013). Identifying personality pathology associated with major depressive episodes: Incremental validity of informant reports. *Journal of Personality Assessment, 95,* 625-632. doi: 10.1080/00223891.2013.825624

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

Gelfand, M.J. (2012). Culture's constraints: International differences in the strength of social norms. *Current Directions in Psychological Science* 2012 21: 420-424. doi: 10.1177/0963721412460048

Gelfand, M.J., & Jackson, J.C. (2016). From one mind to many: the emerging science of cultural norms. *Current Opinion in Psychology, 8*, 175-181. doi: 10.1016/j.copsyc.2015.11.002

Gelfand, M.J., Raver, J.L., Nishii, L., Leslie, L. M., Lun, J., Lim, B.C., . . . Yamaguchi, S. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, *332*, 1100-1104. doi: 10.1126/science.1197754

Gillespie, A., & Cornish, F. (2010). Intersubjectivity: Towards a dialogical analysis. *Journal for the Theory of Social Behaviour, 40,* 19-46. doi: 10.1111/j.1468-5914.2009.00419.x

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

Goodman, K.L., De Los Reyes, A., & Bradshaw, C.P. (2010). Understanding and using

informants' reporting discrepancies of youth victimization: A conceptual model and recommendations for research. *Clinical Child and Family Psychology Review*, *13,* 366-383. doi: 0.1007/s10567-010-0076-x

Gough, B., & Madill, A. (2012). Subjectivity in psychological science: From problem to prospect. *Psychological Methods, 17,* 374-384. doi: 10.1037/a0029313

Greiffenhagen, C., & Sharrock, W. (2008). Where do the limits of experience lie? Abandoning the dualism of objectivity and subjectivity. *History of the Human Sciences, 21,* 70-93. doi: 10.1177/0952695108093954

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

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

Higgins, J.P., Thompson, S.G., Deeks, J.J., & Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557-560. doi: 10.1136/bmj.327.7414.557

Holmbeck, G.N., Li, S.T., Schurman, J.V., Friedman, D., & Coakley, R.M. (2002). Collecting and managing multisource and multimethod data in studies of pediatric populations. *Journal of Pediatric Psychology, 27,* 5-18. doi: 10.1093/jpepsy/27.1.5

Hager, M.G. (1982). The myth of objectivity. *American Psychologist, 37,* 576-579. doi: 10.1037/0003-066X.37.5.576

Hunsley, J., & Lee, C.M. (2014). *Introduction to clinical psychology* (2nd ed.). Hoboken: Wiley & Sons.

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

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

Ioannidis, J.P.A. (2005). Why most published research findings are false. *PLOS Medicine, 2*, e124. doi: 10.1371/journal.pmed.0020124

Kazdin, A.E. (2003). *Research design in clinical psychology* (4th ed.). Boston: Allyn & Bacon.

Klonsky, E. D., Oltmanns, T. F., & Turkheimer, E. (2002). Informant-reports of personality disorder: Relation to self-reports and future research directions. *Clinical Psychology: Science and Practice, 9*, 300-311. doi: 10.1093/clipsy.9.3.300

Kraemer, H.C., Measelle, J.R., Ablow, J.C., Essex, M.J., Boyce, W.T., & Kupfer, D.J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry, 160,* 1566-1577. doi: 10.1176/appi.ajp.160.9.1566

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

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174. doi:10.2307/2529310

Lapouse, R., & Monk, M.A. (1958). An epidemiologic study of behavior characteristics in children. *American Journal of Public Health, 48,* 1134-1144. doi: 10.2105/AJPH.48.9.1134

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

doi: 10.1111/jcpp.12730

Lewin, K. (1947). Frontiers in group dynamics: Concept, method and reality in social science; social equilibria and social change. *Human Relations, 1*, 5-41. doi: 10.1177/001872674700100103

Lippold, M.A., Greenberg, M.T., & Collins, L.M. (2013). Parental knowledge and youth risky behavior: A person oriented approach. *Journal of Youth and Adolescence, 42*, 1732-1744. doi: 10.1007/s10964-012-9893-1

Lippold, M.A., Greenberg, M.T., & Collins, L.M. (2014). Youths' substance use and changes in parental knowledge-related behaviors during middle school: A person-oriented approach. *Journal of Youth and Adolescence, 43*, 729-744. doi:10.1007/s10964-013-0010-x

Marsh, J.K., & De Los Reyes, A. (2018). Explaining away disorder: The influence of context on impressions of mental health symptoms. *Clinical Psychological Science, 6*, 189-202. doi: 10.1177/2167702617709812

Marsh, J.K., Burke, C.T., & De Los Reyes, A. (2016). The sweet spot of clinical intuition: Predictors of the effects of context on impressions of conduct disorder symptoms. *Psychological Assessment, 28*, 181-193. doi: 10.1037/pas0000173

Marsh, J.K., De Los Reyes, A., & Wallerstein, A. (2014). The influence of contextual information on lay judgments of childhood mental health concerns. *Psychological Assessment, 26,* 1268-1280. doi: 10.1037/pas0000012

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102,* 246-268. doi: 10.1037/0033-295X.102.2.246

Norenzayan, A., & Heine, S.J. (2005). Psychological universals: What are they and how can we

know? *Psychological Bulletin, 131,* 763-784. doi: 10.1037/0033-2909.131.5.763

Nosek, B.A., Spies, J.R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7*, 615-631. doi: 10.1177/1745691612459058

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, aac4716. doi: 10.1126/science.aac4716

Oltmanns, T.F., & Turkheimer, E. (2009). Person perception and personality pathology. *Current Directions in Psychological Science, 18,* 32-36. doi: 10.1111/j.1467-8721.2009.01601.x

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

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

Rescorla, L.A., Ginzburg, S., Achenbach, T.M., Ivanova, M.Y., Almqvist, F., Begovac, I., . . . Verhulst, F.C. (2013). Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. *Journal of Clinical Child and Adolescent Psychology, 42,* 262-273. doi: 10.1080/15374416.2012.717870

Rescorla, L.A., Achenbach, T.M., Ivanova, M.Y., Bilenberg, N., Bjarnadottir, G., Denner, S., . . . Verhulst, F.C. (2012). Behavioral/emotional problems of preschoolers: Caregiver/ teacher reports from 15 societies. *Journal of Emotional and Behavioral Disorders, 20*, 68-81. doi: 10.1177/1063426611434158

Richters, J.E. (1992). Depressed mothers as informants about their children: A critical review of

the evidence for distortion. *Psychological Bulletin, 112,* 485-499. doi: 10.1037/0033-2909.112.3.485

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

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54,* 93-105. doi: 10.1037/0003-066X.54.2.93

Shapin, S. (2012). The sciences of subjectivity. *Social Studies of Science, 42,* 170-184. doi: 10.1177/0306312711435375

Silberzahn, R., & Uhlmann, E.L. (2015). Many hands make tight work. *Nature, 526*, 189-191. doi: 10.1038/526189a

Silverman, W. K., & Ollendick, T. H. (2005). Evidence-based assessment of anxiety and its disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *34*, 380-411. doi: 10.1207/s15374424jccp3403_2

Smetana, J. G. (2008). ‘‘It’s 10 o’clock: Do you know where your children are?’’ Recent advances in understanding parental monitoring and adolescents’ information management. *Child Development Perspectives, 2,* 19-25. doi: 10.1111/j.1750-8606.2008.00036.x

South, S.C., Oltmanns, T.F., Johnson, J., & Turkheimer, E. (2011). Level of agreement between self and spouse in the assessment of personality pathology. *Assessment, 18*, 217-226. doi: 10.1177/1073191110394772

Taber, S.M. (2010). The veridicality of children's reports of parenting: A review of factors contributing to parent–child discrepancies. *Clinical Psychology Review, 30,* 999-1010.

doi: 10.1016/j.cpr.2010.06.014

Thomas, S.A., Aldao, A., & De Los Reyes, A. (2012). Implementing clinically feasible psychophysiological measures in evidence-based assessments of adolescent social anxiety. *Professional Psychology: Research and Practice, 43,* 510-519. doi: 10.1037/a0029183

van der Ende, J., Verhulst, F.C., & Tiemeier, H. (2012). Agreement of Informants on emotional and behavioral problems from childhood to adulthood. *Psychological Assessment, 24,* 293-300. doi: 10.1037/a0025500

Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality, 40*, 472-481. doi: 10.1016/j.jrp.2005.03.003

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology, 98*, 281-300. doi: 10.1037/a0017908

Vazire, S., & Carlson, E.N. (2011). Others sometimes know us better than we know ourselves. *Current Directions in Psychological Science, 20*, 104-108. doi: 10.1177/0963721411402478

Vazire, S., & Mehl, M.R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology, 95*, 1202-1216. doi: 10.1037/a0013314

Wakschlag, L.S., Tolan, P.H., & Leventhal, B.L. (2010). Research review: 'Ain't misbehavin': Towards a developmentally-specified nosology for preschool disruptive behavior. *Journal of Child Psychology and Psychiatry, 51,* 3-22. doi: 10.1111/j.1469-7610.2009.02184.x

Weisz, J.R., Jensen Doss, A., & Hawley, K.M. (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annual Review of Psychology, 56*, 337-363. doi: 10.1146/annurev.psych.55.090902.141449

Weisz, J.R., Weiss, B., Alicke, M.D., & Klotz, M.L. (1987). Effectiveness of psychotherapy with children and adolescents: A meta-analysis for clinicians. *Journal of Consulting and Clinical Psychology, 55,* 542-549. doi: 10.1037/0022-006X.55.4.542

Table 1. Key articles on the ubiquity and stability of divergence between multiple informants' subjective reports of psychological behaviors

Author Group (Year)	Methodology	Cross-Informant Correspondence Findings
Quantitative Reviews		
Achenbach, McConaughy, & Howell (1987)	First large-scale meta-analysis of cross-informant correspondence of reports of child mental health; 119 studies spanning over 25 years (1960-1986).	Across multiple informants (parent, teacher, self-report, mental health care worker, peer, trained observer), correspondence ranged from low-to-moderate in magnitude (*r*s in .20s-.60s).
Achenbach, Krukowski, Dumenci, & Ivanova (2005)	First large-scale meta-analysis of cross-informant correspondence of reports of adult mental health; 108 studies spanning over ten years (1993-2003).	Across multiple informants (self-report; clinician; spouse, partner, or other family members; peers), correspondence magnitudes were slightly higher than those for children, but still moderate in magnitude (*r*s in .40-.60s).
Casey & Berman (1985)	First large-scale meta-analysis of the outcomes of psychotherapy studies of children; 75 studies spanning over 30 years (1952-1983)	Across multiple informants (self-report, teacher, peer, parent, trained observer), effects of psychotherapy varied from small-to-large in magnitude (*d*s in .10s to above 1.0).
De Los Reyes et al. (2015)	Second large-scale meta-analysis of cross-informant correspondence of reports of child mental health; 341 studies spanning over 25 years (1989-2014).	Across multiple informants (parent, teacher, self-report), correspondence ranged from low-to-moderate in magnitude (*r*s in .20s to .50s).
Duhig, Renk, Epstein, & Phares (2000)	Meta-analysis of cross-informant correspondence of mother and father reports of child mental health; 60 studies spanning nearly 10 years (1990-1997).	Across multiple domains (internalizing, externalizing, broadband total problems) mother-father correspondence ranged from moderate-to-large in magnitude (*r*s in .40s-.70s).
Renk & Phares (2004)	First large-scale meta-analysis of cross-informant correspondence of reports of child social competence; 74 studies spanning over 60 years (1933-1997).	Across multiple informants (parent, teacher, self-report, peer), correspondence ranged from low-to-moderate in magnitude (*r*s in .20s-.40s).
Weisz, Weiss, Alicke, & Klotz (1987)	Second large-scale meta-analysis of the outcomes of psychotherapy studies of children; 108 studies that included studies published both in and after Casey and Berman (1985) and carefully selected via stringent criteria for study quality (e.g., use of a control group).	Across multiple informants (self-report, teacher, peer, parent, trained observer), effects of psychotherapy varied from small-to-large in magnitude (*d*s in .30s to above 1.0).
Cross-Cultural and Developmental Studies		
Cole et al. (2002)	6-year, 12-wave longitudinal study of individual differences in emergence of child depressive symptoms; 1,570 participants assessed via parent and child reports.	Parent-child correspondence on level of symptoms was moderate in magnitude (*r*s in .40s to .50s); correspondence on rate of change in symptoms was moderate-to-large in magnitude (*r*s. in .50s to .70s).
Rescorla et al. (2013)	Large-scale cross-cultural study of parent-adolescent correspondence in reports of adolescent mental health; 27,861 participants assessed across 25 societies	In all but one society, adolescents self-reported significantly greater concerns than parents reported about adolescents, correspondence across societies ranged from low-to-moderate in magnitude (*r*s .10s to .50s).
Rescorla et al. (2014)	Large-scale cross-cultural study of parent-teacher correspondence in reports of child mental health; 27,962 participants assessed across 21 societies.	In all but one society, parents reported significantly greater concerns in children than teachers reported about children, correspondence across societies ranged from low-to-moderate in magnitude (*r*s of .09 to .40s).
van der Ende, Verhulst, & Tiemeier (2012)	Large-scale developmental study of cross-informant correspondence in reports of mental health; 1,875 participants assessed over 7 waves and 24 years.	Across multiple informants (self-report; parent; teacher, partner), correspondence across development (participant age range of 4 to 40 years) ranged from low-to-moderate in magnitude (*r*s in .30s to .50s).

Table 2. Examples of intersubjectivity studies of participant-participant relations in mental health research

Author Group (Year)	Psychological Behavior	Participant Informants	Validating Measures	Main Findings
De Los Reyes, Henry, Tolan, & Wakschlag (2009)	Preschooler's Disruptive Behavior	Parent and Teacher	Observed Disruptive Behavior	Diverging parent and teacher reports reflected child disruptive behavior within some adult-child interactions and not others (e.g., if parent reported disruptive behavior that teacher did not report, this related to observing disruptive behavior in parent-child and not examiner-child interactions).
Lerner, De Los Reyes, Drabick, Gerber, & Gadow (2017)	Childhood Autism Spectrum Symptoms	Parent and Teacher	Observed Autism Spectrum Concerns	Relative to parent-teacher dyads that diverged in their reports of patients' autism spectrum symptoms, parent-teacher dyads that converged in reporting relatively high or relatively moderate autism spectrum symptoms had patients who displayed relatively greater severity in autism spectrum symptoms within structured observations of patients.
De Los Reyes et al. (2013b)	Parental Knowledge of Adolescent's Whereabouts	Parent and Adolescent	Experimental Manipulations of Assessment Training	Parents and adolescents could be trained to use contextual information when providing likert-type ratings of parental knowledge of adolescents' whereabouts and activities.
De Los Reyes, Lerner, Thomas, Daruwala, & Goepel (2013c)	Discrepant Views about the Family	Parent and Adolescent	Performance-Based Task of Emotion Recognition	Lower parent and adolescent performance on an emotion recognition task significantly related to greater parent and adolescent perceived discrepant views about the family.
De Los Reyes, Salas, Menzer, & Daruwala (2013d)	Parental Monitoring of Adolescents' Whereabouts	Parent and Adolescent	Structured Interview	Parents and adolescents exhibited greatest convergence between their reports of parental monitoring (i.e., reports of parental knowledge of and adolescent disclosure about adolescent's whereabouts) when exhibiting low levels of discrepant views about daily life events via structured interview (e.g., completing household chores and homework).
De Los Reyes, Alfano, Lau, Augenstein & Borelli. (2016)	Adolescent Mental Health	Parental Caregivers	Observed Adolescent Hostility	Relative to caregiver dyads that diverged in their reports of adolescent mental health, caregiver dyads that converged in reporting relatively high adolescent mental health concerns had adolescents who evidenced greater levels of hostility within structured interactions with caregivers.
Laird & De Los Reyes (2013)	Parent-Adolescent Relationship Quality	Parent and Adolescent	Adolescent Depressive Symptoms	Convergence between parent and adolescent reports indicating positive parent-adolescent relationship quality predicts low levels of adolescent depressive symptoms, relative to other reporting combinations.

Table 3. Examples of intersubjectivity studies of participant-researcher relations in mental health research

Author Group (Year)	Psychological Behavior	Participant Informants	Researcher Informants	Validating Measures	Main Findings
De Los Reyes et al. (2012)	Adolescent Social Anxiety	Adolescent	Physiology	Clinical Referral Status	Even when adolescents' self-reported social anxiety did not correspond with objective psychophysiological indices, self-reports validly distinguished adolescents assessed via clinical referral from adolescents assessed as part of a community based study.
De Los Reyes, Bunnell, & Beidel (2013e)	Adult Social Anxiety	Patient	Clinician	Observed Social Skills	Converging patient and clinician reports reflected patients displaying social skills deficits across social interaction tasks (e.g., one-on-one interactions and public speaking).
Deros et al. (2018)	Adolescent Social Anxiety	Adolescent and Parent	Unfamiliar Peer Confederates	Adolescent Self-Reported State Arousal	Adolescents' self-reports correlated with reports on parallel measures from parents in the .30s and with peer confederates in the .40s-to-.50s, whereas reports from parent-confederate dyads correlated in the .07-to-.22 range. Adolescent and peer confederate (but not parent) reports predicted adolescents' state arousal in social interactions with peer confederates.
Fjermestad et al. (2016)	Childhood Anxiety	Patient	Clinician	Treatment Outcomes	Higher levels of agreement on change in patient and clinician reports of therapeutic alliance early-to-late in treatment predicted loss of diagnosis and reduction in clinical severity at 1-year follow-up.

Table 4. Frequency counts of countries represented in previous meta-analyses of cross-informant correspondence

	Achenbach et al. (1987) (*n*=119)	**Duhig et al. (2000) (*n*=57)**	**Renk & Phares (1994) (*n*=74)**	**Achenbach et al. (2005) (*n*=108)**		**Rescorla et al. (2012) (*n*=6)**	**Rescorla et al. (2013) (*n*=21)**		**Rescorla et al. (2014) (*n*=20)**		**De Los Reyes et al. (2015) (*n*=338)[a]**		
Countries	US(97)	US(42)	US(61)	US(72)	CA(9)	DK(1)	CN(2)	US(2)	CN(2)	US(2)	US(194)	NL(39)	CA(15)
	UK(9)	UK(5)	CA(7)	AU(5)	DE(4)	DE(1)	DZ(1)	AU(1)	HR(1)	DK(1)	UK(12)	AU(11)	CN(7)
	CA(8)	CA(4)	CN(2)	UK(4)	NZ(3)	IT(1)	HR(1)	DK(1)	FI(1)	FR(1)	NO(7)	DE(6)	IL(6)
	AU(2)	AU(3)	NL(2)	IN(2)	IT(2)	US(1)	FI(1)	DE(1)	GR(1)	IR(1)	BE(5)	FI(4)	IT(3)
	NO(1)	IL(2)	UK(2)	SE(2)	CN(1)	LT(1)	IS(1)	IR(1)	IT(1)	LT(1)	NZ(3)	RU(3)	SE(3)
	NL(1)	NL(1)		FR(1)	CH(1)	NL(1)	JM(1)	JP(1)	NL(1)	PL(1)	GR(2)	ES(2)	CH(2)
	NZ(1)			FSB(1)	UC(1)		LT(1)	NL(1)	PT(1)	SG(1)	DZ(1)	AT(1)	BR(1)
							NO(1)	PL(1)	TR(1)	UT(2)	CL(1)	DK(1)	IS(1)
							KR(1)	CH(1)	UJ(1)		JP(1)	ZA(1)	TR(1)
							TR(1)				UG(1)	AB(1)	AN(1)
											CHI(1)	UCN(1)	
Number of Unique Countries	7	6	5	13		6	19		17		30		

Note. US = United States; UK = United Kingdom; CA = Canada; AU = Australia; NL = Netherlands; NO = Norway; NZ = New Zealand; IL = Israel; CN = China; DE = Germany; IN = India; IT = Italy; FR = France; CH = Switzerland; FSB = Multi-Country (France, Switzerland, Belgium); UC = Multi-Country (United States, Canada); DK = Denmark; LT = Lithuania; DZ = Algeria; HR = Croatia; FI = Finland; IS = Iceland; IR = Iran; JM = Jamaica; JP = Japan; PL = Poland; KR = South Korea; TR = Turkey; GR = Greece; PT = Portugal; SG = Singapore; UJ = Multi-Country (United States, Jamaica); UT = Multi-Country (United States, Thailand); BE = Belgium; SE = Sweden; ES = Spain; AT = Austria; BR = Brazil; CL = Chile; ZA = South Africa; UG = Uganda; AB = Multi-Country (Australia, Brazil); AN = Multi-Country (Australia, Netherlands); CHI = Multi-Country (Czech Republic, Hungary, Italy); UCN = Multi-Country (United States, China); [a] The original study sample included the three Rescorla et al. reviews (2012, 2013, 2014), which we excluded from data reported here to reduce overlap among studies used by the reviews reported herein.

Table 5. Descriptive statistics for cultural tightness and cross-informant correspondence

Country	*N Samples*	*Cultural Tightness*	*r [95% CI]*	*Z*
North America				
United States	220	5.1	.28 [.22, .34]	8.46***
Canada	18	—	.28 [.18, .38]	5.12***
South America				
Chile	2	—	.50 [-08, .83]	1.70 (*p*=.09)
Europe				
Austria	2	6.8	.37 [.35, .39]	33.28***
Belgium	6	5.6	.41 [.31, .50]	7.34***
Finland	4	—	-.03 [-.36, .30]	-0.19 (*p*=.85)
Germany [a]	12	7	.32 [.24, .38]	8.07***
Greece	1	3.9	.48 [.43, .53]	15.13***
Iceland	1	6.4	-.19 [-.44, .10]	-1.30 (*p*=.19)
Italy	7	6.8	.21 [.11, .31]	3.97***
Netherlands	40	3.3	.30 [.17, .42]	4.43***
Norway	8	9.5	.30 [-.12, .63]	1.40 (*p*=.16)
Poland	2	6	.30 [-.02, .56]	1.85 (*p*=.06)
Portugal	2	7.8	.40 [.33, .47]	10.16***
Romania	2	—	-.17 [-.48, .18]	-0.95 (*p*=.34)
Russia	3	—	.15 [-.06, .34]	1.41 (*p*=.16)
Serbia	1	—	.37 [.29, .44]	8.75***
Spain	2	5.4	.29 [.22, .35]	7.98***
Sweden	4	—	.26 [.03, .46]	2.19*
Switzerland	2	—	.18 [-.03, .38]	1.72 (*p*=.09)
United Kingdom	9	6.9	.29 [.17, .40]	4.56***
Belgium, Netherlands [a`]	1	4.45	.63 [.49, .74]	6.91***
Czech Republic, Hungary, Italy [a]	3	4.85	.38 [.32, .44]	11.58***
Africa				
Algeria	1	—	.40 [.25, .53]	4.81***
Ghana	1	—	.07 [-.04, .19]	1.24 (*p*=.22)
South Africa	1	—	.09 [-.001, .18]	1.94 (*p*=.052)
Asia				
China	5	7.9	.36 [.11, .57]	2.80**
Israel	8	3.1	.38 [.26, .49]	5.87***
Japan	1	8.6	.29 [.19, .39]	5.28***
Singapore	2	10.4	.29 [.19, .38]	5.59***
Turkey	1	9.2	.36 [.28, .43]	8.52***
Oceania				
Australia	14	4.4	.15 [-.05, .33]	1.47 (*p*=.14)
New Zealand	3	3.9	.10 [-.004, .21]	1.89 (*p*=.06)
Intercontinental Studies				
Australia, Netherlands [a`]	2	3.85	.48 [.42, .54]	14.14***
Canada, Netherlands [`]	3	—	-.20 [-.30, -.10]	-3.93***
China, United States [a]	1	6.5	.34 [.25, .43]	6.90***
Italy, China [a]	1	7.35	.27 [.18, .35]	5.86***

Note. — = Cultural tightness scores not reported in Gelfand and colleagues (2011); [a] = For studies of multiple countries, we calculated an average cultural tightness score for the study; * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 6. Relation between cultural tightness and cross-informant correspondence in the total sample

Statistical test	*Test statistic [95% CI]*	*Z*	*N Data Points*	*N Samples*
Cross-informant correspondence: **Total Sample**	r = .28 [.24, .31]	13.89***	1560	396
Relation between magnitude of cross-informant correspondence and cultural tightness: **Total Sample**	b = 0.27, *SE* b = 0.07 [0.13, 0.42]	3.79***	1366	355
Relation between magnitude of cross-informant correspondence and cultural tightness: **Reports About Internalizing Symptoms**	b = 0.31, *SE* b = 0.08 [0.16, 0.47]	3.88***	743	286
Relation between magnitude of cross-informant correspondence and cultural tightness: **Reports About Externalizing Symptoms**	b = 0.28, *SE* b = 0.09 [0.11, 0.45]	3.22***	623	258

* $p < .05$, ** $p < .01$, *** $p < .001$.

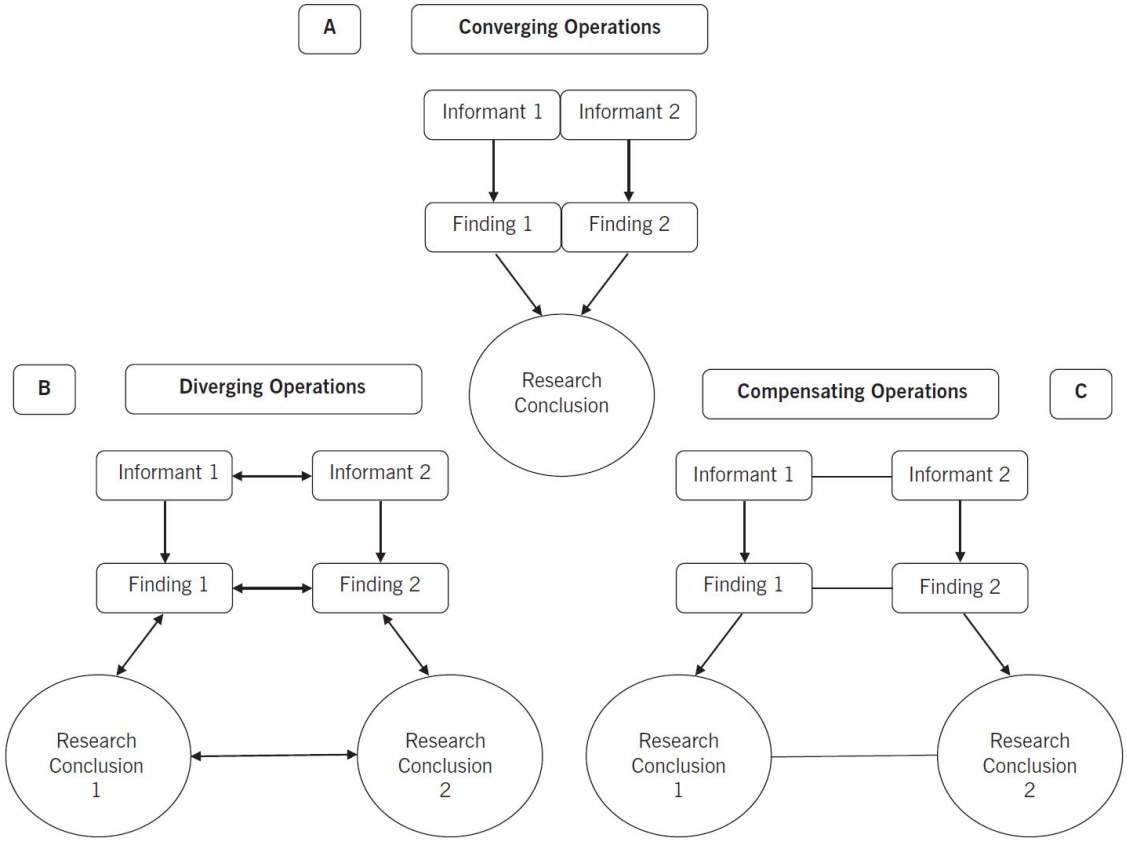

Figure 1. Graphical representation of the research concepts that comprise the Operations Triad Model. The top half (A) represents Converging Operations: a set of measurement conditions for interpreting patterns of findings based on the consistency within which findings yield similar conclusions. The 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 (B) 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 (C) 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. Originally published in De Los Reyes, Thomas, et al. (2013).

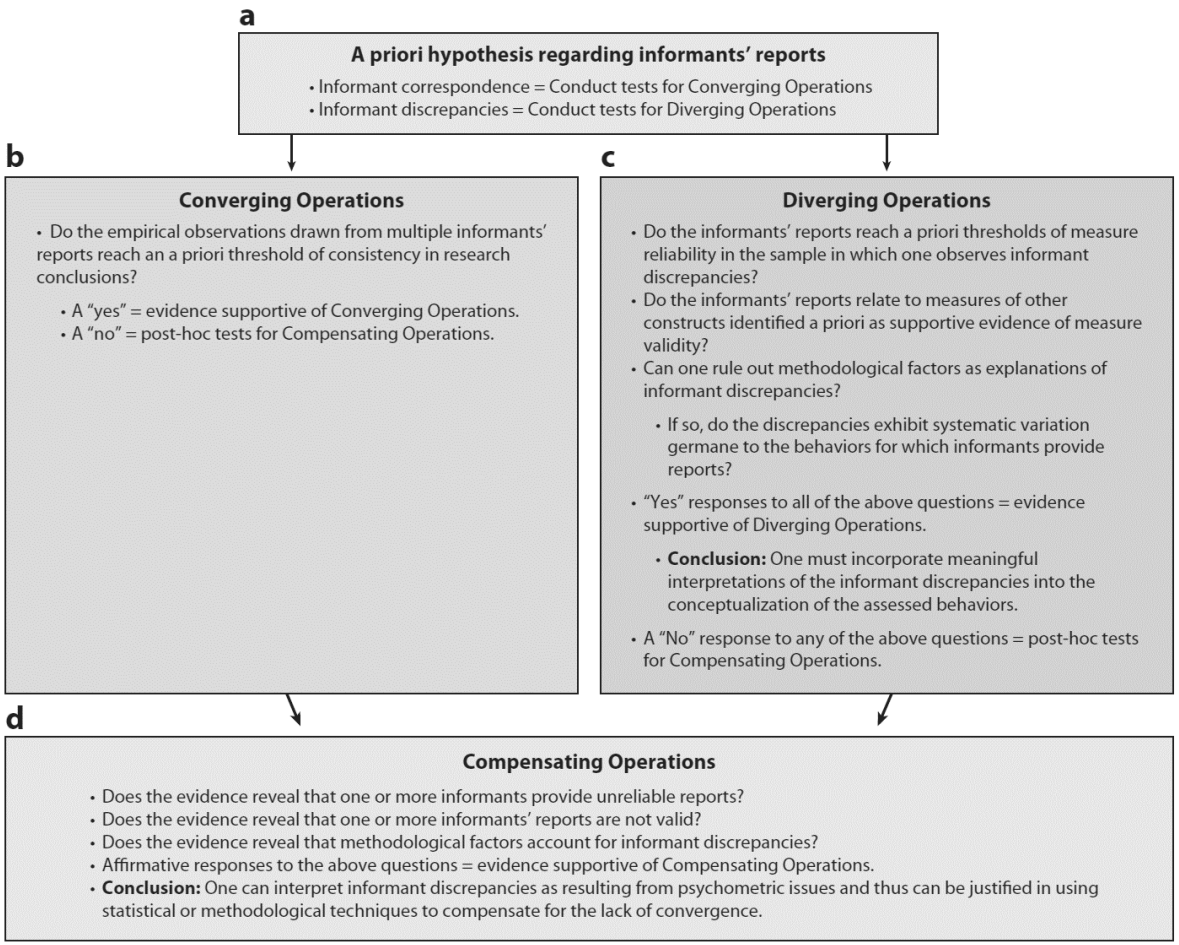

Figure 2. Graphical display of decision-making processes based on the Operations Triad Model. To use this process, researchers must pose *a priori* hypotheses as to whether they expect converging or diverging findings (*a*). Empirical questions outlined in the figure can then guide researchers' tests 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; *b*) or diverging findings (i.e., Diverging Operations; *c*) as yielding meaningful information about behavior. 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; *d*). Originally published in De Los Reyes, Thomas, et al. (2013).

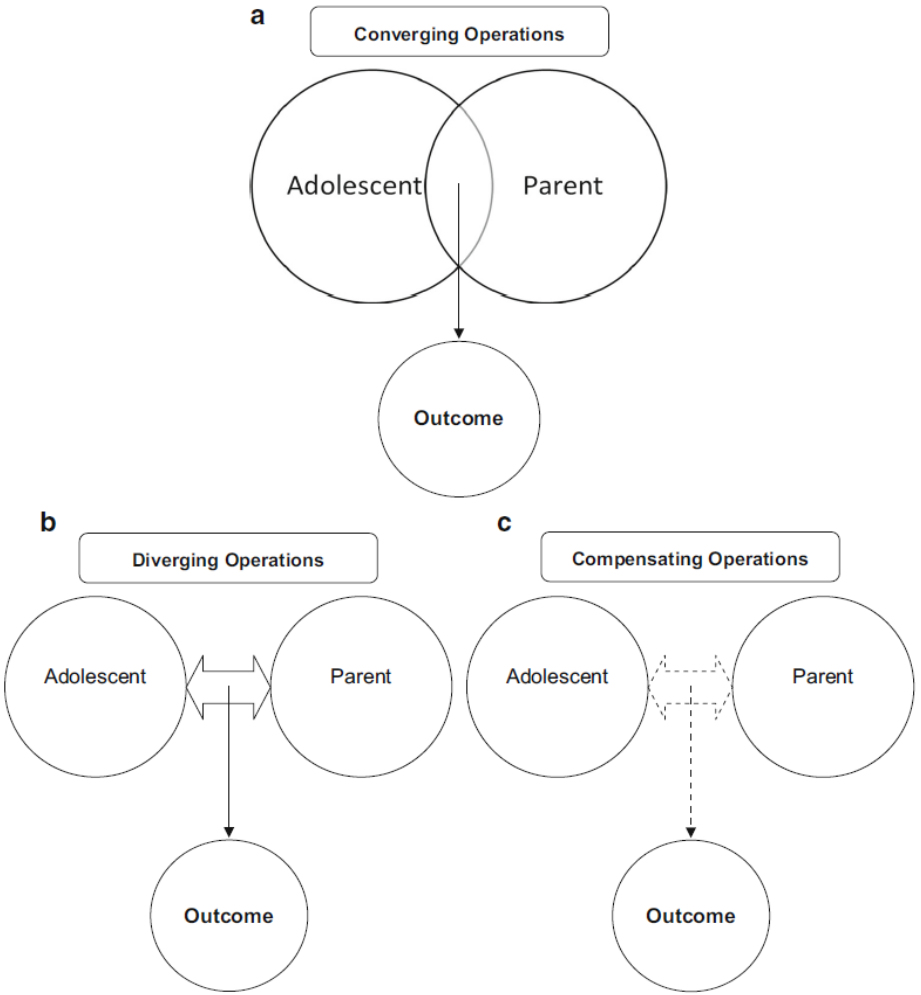

Figure 3. Graphical depiction of adaptations to the Operations Triad Model for use in interpreting adolescent-parent assessments of family functioning and their links to criterion variables reflecting adolescent adjustment. Consistent with Figure 1, we graphically depict interpretations of adolescents' and parents' reports and their links to adolescent adjustment consistent with Converging Operations (a), Diverging Operations (b), and Compensating Operations (c). Originally published in De Los Reyes and Ohannessian (2016).