



HHS Public Access

Author manuscript

Psychol Bull. Author manuscript; available in PMC 2016 July 01.

Published in final edited form as:

Psychol Bull. 2015 July ; 141(4): 858–900. doi:10.1037/a0038498.

The Validity of the Multi-Informant Approach to Assessing Child and Adolescent Mental Health

Andres De Los Reyes^{1,*}, Tara M. Augenstein¹, Mo Wang², Sarah A. Thomas¹, Deborah A.G. Drabick³, Darcy E. Burgers³, and Jill Rabinowitz³

¹Comprehensive Assessment and Intervention Program, Department of Psychology, University of Maryland at College Park, 20742, USA

²Department of Management, Warrington College of Business Administration, University of Florida, Gainesville, FL, 32611, USA

³Department of Psychology, Temple University, Philadelphia, PA, 19122, USA

Abstract

Child and adolescent patients may display mental health concerns within some contexts and not others (e.g., home vs. school). Thus, understanding the specific contexts in which patients display concerns may assist mental health professionals in tailoring treatments to patients' needs. Consequently, clinical assessments often include reports from multiple informants who vary in the contexts in which they observe patients' behavior (e.g., patients, parents, teachers). Previous meta-analyses indicate that informants' reports correlate at low-to-moderate magnitudes. However, is it valid to interpret low correspondence among reports as indicating that patients display concerns in some contexts and not others? We meta-analyzed 341 studies published between 1989 and 2014 that reported cross-informant correspondence estimates, and observed low-to-moderate correspondence (mean internalizing: $r = .25$; mean externalizing: $r = .30$; mean overall: $r = .28$). Informant pair, mental health domain, and measurement method moderated magnitudes of correspondence. These robust findings have informed the development of concepts for interpreting multi-informant assessments, allowing researchers to draw specific predictions about the incremental and construct validity of these assessments. In turn, we critically evaluated research on the incremental and construct validity of the multi-informant approach to clinical child and adolescent assessment. In so doing, we identify crucial gaps in knowledge for future research, and provide recommendations for “best practices” in using and interpreting multi-informant assessments in clinical work and research. This paper has important implications for developing personalized approaches to clinical assessment, with the goal of informing techniques for tailoring treatments to target the specific contexts where patients display concerns.

*Address all correspondence 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; adlr@umd.edu.

Andres De Los Reyes, Tara M. Augenstein, and Sarah A. Thomas, Comprehensive Assessment and Intervention Program, Department of Psychology, University of Maryland. Mo Wang, Department of Management, University of Florida. Deborah A.G. Drabick, Darcy E. Burgers, and Jill Rabinowitz, Department of Psychology, Temple University.

Keywords

Construct Validity; Incremental Validity; Informant Discrepancies; Multiple Informants; Operations Triad Model

Child and adolescent mental health patients lead complex lives (i.e., collectively referred to as “children” unless otherwise specified). Indeed, mental health concerns arise out of an interplay among biological, psychological, and socio-cultural factors that pose risk for, or offer protection against, developing maladaptive reactions to environmental or social contexts (e.g., Cicchetti, 1984; Luthar, Cicchetti, & Becker, 2000; Sanislow et al., 2010). However, not all contexts elicit mental health concerns to the same degree (e.g., Carey, 1998; Kazdin & Kagan, 1994; Mischel & Shoda, 1995). Therefore, patients may display concerns within some contexts, such as home and school, but not others, such as within peer interactions. In fact, these contextual variations in displays of mental health concerns occur within a variety of mental health domains including social anxiety, attention and hyperactivity, and conduct problems (e.g., Bögels et al., 2010; Dirks, De Los Reyes, Briggs-Gowan, Cella, & Wakschlag, 2012; Drabick, Gadow, & Loney, 2007, 2008; Kraemer et al., 2003). Thus, identifying the specific contexts in which patients display concerns may facilitate treatment planning and boost treatment efficacy (e.g., De Los Reyes, 2013; National Institute of Mental Health [NIMH], 2008).

The most prevalent strategy for assessing contextual variations in mental health is the multi-informant assessment approach (Kraemer et al., 2003). Specifically, this approach involves taking reports from informants who share close relationships with the patients about whom they are providing reports, or at minimum spend a significant amount of time observing patients' behavior (Achenbach, 2006). By considering reports from informants who vary among each other in the specific contexts in which they observe patients' behavior (e.g., home vs. school vs. peer interactions), mental health professionals may gain an understanding as to how consistently or inconsistently patients display concerns across contexts (Dirks et al., 2012). For child patients, these informants most often include parents, teachers, and patients themselves (Hunsley & Mash, 2007). Further, trained raters might also complete reports, such as clinical interviewers and independent observers of patients' behavior on standardized clinical tasks (e.g., structured social interactions) or unstandardized home or school observations (Groth-Marnat, 2009).

Clinicians may use informants' reports to make decisions about mental health care, such as assigning diagnoses and planning treatment (e.g., Hawley & Weisz, 2003). Researchers may also use these reports to draw conclusions from empirical work on such topics as identifying treatments that successfully ameliorate mental health concerns (i.e., evidence-based treatments; Weisz, Jensen Doss, & Hawley, 2005). In both practice and research settings, collecting multiple informants' reports generates a great deal of information about patients' concerns. However, the individual reports often yield inconsistent conclusions (i.e., *informant discrepancies*; Achenbach, 2006; De Los Reyes & Kazdin, 2004, 2005, 2006; Goodman, De Los Reyes, & Bradshaw, 2010).¹ For instance, a female adolescent patient in

a pre-treatment assessment may be identified as experiencing “low” positive mood based on a parent or teacher report whereas the adolescent self-reports her mood as “elevated.”

Informant discrepancies often create considerable uncertainties in delivering services to patients and drawing conclusions from research (De Los Reyes, Kundery, & Wang, 2011). A key reason for these uncertainties originates from the near-exclusive focus in mental health research on whether informant discrepancies reflect measurement error or reporting biases (e.g., De Los Reyes, 2011; Richters, 1992). Consequently, what remains unclear is whether multi-informant approaches to assessment validly capture contextual variations in displays of patients' mental health concerns, and which informants ought to be included in mental health assessments. As we explain below, these decisions might involve considering data from multiple assessment literatures, chief among them research on multi-informant clinical child assessments and their (a) correspondence, (b) incremental validity, and (c) construct validity. Prior work has addressed each of these issues as they relate to clinical child assessment broadly (e.g., Achenbach, McConaughy, & Howell, 1987; Johnston & Murray, 2003; Mash & Hunsley, 2005), and within assessments in other clinical literatures (e.g., Achenbach, 2006; Campbell & Fiske, 1959; Dawes, 1999; Garb, 2003; Hunsley & Mash, 2005). However, with regard to multi-informant clinical child assessments, these literatures have advanced largely in isolation of one another.

The Present Review

The purpose of this paper is to review and synthesize research on using and interpreting multiple informants' reports in clinical child assessments. We expand the literature on this topic in five ways. First, as an update to the seminal meta-analysis on cross-informant correspondence of clinical child assessments carried out by Achenbach and colleagues (1987), we conduct a quantitative review of 341 studies of cross-informant correspondence in reports of children's mental health published over the last 25 years (1989-2014). Second, we use this updated quantitative review as a backdrop for discussing recent work on the conceptual rationale for conducting multi-informant assessments (De Los Reyes, Thomas, Goodman, & Kundery, 2013). We expand upon this conceptual work by developing specific predictions about the incremental and construct validity of the multi-informant approach. Third, we review research on the *incremental validity* of multi-informant assessments, or work examining whether a multi-informant approach yields reports that, relative to one another, incrementally contribute information in the prediction of relevant criterion variables (e.g., diagnostic status or treatment response; see Dawes, 1999; Garb, 2003; Hunsley & Meyer, 2003). Fourth, we summarize recent work on the construct validity of multi-informant assessments. By *construct validity*, we mean research examining whether it is valid to interpret patterns of convergence and divergence among multiple informants' reports as accurate reflections of contextual variations in children's mental health concerns (e.g.,

¹We use the term *informant discrepancies* to generally denote studies of differences and similarities among informants' reports. Yet, researchers have used other terms to characterize this research (e.g., *informant agreement* and *informant correspondence*). Importantly, these terms are not interchangeable and, where possible, we use different terms to denote disparate lines of research. That is, these terms distinguish examinations of divergence between reports (discrepancies) from examinations of convergence between reports (agreement and correspondence), and researchers often use disparate statistical techniques within and between studies (e.g., statistical interactions between reports to study divergence vs. correlations between reports to study convergence; De Los Reyes, Salas, Menzer, & Daruwala, 2013).

Borsboom, Mellenbergh, & van Heerden, 2004; Campbell & Fiske, 1959). Lastly, we synthesize work on multi-informant assessments to provide recommendations for both future research and “best practices” for using and interpreting multi-informant assessments administered in clinic settings.

Correspondence in Multi-Informant Clinical Child Assessments

One of the most normative observations in assessments of child patients is that of low correspondence between informants' reports about patients' mental health (Achenbach, 2011). Over five decades ago, Lapouse and Monk (1958) were the first to report about this phenomenon. Their primary aim was to estimate the prevalence of children's mental health concerns in an epidemiological study of 482 6- to 12-year-old children representative of the community from which they were sampled (i.e., Buffalo, New York, USA). To estimate prevalence, the researchers relied on structured clinical interviews administered to mothers about a single child from the household. In this study, the researchers also wanted to address what they termed “the problem of validity” with respect to mothers' interview responses. Thus, the researchers conducted a secondary study using a separate convenience sample of 193 children ages 8-12 and their mothers, recruited from local outpatient hospitals and pediatricians' offices. In this sample, researchers conducted structured clinical interviews with the mothers and the identified child separately and simultaneously, using two different interviewers. To assess correspondence between mother and child reports, researchers examined responses on a subset of behavioral domains assessed in the interview, including fears, nightmares, bed wetting, restlessness, and repetitive behaviors (e.g., tics, thumb sucking, skin picking). Using percent agreement to assess correspondence, the authors reported agreement between 46% (amount of food intake) and 84% (bed wetting). Broadly, the authors observed greater correspondence between reports of relatively more observable behaviors (e.g., bed wetting, stuttering, thumb sucking) than relatively less observable behaviors (e.g., fears and worries, nightmares, restlessness). Further, with two exceptions (bed wetting and overactivity), lack of correspondence was primarily due to children self-endorsing behaviors that the mother did not endorse about the child.

Highlighting the robust nature of cross-informant correspondence patterns, in their seminal meta-analysis of 119 studies of cross-informant correspondence in reports of children's mental health, Achenbach and colleagues (1987) reported, similar to Lapouse and Monk (1958), that correspondence levels exhibited a range from low-to-moderate in magnitude (i.e., r s ranging from .20 to .60). Importantly, Achenbach and colleagues (1987) noted that the low correlations they observed among informants' reports of children's mental health did not necessarily indicate that such reports carried poor psychometric qualities, as they also observed satisfactory test-retest reliability estimates for the informants' reports they examined. In fact, levels of correspondence systematically varied as a function of three key factors. First, pairs of informants who observed children in the same context (e.g., pairs of parents or pairs of teachers) tended to exhibit greater levels of correspondence than pairs of informants who observed children in different contexts (e.g., parent and teacher). Similarly, a second key finding was that greater levels of cross-informant correspondence occurred when informants provided reports about younger children relative to reports about older children. Younger children may be relatively more constrained than older children in the

contexts in which they display mental health concerns (e.g., De Los Reyes & Kazdin, 2005; Smetana, 2008). Additionally, Achenbach and colleagues (1987) conducted their meta-analysis at a time when cross-informant correspondence studies predominantly focused on comparing adults' reports of children's concerns (e.g., parent vs. teacher; mother vs. father; see Table 2 of Achenbach et al., 1987). Thus, this age effect might have also reflected the idea that less variation among informants in contexts of observation pointed to greater correspondence between informants' reports, particularly for adult informants' reports of children's concerns. Third, Achenbach and colleagues (1987) observed larger correspondence levels between informants' reports of children's externalizing (e.g., aggression and hyperactivity concerns) versus internalizing (e.g., anxiety and mood) concerns. This third finding may have reflected greater correspondence between reports of directly observable concerns, relative to concerns that are internally experienced by the child and thus relatively less observable in nature.

Meta-Analysis of the Last Quarter-Century of Cross-Informant Correspondence Studies

Since the seminal work of Lapouse and Monk (1958) and Achenbach and colleagues (1987), researchers have conducted many additional cross-informant correspondence studies. As in earlier work, more recent studies find that low-to-moderate levels of cross-informant correspondence characterize such varied assessment settings and mental health domains as inpatients' depressive mood symptoms (Frank, Van Egeren, Fortier, & Chase, 2000), outpatients' anxiety symptoms (Rapee, Barrett, Dadds, & Evans, 1994), and disruptive behavior assessments taken from representative non-clinic samples (Offord et al., 1996). Further, the Lapouse and Monk (1958) finding of greater endorsements among children's self-reports relative to mothers' reports was corroborated by comparisons between mother and child behavioral checklist reports from 25 countries (Rescorla et al., 2013). Since the Achenbach et al. (1987) review, researchers have expanded the range of domains examined for cross-informant correspondence (e.g., anxiety, conduct problems, hyperactivity, mood; De Los Reyes, 2011). Further, studies conducted after the Achenbach et al. (1987) review collectively examined to a greater extent variations in cross-informant correspondence across multiple (a) measurement and scaling methods (e.g., behavioral checklists, clinical interviews, symptom rating scales), (b) patients' developmental periods, and (c) informant pairs (e.g., greater focus on children's self-reports for assessing internalizing concerns) (e.g., Achenbach, 2006; De Los Reyes & Kazdin, 2005).

In light of the continued attention to estimating cross-informant correspondence, have the findings of Achenbach and colleagues (1987) stood the test of time? To address this question, we conducted a meta-analysis of the last 25 years of cross-informant correspondence studies for assessments of children's internalizing and externalizing mental health concerns. We focused on these domains because they encompass the most commonly assessed and treated childhood concerns (i.e., anxiety, attention and hyperactivity, conduct, mood; Hunsley & Mash, 2007; Weisz et al., 2005). Further, we took two approaches to identifying studies for our review. First, we sampled studies included in meta-analyses of cross-informant correspondence published since Achenbach et al. (1987), which is a variant of the *second-order* meta-analytic approach (e.g., Butler, Chapman, Forman, & Beck, 2006; Cuijpers & Dekker, 2005; Lipsey & Wilson, 1993; Møller & Jennions, 2002; Peterson,

2001; Tamim, Bernard, Borokhovski, Abrami, & Schmid, 2011). Second, we searched for studies published in the years (i.e., 2000-2014) following the meta-analyses sampled in our review. Importantly, Achenbach and colleagues (1987) reviewed 119 studies published over roughly a quarter-century (i.e., 1960-1986). Similarly, as we explain below, we reviewed 341 studies published in the most recent quarter-century (i.e., 1989-2014). Thus, we were well-positioned to assess whether more recent work replicated the findings of Achenbach and colleagues (1987). To this end, we focused our review on studies that examined correspondence among parent, teacher, and child reports of children's mental health. We focused on these three informants because these are the reporters on which mental health professionals most commonly rely when administering and interpreting the outcomes of clinical child assessments (e.g., Hunsley & Mash, 2007; Kraemer et al., 2003).

Method

Literature review

We identified meta-analyses and empirical studies published since Achenbach and colleagues (1987). We conducted two searches. First, to identify relevant meta-analyses we searched via Google Scholar of all peer-reviewed scholarly work citing the Achenbach et al. (1987) review (N = 3,978 citations; search conducted March 2, 2014). We searched within these cited articles using the search terms “informant” and “quantitative review.” We augmented this search with an additional Google Scholar search of cited articles using the terms “meta-analysis OR quantitative review OR systematic review,” and conducted this same literature search using the Web of Science search engine. Combined across searches, we identified 1,799 articles. This search yielded an initial set of four quantitative reviews (i.e., Crick et al., 1998; Duhig, Renk, Epstein, & Phares, 2000; Meyer et al., 2001; Renk & Phares, 2004), to which we applied the inclusion and exclusion criteria described below. As an additional check, we conducted Google Scholar searches using the above-listed terms of all work citing these four quantitative reviews, yielding an additional set of 858 articles. This search yielded no additional quantitative reviews.

Second, to identify empirical articles of cross-informant correspondence conducted in the years (i.e., between 2000 and 2014) following the meta-analyses included in our review, we searched via Google Scholar for all peer-reviewed scholarly work citing the Achenbach et al. (1987) review between 2000 and 2014 using the search terms “internalizing symptoms/problems/difficulties OR externalizing symptoms/problems/difficulties” (N = 1,440 citations; search conducted September 9, 2014). Additionally, we searched the reference lists of narrative reviews of cross-informant correspondence research published since 2000.

Inclusion and exclusion criteria

To be included in our quantitative review, meta-analyses and studies must have (a) focused on informants' reports of children at or under the age of 18 years; (b) examined correspondence between informants' reports of children's mental health concerns (i.e., internalizing and/or externalizing concerns); (c) examined correspondence between reports completed by pairs of parents, teachers, and/or children (i.e., mother-father, parent-child, parent-teacher, teacher-child); and (d) been published in English. In addition to these

criteria, included meta-analyses must have provided a list of individual studies used to calculate metrics of cross-informant correspondence. We employed this criterion to ensure that articles examined in our review were published after 1987 (i.e., not included in the original review by Achenbach et al., 1987). Further, knowledge of individual studies within each of the reviews allowed us to identify any articles examined by more than one review and multiple articles that examined the same sample or cohort of children. This step allowed us to ensure that effects observed did not occur because reviews examined the same studies or sample(s) across studies.

Among individual empirical studies identified either via the reference lists of meta-analyses or our additional search for individual studies published between 2000 and 2014, to be included in our review studies must have provided sufficient data to code estimates of cross-informant correspondence. Specifically, we required studies to report between-informant correspondence metrics (e.g., Pearson correlations for dimensional measures or Kappa coefficients for categorical measures) on measures completed by one or more informant pairs (e.g., mother-father, parent-child, parent-teacher, teacher-child). Measures completed by these informants must have assessed the same construct at the same time point (e.g., measures about internalizing problems completed by parent and child when the child was 10 years old). We excluded studies that only reported correspondence metrics as an average or range across multiple informant pairs (e.g., *correlations ranged from .15-.40 depending on informant pair*), as this information did not allow us to code moderator variables. Similarly, we excluded studies that only reported correspondence metrics as an average or range across informants' reports of different mental health domains. Finally, studies needed to report cross-informant correspondence estimates for measures of mental health on the internalizing (e.g., anxiety and mood) and/or externalizing (e.g., aggression and hyperactivity) spectrum. We excluded studies focused on related constructs, namely studies on risk and protective factors of mental health (e.g., emotion regulation, parenting practices, personality traits, resiliency, self-esteem).

For meta-analyses identified in our search, employing inclusion criteria led to our excluding the Crick et al. (1998) and Renk and Phares (2004) reviews. We excluded Crick et al. (1998) because it did not list the individual studies the authors used to calculate cross-informant correspondence estimates. We excluded the Renk and Phares (2004) review because it focused on correspondence between reports on a construct (i.e., social competence) that fell outside of the spectra of internalizing and externalizing mental health concerns. Thus, we identified relevant studies via quantitative review based on two meta-analyses that met our inclusion criteria: Duhig et al. (2000) and Meyer et al. (2001). Across studies identified via these meta-analyses and our search of individual studies published between 2000 and 2014, collectively we coded effect sizes on a final sample of 1218 data points taken from 341 studies published between 1989 and 2014. A complete study list can be retrieved online at <https://sites.google.com/site/caipumaryland/Home/people/director>.

Data extraction and covariates coded from each study

Two doctoral graduate students served as coders for our quantitative review, and received coding training by the first-author and fifth-author through discussion and practice coding.

After the two coders completed study coding, the fifth-author served as an independent assessor on 50% of the studies coded. This independent assessment resulted in 100% inter-rater agreement on effect sizes coded (i.e., effect size magnitude and metrics), and Cohen's Kappa coefficients of 1.0 in terms of inter-rater agreement on coding of each of the covariates described below (i.e., child age, informant pair, measurement method, mental health domain).

Outcomes representing mean estimates of cross-informant correspondence were presented as Pearson r correlations, Cohen's Kappa coefficients, Cohen's d , Hedge's g , means and standard deviations of informants' reports, or odds ratios that we converted to r for the meta-analyses. In addition to mean estimates of cross-informant correspondence, we also coded for four covariates, consistent with effects observed by Achenbach and colleagues (1987): (1) mental health domain (internalizing vs. externalizing); (2) informant pair (mother-father, parent-child, parent-teacher, teacher-child); (3) child age (younger [10 years and younger] vs. older [11 years and older]); and (4) measurement method (categorical vs. dimensional). We coded studies based on measurement method because recent work indicates that, relative to categorical scaling, dimensional scaling results in measures that evidence greater estimates of reliability and validity (Markon, Chmielewski, & Miller, 2011). Thus, we would expect greater levels of cross-informant correspondence for reports taken using dimensional scales (e.g., behavioral checklists), relative to categorical scales (e.g., diagnostic interview endorsements).

Data-analytic strategy

As described below, we observed significant differences between cross-informant correspondence estimates for reports of internalizing concerns versus externalizing concerns (Table 1). Thus, we performed two primary meta-analyses, one for internalizing concerns and one for externalizing concerns, using 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. Because the studies included in the meta-analyses varied in methodology and design, we calculated a random-effects model. In addition, for some studies and samples, we observed multiple effect sizes that varied across our moderator variables (e.g., multiple effect sizes reported for different informant pairs, categorical vs. dimensional approaches). We accounted for this nesting in the data by calculating (a) effect sizes for each cohort or sample and then (b) an overall effect size. Specifically, we computed effect sizes and variances for each cohort. Next, we calculated 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 informant pairs).

Additionally, we addressed the issue of statistical stability in two ways. First, we individually excluded each study from the analysis and recalculated the pooled r and 95% confidence interval (CI). If an individual study contributed heavily to the pooled r , we would observe a change in the magnitude or significance of the pooled r . Second, 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 as an r of .10).

Following determination of the combined effect size across studies, we considered heterogeneity or observed variance across studies. Specifically, we calculated τ , a metric that is sensitive to the unit of measurement and acts as the standard deviation of the summary effect, and I^2 , which provides an index of the proportion of observed variability that is attributable to heterogeneity among the data points and reflects “real” differences among studies (Borenstein et al., 2009; Higgins, Thompson, Deeks, & Altman, 2003). To explore the impact of the categorical variables or covariates on this observed variability among studies, we conducted ancillary, subgroup analyses to determine effect sizes (r s) for each level of the categorical variable (e.g., separately for studies that used categorical vs. dimensional assessment approaches or examined younger vs. older children). For these analyses, we calculated r s and associated p -values for each level of the categorical variable (Borenstein et al., 2009). We then examined the Q statistic, which is based on the weighted sum of squares for each level of the covariate and thus provides an index of dispersion (Borenstein et al., 2009), to determine whether the magnitude of correspondence differed between levels of the covariates.

Results and Discussion

Overall effect sizes

Combined across 341 studies and 1218 data points, we observed an overall cross-informant correlation of .28 (95% CI [.22, .33]; $p < .001$). These estimates are virtually identical to the mean cross-informant correlations reported by Achenbach and colleagues (1987). In Table 1, we report cross-informant correlations for reports of children's internalizing and externalizing concerns. Consistent with Achenbach and colleagues (1987), we observed low-to-moderate and statistically significant magnitudes of cross-informant correspondence for both domains. As seen in Table 1, we observed non-overlapping 95% CI for cross-informant correspondence estimates for reports of children's internalizing concerns versus externalizing concerns. Overall, informant correspondence rates tended to be larger for reports of children's externalizing concerns, relative to reports of children's internalizing concerns. Thus, below we report effects of covariates on levels of cross-informant correspondence separately for reports of children's internalizing concerns and externalizing concerns.

Heterogeneity in effect sizes and stability of results

Consistent with Achenbach et al. (1987), we observed an overall large variance in results for internalizing and externalizing concerns, respectively ($I^2 = 99.25$ and 99.40), suggesting

heterogeneity among studies in effect sizes. That is, the percentage of total variability that is attributable to heterogeneity among the data points included in the meta-analysis is approximately 99% for reports of internalizing and externalizing concerns. Further, removal of any individual study from the analysis did not affect relations between magnitudes of cross-informant correspondence and covariates (r s with each individual study removed ranged from .27 to .28 for reports of internalizing concerns, r with each individual study removed was .36 for reports of externalizing concerns; all p s < .001). Using the respective effect sizes for these data points, defining a trivial effect size as an r of .10, and a threshold of p < .05, Orwin's Fail-safe N was 1,875 for reports of internalizing concerns and 2,292 for reports of externalizing concerns. This indicates that one would have to include in the meta-analysis over 1,800 data points of reports of children's internalizing concerns and well over 2,000 data points of reports of children's externalizing concerns with a mean r of .00 before the cumulative effects would become trivial.

Effects of covariates

To explore whether the covariates could account for some of the variability among the effect sizes for each of the data points (697 for internalizing, 521 for externalizing), we conducted ancillary analyses where we calculated separate r and p -values for each level of the categorical covariates. Because the data points within studies and cohorts often differed in these covariates (e.g., child age, measurement method), these supplementary analyses considered the data points independently. Beyond the effects of mental health domain reported previously, we observed significant effects for two other covariates. First, pairs of mother and father informants (i.e., informants who observe child in the same setting) yielded larger magnitudes of cross-informant correspondence, relative to all other informant pairs (i.e., parent-child, parent-teacher, teacher-child). We observed this effect for both reports of internalizing concerns (r s: .48 vs. .24; $Q = 72.42$, $p < .001$) and externalizing concerns (r s: .58 vs. .30; $Q = 96.26$; $p < .001$). Second, informants completing reports on a dimensional scale tended to yield higher magnitudes of correspondence, relative to informants completing reports on a categorical scale. We observed this effect of measurement method for both reports of internalizing concerns (r s: .29 vs. .06; $Q = 38.54$, $p < .001$) and externalizing concerns (r s: .37 vs. .06; $Q = 30.86$, $p < .001$). Further, we observed non-significant effects of child age on levels of correspondence for reports about younger children (i.e., 10 years and younger) versus older children (i.e., 11 years and older) for both reports of internalizing concerns (r s: .32 vs. .26; $Q = 1.42$, $p = .23$) and externalizing concerns (r s: .38 vs. .35; $Q = 0.48$, $p = .49$).

In sum, we made three findings consistent with previous reviews (Achenbach et al., 1987; Markon et al., 2011). First, informants' reports of relatively more observable or externalizing concerns tended to evidence greater levels of cross-informant correspondence, relative to reports of internalizing concerns. Second, informant pairs for which both informants observed the child in the same setting (i.e., mother-father) tended to yield the highest levels of correspondence, relative to all other informant pairs. Third, when informants completed reports using measures that tended to evidence greater reliability and validity estimates (i.e., dimensional), we observed greater levels of cross-informant correspondence, relative to when informants completed reports using categorical measures.

One finding from our review that was inconsistent with the Achenbach et al. (1987) meta-analysis was our null finding regarding the effects of child age on magnitudes of cross-informant correspondence. These inconsistencies may reflect changes since Achenbach et al. (1987) with regard to evidence-based assessment practices for clinical child assessments. Specifically, the Achenbach et al. (1987) review consisted of a relatively small proportion of studies comparing parent and teacher reports to children's self-reports (see Table 2 of Achenbach et al., 1987). Since 1987, a proliferation of evidence-based assessment research focused on understanding children's perspectives on their own mental health concerns, and this increased focus resulted in an increased number of options available for taking self-reports of mental health, particularly for older children and adolescents (for reviews, see Klein, Dougherty, & Olino, 2005; McMahon & Frick, 2005; Silverman & Ollendick, 2005). Consequently, we based our cross-informant correspondence estimates for reports of children's mental health concerns on a sample of studies in which over 50% compared children's self-reports to reports from other informants (Table 1). For some mental health domains and contexts (e.g., worry and anxiety displayed within peer interactions; covert delinquent behaviors displayed within peer interactions), children may be in a unique position to observe displays of these concerns, relative to parents and teachers (e.g., Comer & Kendall, 2004; McMahon & Frick, 2005). Thus, it may be that the null effects of child age on magnitudes of cross-informant correspondence that we observed in fact reflected a greater use since Achenbach et al. (1987) of self-reports to assess children's mental health concerns.

Conceptual Foundations of Multi-Informant Mental Health Assessments

Operations Triad Model

Collectively, our meta-analytic findings indicated that clinical child assessments tend to yield low-to-moderate levels of cross-informant correspondence, with higher levels of correspondence occurring when informants complete reports about mental health concerns that both informants either have relatively greater opportunities to observe (e.g., externalizing vs. internalizing) or observe within the same context (e.g., mother-father vs. parent-teacher). The main findings from the Achenbach and colleagues (1987) review, as well as the stability of these findings in research published since their review (Table 1), have greatly informed the development of theoretical principles about using and interpreting multi-informant assessment outcomes. Indeed, an underlying assumption of the multi-informant approach is that informants each carry a unique and valid perspective of the patients about whom they provide reports (De Los Reyes, 2013). Thus, to maximize the clinical value offered by the multi-informant approach, the informants selected to provide reports ought to differ in their opportunities for observing patients' mental health concerns (e.g., observations that vary across home and school contexts; Kraemer et al., 2003). Therefore, if patients contextually vary in where they display concerns, then discrepancies among informants' reports should, in part, reflect these contextual variations.

These basic assumptions underlying the multi-informant assessment approach have been elaborated upon and codified in recent theoretical work. Specifically, the idea that integrating multiple distinct pieces of information (e.g., research evidence and clinical

expertise) improves clinical decision-making relative to relying on limited information (e.g., only clinical expertise) is a foundational principle of evidence-based practice (e.g., Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). Yet, much of multi-informant assessment research has focused on whether informant discrepancies reflect error or rater biases (for a review, see De Los Reyes, 2013), and this research often conflicts with the key rationale for taking a multi-informant approach (De Los Reyes, Thomas et al., 2013). To address these inconsistencies between theory and interpretations of multi-informant assessments, researchers developed the Operations Triad Model (OTM; Figure 1; De Los Reyes, Thomas et al., 2013). The OTM addresses a key issue underlying these inconsistencies, namely, that research methodology in clinical psychology has largely relied on the concept of Converging Operations to interpret research conclusions and assessment outcomes in practice. *Converging Operations* is a set of measurement conditions within which one draws inferences of the veracity of multiple methodologically distinct research observations based on whether the observations yielded similar conclusions (Garner, Hake, & Eriksen, 1956). Under this concept, one draws stronger inferences from studies within which observations converged on a common conclusion (e.g., multiple informants' reports all supported treatment efficacy), and in turn, one draws weaker inferences from studies within which observations diverged toward different conclusions (De Los Reyes, Thomas et al., 2013).

To expand upon the Converging Operations concept, the OTM delineates conditions for two additional research concepts for interpreting discrepancies among informants' reports. First, the OTM includes a concept, Diverging Operations, developed for interpreting instances in which multiple informants' reports diverge from each other for meaningful reasons. An example of Diverging Operations may involve a circumstance in which a clinician expects two informants' reports about a patient's mental health concerns to meaningfully diverge because (a) the informants observe the patient's behavior within completely different contexts (e.g., home vs. school) and (b) the patient displays mental health concerns to a greater degree in one context than the other (e.g., home to a greater extent than school). Second, the OTM also includes conditions for a research concept, Compensating Operations, reflecting those circumstances in which multiple informants' reports diverge from each other because of measurement error or some other methodological process. For instance, two informants' reports may yield divergent information because they completed reports using distinct measures (e.g., different item content and scaling) and the measures may have varied in their psychometric properties (e.g., internal consistency estimates). Thus, within instances that reflect Compensating Operations, methodological features of the assessment process provide a parsimonious account for why informants' reports yielded different outcomes. In sum, the OTM promotes an evidence-based approach to testing whether meaningful data can be gleaned from interpreting the consistencies and discrepancies observed among multiple informants' reports.

Predictions about the Incremental and Construct Validity of Multi-Informant Assessments

Mental health professionals can make a strong conceptual case for taking a multi-informant approach to assessing child patients. Further, the OTM may facilitate linking the concepts underlying administering multi-informant assessments with interpreting their outcomes (Figure 1). Yet, unknown is how to use the OTM to make specific predictions about the

incremental and construct validity of multi-informant assessments. That is, when multi-informant assessments yield low correspondence among informants' reports, how should the evidence appear if the low correspondence reflects contextual variations in patients' mental health? An illustrative example of multi-informant assessment outcomes may be helpful.² For the purposes of this illustration, we focus on interpreting reports from informants who differ in the contexts in which they observe patients, specifically parents at home versus teachers at school. We also focus on assessments of a mental health domain for which patients' concerns may significantly vary across home and school contexts, namely preschoolers' disruptive behavior (for a review, see Wakschlag, Tolan, & Leventhal, 2010).

To assess a sample of patients experiencing disruptive behavior, consider a research team conducting a multi-informant assessment that involved collecting parent and teacher reports. In Figure 2, we present graphical depictions of the outcomes of these assessments. Figure 2 demonstrates links among parent and teacher reports of disruptive behavior and “actual” disruptive behavior as displayed by the patient within home and/or school contexts. We include this measure of “actual” disruptive behavior in Figure 2 to illustrate the construct validity of parent and teacher reports. Specifically, we graphically depict three different patients, or patients who display concerns in home but not school (Figure 2a), school but not home (Figure 2b), and across home and school (Figure 2c). For each example, arrows link informants' reports to patients' behavior. These arrows depict the strength of relations between informants' reports and patients' behavior. That is, solid-line arrows linking informants' reports to patients' behavior indicate stronger relations between reports and behavior, relative to the dashed-lines that indicate weaker relations between reports and behavior. Thus, Figures 2a and 2b depict examples of Diverging Operations, in that the parent and teacher reports differ because the patient displays concerns in either the home context or school context, but not both contexts. In contrast, Figure 2c depicts an example of Converging Operations, in that the parent and teacher reports corroborate each other because the patient displays concerns across home and school contexts.

In Figure 2, we also describe the strength of relations between parent and teacher reports and an independently administered criterion variable; an example of such a variable might be a consensus diagnostic assessment by an experienced clinical team. We include this criterion variable in Figure 2 to illustrate the incremental validity of parent and teacher reports. Here, we use the same “solid-line/dashed-line” approach as with the links between informants' reports and measures of “actual” patients' behavior that we used to illustrate construct validity. That is, in illustrating relations between parent and teacher reports and the criterion variable, solid-line (relative to dashed-line) arrows linking informants' reports to each other indicate stronger relations between reports. Further, solid-line (relative to dashed-line) arrows linking an informant's report to the validity criterion indicate stronger incremental validity for the “solid-line” informant's report, relative to the “dashed-line” informant's report. Therefore, in Figure 2a parent (and not teacher) reports incrementally predict variance in the criterion. In contrast, in Figure 2b teacher (and not parent) reports

²The predictions we highlight below may also apply to interpreting distinctions between observer reports (e.g., parent, teacher, and clinician reports) and patient self-reports, with the one caveat that the contexts in which patients observe their own behavior may overlap with those of other informants' reports (see De Los Reyes, Bunnell et al., 2013; Kraemer et al., 2003).

incrementally predict variance in the criterion. In Figure 2c, the overlap between parent and teacher reports results in neither report incrementally predicting variance in the criterion.

If low correspondence among reports validly reflected contextual variations in patients' disruptive behavior, the patterns of assessment outcomes would be likely to reflect two predictions, each of which are graphically depicted in Figure 2. The first prediction is that each informant's report should “build upon the other” in terms of predicting or relating to relevant clinical metrics (e.g., consensus diagnoses, referral status, treatment response). The uniqueness of informants' reports cannot merely reflect “noise” and must reveal something important about patients' clinical presentations, prognoses, and/or responses to treatment protocols (see also Youngstrom, 2008). Thus, in Figure 2 one can see how conducting a study on the incremental validity of the multi-informant approach depends heavily on the contexts in which patients displayed concerns. Specifically, a sample of patients whose concerns were specific to the home context may result in teacher reports that did not yield incrementally valid information relative to parent reports (Figure 2a). Conversely, a sample of patients who exhibited school-specific concerns may yield parent reports that did not evidence incremental validity relative to teacher reports (Figure 2b). Thus, for multiple informants' reports to be considered incrementally valid relative to each other, patients in a sample should be heterogeneous in terms of the contexts in which they displayed concerns (i.e., a sample of Figure 2a, 2b, and 2c patients). That is, if a sample of patients exhibited individual differences in the context(s) in which they displayed concerns, then each kind of informant who provided reports about these patients (e.g., parents and teachers) would have had the opportunity to observe at least a subgroup of patients' concerns in the sample. Without an informant having the *opportunity* to observe patients' concerns, one cannot expect that informant to provide incrementally valid reports about such concerns.

The second prediction is that patterns of cross-informant correspondence should reflect contextual variations inherent in patients' concerns. For instance, a pattern of parents reporting high levels of disruptive behavior for which teachers' reports do not corroborate should indicate that patients exhibited disruptive behavior to a greater extent at home than school (Figure 2a). Alternatively, consistently high levels of disruptive behavior reported by both parents and teachers should indicate that patients experienced concerns across contexts (Figure 2c).

Considering Constructs Other than Context When Interpreting Informants' Reports

One key issue warrants comment. Specifically, a great deal of research has focused on testing whether low versus high cross-informant correspondence reflects constructs other than contextual variations in displays of patients' concerns (for reviews, see Achenbach, 2006; De Los Reyes & Kazdin, 2005; Kraemer et al., 2003). Thus, drawing inferences about patients' concerns from multi-informant assessments ought to involve examining plausible rival hypotheses for why informants' reports may converge or diverge.

Clinical severity—First, one interpretation of the patterns of mental health concerns illustrated in Figure 2 is that consistencies and inconsistencies between informants' reports reflect individual differences in severity of patients' clinical presentations, and not

necessarily contextual variations in patients' concerns. Importantly, the literature is quite equivocal as to whether cross-contextual consistency in patients' mental health concerns can be treated as a proxy for high levels of clinical severity. Indeed, prior work identified relations between informant discrepancies and context-specific displays of mental health concerns, even when accounting for participants' clinical severity (e.g., De Los Reyes, Bunnell, & Beidel, 2013; De Los Reyes, Henry, Tolan, & Wakschlag, 2009). Additionally, studies of patients from a variety of clinical domains (e.g., attention and hyperactivity, conduct problems, social anxiety) are quite inconsistent as to whether patients evidencing concerns across contexts also evidence significantly greater clinical severity or impairment levels than patients who display concerns within specific contexts (e.g., Bögels et al., 2010; Dirks et al., 2012; Drabick et al., 2007). In contrast, studies of informant discrepancies in assessments of large samples of psychiatric inpatients (Carlson & Youngstrom, 2003) and outpatients (Thuppal, Carlson, Sprafkin, & Gadow, 2002) found that patients whose concerns were endorsed by at least two informants evidenced greater levels of clinical impairment, relative to patients whose concerns were endorsed by a single informant. Overall, in informant discrepancies studies, one ought to statistically account for the clinical severity of patients' concerns when examining the relations between informants' reports and contextual variations in patients' concerns.

Informants' perspectives and rater biases—Perhaps the most frequently studied set of constructs in informant discrepancies research are those that reflect variations in informants' perspectives about patients' concerns and possible biases in informants' reports. Research on these informant characteristics has been extensively reviewed in prior work (e.g., De Los Reyes, 2013; De Los Reyes & Kazdin, 2005). Consequently, we will focus on the evidence regarding the most often-studied characteristics. For instance, informants may vary in their perspectives as to which mental health concerns in child patients warrant care (Brookman-Fraze, Haine, Gabayan, & Garland, 2008; Hawley & Weisz, 2003; Jensen-Doss & Weisz, 2008; Yeh & Weisz, 2000). In fact, informants may vary in perceiving that a child's mental health concerns distress the child and thus warrant care, even when all informants converge on endorsing that concerns exist (e.g., Phares & Compas, 1990; Phares & Danforth, 1994). Consequently, informants may vary considerably in their perspectives of patients' mental health concerns, and these discrepant perspectives may impact such aspects of mental health care as access to services and therapeutic engagement (see also Hawley & Weisz, 2003; Yeh & Weisz, 2000). Similarly, informants from different ethnic or racial backgrounds may vary in whether or which behaviors reflect mental health concerns and warrant care, although research on the relation between such backgrounds and informant discrepancies has yielded inconsistent findings (see De Los Reyes & Kazdin, 2005; Duhig et al., 2000). Further, low-to-moderate correspondence levels and directional differences in reporting (e.g., whether children self-report greater concerns than parents report about children) similarly characterize cross-informant correspondence in reports examined in over 20 countries (Rescorla et al., 2013; Rescorla et al., 2014).

Few characteristics in the informant discrepancies literature have been given more research attention than the impact of informants' mental health concerns on their reports of children's mental health. Broadly, the conceptual rationale for this work can be encapsulated in what

researchers have termed the *depression*→*distortion* hypothesis: When an informant experiences low mood, this causes the informant to attend to, encode, and thus rate children's behavior with greater negative descriptors, relative to use of positive or neutral descriptors (see Richters, 1992; Youngstrom, Izard, & Ackerman, 1999). Importantly, depression→distortion effects should not be seen as mutually exclusive from the previously described characteristics. For instance, depression→distortion effects may result in depressed informants reporting greater levels of patients' mental health concerns than non-depressed informants. These effects potentially impact patients' access to care if mental health professionals encounter discrepant reports, and these discrepant reports result in uncertainties when making clinical decisions. Among the informants providing reports, parents' reports have been the most frequently studied in terms of depression→distortion effects, perhaps because a common risk factor of children's mental health concerns is a family history of such concerns (e.g., Goodman & Gotlib, 1999). Thus, within clinic samples of child patients, parents providing reports of patients' concerns may often experience a key characteristic thought to bias informants' reports.

A key concern with research on the depression→distortion hypothesis is that the hypothesis does not have strong empirical support on its behalf. As mentioned previously, many studies have tested the depression→distortion hypothesis, with some studies finding that informants experiencing depressive symptoms provide reports that indicate greater levels of children's mental health concerns, relative to reports taken from other informants (De Los Reyes & Kazdin, 2005). However, a number of studies have found no such support (e.g., Conrad & Hammen, 1989; De Los Reyes, Goodman, Kliwer, & Reid-Quiñones, 2010; De Los Reyes, Youngstrom et al., 2011a; Hawley & Weisz, 2003; Weissman et al., 1987). Further, most depression→distortion studies do not include independent ratings of children's behavior in the same situation (see also Richters, 1992), and interactions between informants (e.g., parents and teachers) and the subject of the assessment (e.g., children receiving treatment) are often context-dependent (e.g., occur exclusively in home and/or school contexts). Consequently, observing a child behaving differently with their depressed parent at home than they did with teachers at school might be a logical consequence of the nature of interactions between patients and informants.

Among those studies that have used experimental designs and constrained other possible confounding factors (e.g., context of observation), researchers have observed, at best, modest support for the depression→distortion hypothesis. For instance, Jouriles and Thompson (1993) experimentally induced negative mood states in parents before they viewed a previously recorded task involving a “cleanup” activity with their child, and had parents and independent observers rate the child's behavior during the task. The researchers observed non-significant differences between parents' and independent observers' reports of children's behavior during the task. Other studies using mood induction procedures have also failed to support the depression→distortion hypothesis (Youngstrom, Kahana, Starr, & Schwartz, 2004).

The depression→distortion hypothesis's strongest empirical support comes from a quasi-experimental study examining the relation between mothers' depressive mood symptoms and their reports of children's behavior during completion of a frustrating task (Youngstrom et

al., 1999). Youngstrom and colleagues (1999) controlled for the context on which informants based their reports by comparing mothers' reports of children completing a frustrating task to independent observers' reports of children during the same task. The researchers observed statistically significant depression→distortion effects, and yet these effects only accounted for between 2%-20% of incremental variance in mothers' reports, relative to independent observers' reports. Consequently, one would be hard-pressed to discard mothers' reports based on even the highest observed figure of incremental variance (i.e., 20%), as even this figure constitutes a minority of the variance in the mothers' report that could possibly be “afflicted” with rater bias. Overall, the lack of strong support for depression→distortion effects indicate that these effects cannot fully account for the presence of informant discrepancies.

Measurement error—A construct related to that of informants' perspectives and rater biases is measurement error. Measurement error has key implications for research on the validity of the multi-informant approach. Indeed, measurement error truncates the ability of informants' reports to predict criterion variables, thus hindering the ability of multi-informant assessments to evidence incremental and construct validity (see Dirks, Boyle, & Georgiades, 2011; Dirks et al., 2012). Thus, if measurement error explained informant discrepancies, then mental health professionals could not validly infer that discrepancies among informants' reports reflect changes in displays of patients' mental health concerns across contexts.

The psychometrics literature points to three important components of measurement error that have implications for the multi-informant assessment approach: (a) transient error, (b) random error, and (c) systematic error. The first, transient error refers to characteristics of the rater that might hinder measurement reliability (Schmidt & Hunter, 1996). In many respects, the literature reviewed previously on the depression→distortion hypothesis represents a case of research on transient error in multi-informant assessment. Second, randomly distributed error variance, particularly within the individual informants' reports in a multi-informant assessment, may pose challenges to reliably assessing discrepancies among informants' reports. This is because assessments of discrepancies between informants' reports can only be as reliable as the individual informants' reports from which one assesses discrepancies (e.g., Rogosa, Brandt, & Zimowski, 1982; Rogosa & Willett, 1983). Third and with regard to systematic error, recent work indicates that multi-informant assessments may yield consistent differences between informants' reports across multiple measures (De Los Reyes, Alfano, & Beidel, 2010, 2011; De Los Reyes, Bunnell et al., 2013; De Los Reyes, Goodman et al., 2008, 2010; De Los Reyes, Youngstrom et al., 2011a, b). Yet, the presence of systematic error may result in giving consistent differences between informants' reports the appearance of meaningful differences (see also De Los Reyes, 2013).

Overall, measurement error may pose challenges to validly interpreting the outcomes of multi-informant assessments as reflecting contextual variations in patients' concerns. Thus, researchers have examined components of the measurement process that either constrain levels of such error or improve interpretations of informants' reports. Two research literatures appear particularly pertinent in this regard. First, even subtle changes in measurement scaling can have profound effects on informants' reports. For example, on a

positively valenced self-report measure of life success, scaling items on a “-5 to 5” range produced greater proportions of respondents rating high life success than the proportion observed when the scale ranged from “0 to 10” (for a review, see Schwarz, 1999). Along these lines, a key element of best practices in multi-informant assessment involves ensuring that assessors hold such measurement features as item content, scaling, and response labeling constant across multiple informants' reports. In this way, an assessor can decrease the likelihood that discrepancies among reports are the result of methodological artifacts of the measurement process (for a review, see De Los Reyes, Thomas et al., 2013). In fact, when researchers hold these measurement factors constant across reports, and the measures informants complete exhibit the same factor structures and similar levels of internal consistency, informants nevertheless provide reports that correspond with each other at low-to-moderate magnitudes (e.g., Achenbach & Rescorla, 2001; Baldwin & Dadds, 2007).

At the same time, holding measurement factors constant across informants' reports does not guarantee that the resulting discrepancies reflect contextual variations in patients' concerns. Indeed, informants may nonetheless react differently to the same measurement instrument, and the reasons for these different reactions may or may not reflect differences in the contexts in which informants observe patients' behavior. For instance, in recent work researchers administered assessments of parental knowledge of adolescents' whereabouts and activities (i.e., a risk factor for adolescent delinquency) to a community-based sample of parents and adolescents (De Los Reyes, Ehrlich et al., 2013). Researchers randomly assigned parents and adolescents to the order in which they completed two different assessments of parental knowledge. Within one assessment protocol, parents and adolescents received training on how to use the number of contexts in which they tend to observe behaviors indicative of parental knowledge to provide Likert-type ratings on survey reports about such knowledge (i.e., greater number of contexts → greater ratings; De Los Reyes, Ehrlich et al., 2013). Within a second protocol, parents and adolescents received instructions to complete the assessments that did not involve any training (i.e., a typical questionnaire completion protocol). Interestingly, the training received by parents and adolescents increased the differences between their reports, relative to the no-training condition. These findings suggest that parents and adolescents differ in reports about parental knowledge, in part, because they view behaviors indicative of parental knowledge in different contexts.

In contrast to the findings of De Los Reyes, Ehrlich et al. (2013), in another study clinicians with experience in assessing and treating child patients read vignettes of hypothetical children receiving a screening evaluation for conduct disorder (De Los Reyes & Marsh, 2011). The clinicians read vignettes that described contextual features in a child's life that the researchers randomly manipulated to reflect either risks for conduct disorder (e.g., associating with deviant peers) or no such risks. The vignettes also described the child as evidencing a single symptom within the fourth edition of the *Diagnostic and statistical manual of mental disorders* (DSM-IV) diagnostic criteria for conduct disorder (American Psychiatric Association [APA], 2000). Researchers prompted clinicians to make likelihood ratings of the children that reflected their impressions of whether the children would meet DSM-IV criteria for conduct disorder if a clinician were to administer a full diagnostic

evaluation to the child. Further, the clinicians made judgments for 30 hypothetical children, in order to make likelihood judgments for all 15 DSM-IV conduct disorder symptoms rated in both high-risk and low-risk contexts. Thus, all clinicians in the sample made ratings about the same hypothetical children and conduct disorder symptoms, were exposed to the same rating instructions, and had access to the same information about the hypothetical children's environments. Overall, clinicians made greater likelihood ratings for children described in high-risk contexts versus those described in low-risk contexts. These findings suggest that clinicians tended to apply contextual information to their judgments about the children's conduct disorder symptoms. However, clinicians exhibited little correspondence in terms of the specific symptoms for which they applied contextual information. For instance, whereas high-risk contextual information might have influenced one clinician's judgments about a child's truancy, the same high-risk contextual information might have had little influence on another clinician's judgments about the child's truancy. These findings have since been replicated when examining judgments about attention-deficit/hyperactivity concerns, as well as judgments completed by laypeople (Marsh, De Los Reyes, & Wallerstein, 2014).

The findings of De Los Reyes, Ehrlich et al. (2013) and De Los Reyes and Marsh (2011) indicate that holding measurement factors constant alone does not guarantee interpretability of any one informant's report or the discrepancies between informants' reports. Thus, a second literature involves identifying and using validity criteria for assessing context-specific displays of mental health concerns. Using context-specific criterion measures, mental health professionals might improve understanding of the reasons why informants provide discrepant reports. For example, to assess cross-contextual variations in preschool children's disruptive behavior, researchers recently developed the Disruptive Behavior Diagnostic Observation Schedule (DB-DOS; Wakschlag et al., 2010). This behavioral observation measure consists of assessments of children's disruptive behavior as displayed within interactions between children and adult authority figures. The adult authority figures consist of the assessed child's parent and an unfamiliar clinical examiner. Assessors hold the nature of the interactions constant across adult-child interactions. For instance, the parent prompts the child in one interaction to help with cleaning up toys, and in a separate interaction the clinical examiner administers a similar cleanup prompt to the child. The consistency in the structure of adult-child interactions on the DB-DOS allows one to assess in an analogue sense disruptive behavior as displayed in home contexts (e.g., with parental adults) and/or non-home contexts (e.g., with non-parental adults such as teachers). Thus, the DB-DOS might serve as an independent assessment of whether patients display disruptive behavior consistently across contexts or within specific contexts. Such an assessment could serve as a criterion for assessing the validity in interpreting patterns of informants' reports (e.g., parent and teacher reports) as reflective of contextual variations in patients' concerns. Below, we describe a study that took exactly this approach.

Another example of a context-sensitive validity criterion comes from research on the social interactions that tend to elicit displays of children's behavioral and emotional concerns (Hartley, Zakriski, & Wright, 2011). In this work, researchers administered behavioral checklists for informants (e.g., parents and teachers) to complete about children. The informants also completed measures about how the same children tend to react (e.g., aggressive behavior) within interactions with peers (e.g., peer bosses child around) and adult

authority figures (e.g., adult giving child instructions). This second measure allows researchers to examine whether convergence between informants' reports about children's behavioral and emotional concerns signals similarities in the kinds of interactions encountered across contexts (e.g., child encounters teasing by peers at home and school). Alternatively, one could examine whether divergence between informants' reports about children's concerns reflects the idea that interactions that tend to elicit children's concerns are present in one context (e.g., school) and not another (e.g., home).

The previous examples of context-sensitive validity criteria relied on behavioral observations or survey methods. One final example leverages the time- and context-sensitive properties of physiological arousal. Specifically, individuals experiencing social anxiety vary in that some experience social anxiety in different ways, depending on the context (e.g., Bögels et al., 2010). Direct assessments of physiological arousal may distinguish social anxiety patients' experiences within and across these contexts (e.g., arousal during public speaking vs. arousal in anticipation of public speaking; De Los Reyes, Augenstein et al., 2015). These assessments have great potential as validity indicators to interpret informants' reports (De Los Reyes & Aldao, 2015). This is because technology now allows mental health professionals to assess arousal using wireless, ambulatory devices (e.g., heart rate monitors), and implement them *in vivo* within the laboratory as well as routine clinic settings (see also Thomas, Aldao, & De Los Reyes, 2012).

For example, in one study researchers recruited adolescents who met DSM-IV criteria for social anxiety disorder as well as adolescents who did not meet criteria for any mental disorder (Anderson & Hope, 2009). In this study, adolescents engaged in a series of laboratory social interaction tasks totaling 20 minutes (e.g., one-on-one social interaction and public speaking tasks with trained confederates), to assess adolescents' self-reported changes in arousal in reference to arousal indices of physiological habituation taken during these interactions (i.e., wireless, ambulatory heart rate monitors). Regardless of diagnostic status, adolescents experienced physiological habituation to the social interactions (i.e., sharp increase in heart rate at beginning of task, followed by gradual decrease in heart rate over course of task). Yet, adolescents' self-reports varied as a function of diagnostic status. Specifically, adolescents experiencing social anxiety disorder were more likely than adolescents not meeting criteria for any mental disorder to self-report stable and high levels of arousal from pre-to-post tasks. This observation is consistent with key components of exposure-based therapies for social anxiety. Indeed, these therapeutic approaches often involve training patients to subjectively perceive decreases in physiological arousal within and across social situations that at the outset of therapy patients find anxiety provoking (e.g., Beidel et al., 2007). Overall, these three examples illustrate how mental health professionals might use independent behavioral, survey, and/or physiological assessments to use and interpret informants' clinical reports of patients' concerns.

False-positive and false-negative reports—A set of constructs related to measurement error but with different implications for the multi-informant approach involves concepts drawn from judgment and decision-making research, namely signal detection theory (SDT; for a review, see Swets, Dawes, & Monahan, 2000). Specifically, let us assume that the parent and teacher reports illustrated in Figures 2a and 2b reflected error-

free representations of children's behavior, and the parents and teachers systematically varied in terms of exclusive access to observations of children's behavior as displayed at home and school, respectively. If so, then parents and not teachers would endorse every instance of disruptive behavior displayed at home and not school, whereas teachers and not parents would endorse every instance of disruptive behavior displayed at school and not home. Stated another way, both parents and teachers would evidence perfect rates of positive endorsements of “true cases” of disruptive behavior as displayed within their own context of observation. Parents and teachers would also refrain from endorsing disruptive behavior that only occurred outside of their observational context.

Within SDT, researchers refer to true cases of endorsement as *true positives* and true cases of non-endorsement as *true negatives* (see also Swets, 1992). However, one should not expect any measure (e.g., informant's report), to provide an error-free representation of the construct being assessed. Thus, SDT delineates concepts for representing incorrect positive endorsements for non-cases (i.e., *false positives*) and incorrect non-endorsements for true cases (i.e., *false negatives*). For instance, in a case in which a parent endorsed disruptive behavior that a teacher did not endorse, if in reality the patient did not display any disruptive behavior in any context, the parent report would reflect a false positive. Conversely, both parent and teacher might fail to endorse disruptive behavior in a case in which the patient actually did display disruptive behavior in both contexts; in this case both reports would reflect false negatives.

As in addressing measurement error, one might address false positive and false negative outcomes with an independent, context-sensitive assessment of patients' concerns. Figure 3 provides an illustration of this approach. For example, consider a hypothetical sample of children in which teachers and parents completed standardized behavioral checklists of children's disruptive behavior, and researchers also administered an independent observation of children's disruptive behavior in the home context. In Figure 3, the left-most distribution of scores relates to teacher reports, the center distribution to parent reports, and the right-most distribution to the independent home observation. The x-axis denotes a standardized score (i.e., *T*-score) derived from each assessment to represent children's disruptive behavior. Further, we included a demarcation at the *T*-score “65” to denote the clinical cutoff at which researchers would identify cases of disruptive behavior. The y-axis denotes the frequency for scores at a given *T*-score level. Given standardization of the disruptive behavior scores, all three assessment modalities in the example follow a normal distribution.

In Figure 3, the independent assessment occurred in the home context. Thus, one would expect greater overlap between identified cases of disruptive behavior concerns using the independent assessment and identified cases using parent report (i.e., relative to teacher report). Thus, using the independent assessment as a reference point, researchers could examine whether patients identified via the independent assessment were more likely to be those patients positively endorsed by parents relative to teachers. Stated another way, the independent assessment allows the research team to identify the proportions of parent and teacher reports evidencing true positives and negatives, as well as false positives and negatives. With these proportions, one could estimate how sensitive (i.e., proportion of true positives relative to false positives) and specific (i.e., proportion of true negatives to false

negatives) the parent reports were in detecting home-specific observations of patients' concerns, relative to an informant who observed patients in a context other than home (see also Youngstrom, 2013). One could also structure an assessment similar to that described in Figure 3 to gauge the context-sensitivity of teachers' reports (i.e., independent school observation). Thus, one might incorporate methods from SDT with the OTM to improve use and interpretation of multi-informant assessments.

Informants' cognitive abilities and social desirability—Lastly, how might one interpret the unique qualities of informants' reports, beyond that they reflect contextual variations in patients' concerns, levels of clinical severity, and/or rater biases? That is, discrepancies between adult informants such as parents and teachers might originate from patients' concerns varying across contexts. However, they may also vary in light of the relationship that the patient shares with the informant and the context in which this relationship develops. For instance, a parent may provide a report based on extensive experiences with the patient, and although these experiences may develop over the course of several years, they may not be perceived within the context of normative childhood behavior (De Los Reyes et al., 2009). Conversely, a teacher's relationship with the patient may develop within the context of a single academic year and a large classroom of other students. That is, a teacher's report may reflect a relatively limited time range of experiences with a patient. However, the teacher may have the opportunity to calibrate their report based on normative classroom behavior (see also Drabick et al., 2007).

When researchers compare adult and child informants, informant discrepancies may be attributable, in part, to differences in developing cognitive abilities (e.g., Smetana, Campione-Barr, & Metzger, 2006; Spear, 2000). In the context of patient care, differences between adult reports about patients and patient self-reports may occur as a result of social desirability concerns on the part of the patient, although studies yield inconsistent support for these concerns as contributors to informant discrepancies (for a review, see De Los Reyes & Kazdin, 2005). These differences between adult informants and child informants may indicate, for instance, that adult reports reliably and validly reflect patients' concerns to a greater extent than patient self-reports.

Recent evidence indicates that differences between parent reports and child reports may also reflect meaningful variation between their perceptions, in addition to variations in cognitive abilities and social desirability concerns. For instance, in structured diagnostic interview assessments of child anxiety, interviews based on parent reports and child reports exhibit greater correspondence when reports are about directly observable anxiety behaviors displayed in non-school contexts (e.g., behavioral avoidance displayed at home) relative to internal anxiety behaviors such as worry displayed in school contexts (Comer & Kendall, 2004). Further, recent work in an age- and gender-matched sample of clinic-referred adolescents and community control adolescents found that clinic-referred adolescents self-reported lower levels of social anxiety relative to their parents' reports, and adolescents' self-reports exhibited little to no correspondence with objective measures of psychophysiology (i.e., during a baseline psychophysiological assessment; De Los Reyes, Aldao et al., 2012). Yet, in this study adolescents nonetheless provided internally consistent self-reports that evidenced convergent validity and differentiated clinic-referred from community control

adolescents. This study is an important contribution to the literature on social desirability. Indeed, clinic samples of anxiety patients comprise key settings within which patients' social desirability has been of primary concern to mental health professionals (e.g., Dadds, Perrin, & Yule, 1998; DiBartolo, Albano, Barlow, & Heimberg, 1998; Grills & Ollendick, 2003; Rapee et al., 1994; Silverman & Rabian, 1995). Importantly, De Los Reyes, Aldao et al. (2012) did not directly assess social desirability. Yet, the authors made a key observation that researchers often interpret as indicating social desirability—relative under-reporting via patient self-reports relative to adult informants' reports about patients—while also finding support for the internal consistency and validity of scores taken from patient self-reports. Thus, informant discrepancies alone should not be interpreted as reflecting developmental differences in cognitive abilities or social desirability in informants' reports. Rather, mental health professionals should evaluate whether these discrepancies relate to independent assessments of informants' cognitive abilities or social desirability.

Incremental Validity in Multi-Informant Mental Health Assessments

Theoretical and Methodological Approaches to Incremental Validity

A full account of the validity of multi-informant assessments involves examining the unique contribution of one informant's report over another's, or the incremental validity of each report in the prediction of clinical outcomes (Figure 2). Yet, despite repeated appeals for more investigations of incremental validity in psychological assessment, a dearth of information in this area remains (Hunsley & Meyer, 2003). Nevertheless, prior work on incremental validity in such areas as test construction (Haynes & Lench, 2003; Smith, Fischer, & Fister, 2003), adult psychopathology (Garb, 2003), and use of projective tests (e.g., Dawes, 1999) has key implications for our discussion of incremental validity. In particular, two ideas informed our examination of the incremental validity of multi-informant clinical child assessments.

Expertise in assessment—First, the Rorschach Inkblot Test is a projective test that involves an assessor soliciting patients' open-ended, verbal descriptions of ambiguous stimuli (i.e., a series of blotches of black ink set against a plain white background; for a review, see Hunsley & Lee, 2010). Patients' responses are thought to reveal important aspects of their personality functioning from a psychoanalytic perspective, and thus scoring responses often involves use of complex procedures that require a substantial amount of expertise on the part of the assessor (i.e., knowledge of psychoanalytic theory and practice; see Blais, Hilsenroth, Castlebury, Fowler, & Baity, 2001). However, Rorschach responses can also be scored using less onerous procedures that require little expertise to carry out (e.g., assessing the “match” between a patient's response about the shape of an inkblot and the actual shape of the inkblot). Further, personality functioning can be assessed with paper-and-pencil inventories that can be administered, scored, and interpreted by personnel with little training. Consequently, Dawes (1999) argues that a proper evaluation of incremental validity of the Rorschach involves assessing whether scores that require a great deal of expertise to obtain hold incremental value above-and-beyond the explanatory value of scores that require less expertise to obtain.

In clinical child assessments, the multi-informant approach poses incremental validity issues similar to that of the Rorschach and delineated by Dawes (1999). In essence, do clinical child assessments require the *collective expertise* of multiple informants for one to obtain an adequate sense of the context(s) in which a patient displays mental health concerns? Indeed, the multi-informant approach is a comprehensive assessment procedure that incurs considerable resources in terms of time needed to administer, score, and interpret instruments, and staffing and equipment needed to carry out the assessments (see also Yates & Taub, 2003). A less onerous and comprehensive assessment might effectively address the goals of assessing a patient, such as use of a single informant's report. Thus, demonstrating that the multi-informant approach yields incremental value over, for instance, a single informant's report involves testing whether each informant observes patients' concerns in a particular context that other informants have relatively fewer or no opportunities to observe. That is, the expertise of each informant should result in reports from which scores reflect different *facets* of patients' concerns (i.e., different contexts in which patients display concerns; see also Smith et al., 2003).

Criterion variables and criterion contamination—A second core idea driving our examination of the incremental validity of multi-informant clinical child assessments follows from ideas about expertise advanced by Dawes (1999). Specifically, as in other cases of validity (e.g., construct and predictive), the incremental validity of scores from a given measure may vary as a function of the criterion variable(s) examined (e.g., diagnostic status and treatment response) and the other measures to which it is compared (e.g., Haynes & Lench, 2003). These two factors give rise to two additional challenges to using and interpreting multi-informant assessments. The first challenge is to design incremental validity studies involving multiple predictor variables (e.g., multiple informants' reports). Predictors should evidence bivariate relations with the criterion variable(s) but share little overlapping variance with each other (see also Blais et al., 2001). For instance, use of parent reports and teacher reports to predict clinicians' impressions of patients' behavior across home and school contexts capitalizes on use of the context-specific expertise of each informant (i.e., home-specific vs. school-specific) to predict clinicians' reports.

However, addressing challenges with overlapping variance among predictor variables raises a second challenge. Specifically, if the outcomes of incremental validity studies vary depending on the criterion variable, then these studies should be designed such that differences in findings among the predictor variables should to the greatest extent possible reflect “true differences” among predictors in their incremental validity. For instance, when comparing reports within a multi-informant assessment, one should ensure that measurement content be held constant across the informants' reports. Yet, holding these measurement factors constant across reports may do little to address a key issue raised by Garb (2003): criterion contamination. In incremental validity research, *criterion contamination* refers to differences in incremental validity among predictors that arises, in part, because the criterion variable was based on information from one or more of the predictors. For instance, if a study's validity criterion consisted of clinicians' diagnostic judgments about patients, the study might suffer from criterion contamination if prior to forming their judgments,

clinicians had access to information from the predictors evaluated for incremental validity (see also Bossuyt et al., 2003; Whiting et al., 2011).

Incremental Validity in Multi-Informant Clinical Child Assessments

We evaluate the incremental validity of multi-informant clinical child assessments in light of key theoretical and methodological contributions from prior work. As in the incremental validity literature reviewed previously, the incremental validity of reports about child patients may be influenced by measurement methodology, the mental health domain assessed, and criterion contamination (e.g., Johnston & Murray, 2003; Pelham, Fabiano, & Massetti, 2005). In Table 2, we illustrate key findings from prior work. Importantly, few incremental validity studies exist for internalizing concerns such as anxiety, relative to externalizing concerns such as attention and hyperactivity and conduct problems (for reviews, see Dirks et al., 2012; Silverman & Ollendick, 2005; Tulbure, Szentagotai, Dobrea, and David, 2012). Thus, we focus our attention on the incremental validity of multi-informant assessments of children's externalizing concerns, and where possible we incorporate work on children's internalizing concerns.

Measurement method of informants' reports evaluated for incremental validity

—Across both clinical and research settings, informants may provide reports via a variety of measurement methods, including rating scales and structured or semi-structured interviews. Under certain circumstances (e.g., diagnosing childhood attention-deficit/hyperactivity disorder [ADHD]), research supports the construct validity of these approaches (Pelham et al., 2005). Yet, scores from an informant's report may evidence construct validity and still fail to incrementally contribute information beyond that contributed by other measures within an assessment battery (i.e., incremental validity; Johnston & Murray, 2003). For example, in a review of evidence-based assessments for ADHD, Pelham and colleagues (2005) examined the incremental validity of rating scales and structured interviews for the assessment of childhood ADHD. Despite evidence of the construct validity of each measure, Pelham and colleagues (2005) found rating scales to be the most incrementally valid for assessing ADHD, and use of a DSM-IV-based structured interview did not contribute unique information beyond rating scales. Yet, as we discuss below, issues with criterion contamination hinder the interpretability of prior work on the incremental validity of measures used in clinical child and adolescent assessments.

Mental health domain or focus of assessments—Another factor that can affect the incremental validity of a predictor is the domain being assessed or the goal of assessment. The incremental validity of an informant's report may vary dramatically when assessing externalizing symptoms versus internalizing symptoms, or when the goal of the assessment is to reach a single diagnosis versus ruling out a diagnosis (Table 2). For instance, returning to Pelham and colleagues' (2005) review, research suggests that a single teacher rating of ADHD sufficiently captures the presence of ADHD for research purposes. However, ratings from both parents and teachers are important if the goal of the assessment is to reach a clinical diagnosis, given the diagnostic requirement of symptom displays across multiple contexts. Similarly, rating scales may suffice for identifying ADHD symptoms, and yet ruling out diagnoses other than ADHD (i.e., for purposes of differential diagnosis) requires

gathering additional information regarding other mental health domains (e.g., anxiety and conduct problems; Pelham et al., 2005).³ As with incremental validity research on measurement methods (e.g., rating scales vs. structured interviews), we discuss below issues regarding how criterion contamination hinders the interpretability of prior work examining incremental validity and use of multiple informants' reports (e.g., parents vs. teachers) and measurement of multiple mental health domains.

Criterion variables and criterion contamination—Crucially, estimates of incremental validity for multi-informant clinical child assessments may vary as a function of the validity criterion or predictors (e.g., Johnston & Murray, 2003; Pelham et al., 2005). How might the choice of the validity criterion and predictors influence conclusions about the incremental validity of multi-informant assessments? We address this question in two ways.

First, how do multi-informant assessments perform if one holds measurement method, item content, and an independently assessed correlate constant across informants' reports? To address this, we present in Table 3 findings reported by Achenbach and Rescorla (2001) for cross-informant correspondence for parent report, teacher report, and child self-report forms of the Achenbach System of Empirically Based Assessments (ASEBA), as well as relations among these reports and children's referral status. Briefly, the ASEBA battery consists of parent, teacher, and child reports that exhibit an identical format (i.e., behavioral checklist) and response scaling (i.e., response values of *none*, *some*, and *a lot*), with some minor modifications to item content to match each informant's perspective. Each form assesses syndromes reflecting various children's mental health domains that cluster around two broadband problem domains: internalizing (e.g., anxiety, mood, somatic complaints) and externalizing (e.g., aggressive and rule-breaking behaviors). Further, Table 3 reports relations between the informants' reports and a single correlate, namely children's referral status (i.e., differentiating clinic referred vs. non-referred children).

In Table 3 we report two sets of findings regarding the ASEBA forms. Specifically, we present statistical estimates reported by Achenbach and Rescorla (2001). Specifically, we report Pearson *r* correlation coefficients among parent, teacher, and child reports, which we derived from Table 9-2 from Achenbach and Rescorla (2001). As reported in Achenbach and Rescorla (2001), each value denotes a significant relation between informants' reports, $p < .05$. Consistent with meta-analytic findings reviewed previously, levels of informant correspondence (a) ranged from low-to-moderate in magnitude and (b) were larger for correspondence on reports of externalizing relative to internalizing problem behaviors. Based on these cross-informant correspondence estimates, we calculated the common variance shared by multiple informants for the internalizing, externalizing, and total problem scores on the ASEBA forms. We made these calculations based on regression commonality analyses illustrated by Cohen, Cohen, West, and Aiken (2003) and Nimon (2010). These

³Additional factors such as the patient's age and gender, cultural characteristics of the assessment setting, and the base rates of the symptoms may affect the generalizability of information regarding the incremental validity of assessment methods. Therefore, these other issues must be taken into account before generalizing findings from one study to new samples or patient populations (Johnston & Murray, 2003). An example of this effect can be seen in work examining the incremental validity of multiple informants' reports of peer victimization in the prediction of relational adjustment in elementary school-aged children (Ladd & Kochenderfer-Ladd, 2002). In this study, the unique information contributed by reporters varied as a function of children's grade level (Table 2).

estimates (ranging from 0.8% to 3.0%; Table 3) reveal that the proportion of variance in problem scores that is captured by all informants is very small.

We also report the percentages of variances in informants' reports that are explained by children's referral status. We derived these percentages of variance from Table 10-3 from Achenbach and Rescorla (2001). Specifically, the authors conducted multiple regression analyses using referral status as a predictor variable along with a series of children's demographic variables as predictor variables (i.e., age, socioeconomic status, ethnicity). In separate analyses, each of the parent, teacher, and child reports served as dependent variables. Following these analyses, the authors converted the regression coefficients representing unique effects of children's referral status on informants' reports to Fisher's z , averaged these Fisher's z s across gender and age groups, and converted these final Fisher's z s to Pearson r s. The percentages of variances reported in Table 3 are squared estimates of these Pearson r s. These values denote significant differentiation between referred and non-referred children for all informants' reports ($p < .01$), with lower scores for non-referred children relative to referred children. That is, each report (parent, teacher, child) significantly related to children's referral status. Based on these percentages of variances estimated by Achenbach and Rescorla (2001), we estimated correlations between referral status and the individual informants' reports, as well as a total correlation between referral status and the three informants' reports. Specifically, to estimate correlations for individual informants, we took the square root of the percentage of variance accounted for by referral status separately for parent report, teacher report, and child report. To estimate maximum possible correlations for the multi-informant reports, we took the square root of the sum of the common variance across multiple informants and the percentage of variance accounted for by referral status for the parent, teacher, and child reports (Cohen et al., 2003). By "maximum," we mean that we estimated an upper limit correlation between informants' reports and referral status, assuming unique variance among the informants' reports.

In comparing the coefficients observed for individual and multi-informant correlations, a clear pattern of findings emerges. Specifically, one observes greater magnitudes of correlations between referral status and an approach based on the multi-informant parent, teacher, and child reports, relative to an approach based on use of any one of the individual parent, teacher, or child reports. Overall, when using reports that hold item content and scaling constant across informants and in relation to assessments of children's referral status, multiple informants' reports of children's internalizing and externalizing concerns may each relate to referral status. However, in each case (i.e., relations between referral status and internalizing, externalizing, or total broadband problems), a multi-informant assessment approach yields greater correlations with referral status than an approach based on any one individual informant's report. In light of our findings, our approach to illustrating the incremental validity of multi-informant assessments evidenced a number of key strengths. Specifically, each informant's report was assessed using the same reporting format (i.e., survey checklist) and item content, and incremental validity was evaluated using a criterion measure that evidenced relatively little criterion contamination. Yet, one limitation is that the findings in Table 3 were based on the scale scores reported by Achenbach and Rescorla (2001) for the complete parent, teacher, and child ASEBA scales (Leslie Rescorla, Personal

Communication, September 10, 2014). Although these scales share many items in common, each contain items that the other scales do not (e.g., teacher report contains items to assess attentional concerns that the other reports do not contain). Consequently, we suspect the cross-informant correspondence estimates reported in Table 3 are conservative and would be slightly higher if these estimates were based strictly on items common across the informants' instruments. Nevertheless, previous work examining cross-informant correspondence specifically on ASEBA items that were common across informants' reports found low-to-moderate correspondence between informants' reports (e.g., De Los Reyes, Youngstrom et al., 2011a, b).

In light of the findings in Table 3, a second question arises: What happens when a validity criterion suffers from criterion contamination, as described by Garb (2003)? Indeed, in some of the incremental validity studies in Table 2, the validity criterion consisted of clinicians' diagnoses based, in part, on information provided by multiple informants, indicating non-independence between the informants' reports and the criterion measure. In addition, in some of the studies in Table 2, when clinicians gathered information from multiple informants they also knew which informant referred the patient for clinical services and which informants served as “collateral” informants. Importantly, in these studies clinicians would have been likely to encounter instances in which informants' reports pointed to inconsistent conclusions regarding patients' concerns and whether these concerns warranted additional clinical services (see also Yeh & Weisz, 2001).

Two lines of research indicate that the designs of the incremental validity studies described in Table 2 may compromise interpretations of research on the incremental validity of multi-informant assessments. First, mental health professionals may view specific informants as “optimal” for assessing specific mental health concerns, such as viewing teachers as “better” informants than parents for assessing child hyperactivity (e.g., Bird, Gould, & Staghezza, 1992; Loeber, Green, Lahey, & Stouthamer-Loeber, 1989; Loeber, Green, & Lahey, 1990).

Second, compounding issues about “optimal informants” is the fact that a single informant (e.g., parent) usually initiates a patient's clinical services (Hunsley & Lee, 2010). Among other pieces of clinical information, the referral informant provides an assessor with a referral question (e.g., *Does the patient evidence concerns with depressive mood symptoms?*) (Garb, 1998). This referral question typically informs assessors' initial hypotheses as to the patients' concerns and thus the key goal of the assessment (Croskerry, 2003). Thus, when clinicians make decisions based on multi-informant assessment outcomes, their decisions may align with referral informants' reports more so than other informants' reports, even without compelling evidence to indicate the validity of this approach. Note that clinicians' prior knowledge of which informant served as the referral source and which informants served as collateral informants should be distinguished from clinicians having access to the outcomes of multi-informant assessments prior to making clinical decisions. Studies rarely note whether clinicians were kept blind to all clinical information collected apart from their own clinical ratings or tools used to make clinical decisions (e.g., diagnostic interview information; for an exception, see Carlson & Blader, 2011). That being said, studies using clinicians' reports as a criterion variable when clinicians also had access to the multi-informant assessments evaluated for incremental

validity would have been likely to compound the interpretative problems raised by clinicians having knowledge of the identity of the referral source (e.g., Bossuyt et al., 2003; Whiting et al., 2011).

Regardless of the mechanisms underlying reliance on a single informant to make clinical decisions, this practice has important implications for clinical decision-making. That is, presumably approaches to clinical decision-making that rely on one informant's report more so than other informants' reports most often take the form of *disjunctive* decision-making strategies (i.e., “OR” rule; for a review see Piacentini, P. Cohen, & J. Cohen, 1992), whereby the informant relied upon most, for instance, to plan treatment for a patient might also be identifying mental health concerns that go uncorroborated based on other informants' reports. A strength of this approach is that it can increase access to mental health services for patients with impairing symptoms because only a single informant would have to report concerns for the patient to receive care. However, relying on a single report introduces considerable measurement error in decision-making and may lead to the *false positive* test outcomes described previously (e.g., Hunsley, 2003; Swets et al., 2000; Youngstrom, Findling, & Calabrese, 2003). Yet, to what extent do the data indicate that such clinical decisions occur?

In Table 4, we summarize the findings of several studies on the relation between informant discrepancies and decisions that clinicians make regarding ratings of children's functional impairment, diagnosis, treatment response, and treatment planning. For four of the studies described in Table 4 (De Los Reyes, Alfano et al., 2011; DiBartolo, et al., 1998; Grills & Ollendick, 2003; Youngstrom, Findling, & Calabrese, 2004b), an overall pattern of findings emerges: When correspondence between parent reports and child self-reports is low, clinicians tend to make diagnostic decisions and ratings of treatment response that correspond more with parent reports than child self-reports. For three studies (Brown-Jacobsen, Wallace, & Whiteside, 2011; Hawley & Weisz, 2003; Kramer et al., 2004), whether clinicians corresponded in their clinical decisions more with parent reports or child self-reports depended on the measurement method or problem domain assessed. For instance, given high disagreement among parents, patients, and clinicians in terms of which problems to target in therapy, clinicians' decisions regarding therapeutic targets agreed more with parent reports than patient self-reports when the target concerned a patient problem (e.g., children's anxiety, mood, attention concerns) (Hawley & Weisz, 2003). However, in this same study, clinicians agreed more with patient self-reports than parent reports when the target concerned problems with the family or the environment.

When interpreting the findings summarized in Table 4, it is important to further scrutinize the studies and the quality of the informants' reports. Indeed, one possible interpretation of these findings is that clinicians rightfully identified parents as the “most valid” informants for the particular mental health concerns assessed. A key characteristic of two of the studies in Table 4 is that the measures administered to parents and children were identical to or slightly modified versions of those used in our illustration of the incremental validity of multi-informant assessments (i.e., Hawley & Weisz, 2003; Youngstrom et al., 2004b). One of the other studies administered functional impairment measures to parents and adolescents that share many of the characteristics of the measures used in our Table 3 illustration (e.g.,

similar item content, focus on last six months of functioning, Likert-type scaling; Kramer et al., 2004). Thus, differential measurement validity between parent and child reports may not parsimoniously explain findings from Hawley and Weisz (2003) and Youngstrom and colleagues (2004b).

The remainder of the studies described in Table 4 focused on assessments of child anxiety symptoms and diagnoses (Brown-Jacobsen et al., 2011; De Los Reyes, Alfano et al., 2011; DiBartolo et al., 1998; Grills & Ollendick, 2003). With child anxiety concerns, it is important to note that decades of research attest to the reliability and validity of child self-reports for assessing these concerns (for reviews, see Silverman & Ollendick, 2005; Tulbure et al., 2012). As mentioned previously, the lack of correspondence between parent reports and patient self-reports of child anxiety does not necessarily indicate that patients provide self-reports of poor quality (e.g., Comer & Kendall, 2004; De Los Reyes, Aldao et al., 2012). In fact, even when parent reports and child self-reports of child anxiety yield low correspondence, these reports may all exhibit identical factor structures (Baldwin & Dadds, 2007). Most crucially, recent work suggests that children provide incrementally valid self-reports of anxiety, relative to parent reports, in the prediction of child anxiety disorder diagnoses (Villabø et al., 2012).

Overall, when mental health professionals encounter discrepant assessment outcomes from multiple informants' reports, they make clinical decisions that tend to correspond with a *single* informant's report to a greater extent than the reports of other informants involved in the assessment. Sometimes in the cases of child patients, this informant may be a non-parental informant (e.g., patient); however, most often this informant is the parent, or the informant who mental health professionals rely on most to initiate children's clinical services. Consequently, study design considerations in incremental validity research preclude firm conclusions as to the support for use of the multi-informant approach in clinical child assessments. To further illustrate these issues, in Table 5 we describe the characteristics of the incremental validity studies reviewed in Table 2. These characteristics highlight three issues concerning prior work. First, studies on use and interpretation of multiple informants' reports require, at a minimum, informants to complete reports that hold measurement content, scaling, and response options constant across reports to control the methodology through which informants provide reports (De Los Reyes, Thomas et al., 2013). However, at times incremental validity studies involve informants completing reports via measures that contain different items, scaling, and response options (Table 5). Here, findings may reflect differential psychometric properties among measures, rather than evidence supportive of the incremental validity of reports.

Second, studies have not adequately controlled for criterion contamination (Table 5). For instance, many studies evaluate incremental validity using clinicians' diagnostic decisions (i.e., via structured interviews administered to informants) as the key criterion by which to compare informants' reports. Yet, these criterion measures were derived through clinical interviews made by evaluators who had access to the reports of the same informants being evaluated for incremental validity (Table 5). Based on the findings reported in Table 4, this is an important consideration: In the presence of discrepant informants' reports, clinical evaluators tend to make clinical decisions that most closely correspond with the informant

initiating clinical services. In these studies, one interpretation is that the referral informant (e.g., parent) provided incrementally valid information not because of the validity of their reports per se, but because they informed assessors' hypotheses regarding the referral question and thus rationale and structure of the clinical evaluation (Croskerry, 2003; Garb, 1998).

Third, contextual variations in patients' mental health concerns may influence the extent to which a given informant's report is found to be incrementally valid (Figure 2). For example, consider the idea that a given mental health concern, such as social anxiety (Bögels et al., 2010), may manifest itself differently from one patient to the next (see Figures 2a, 2b, and 2c). Further, consider that a researcher conducts an incremental validity study using three informants' reports, namely parents, teachers, and peers. In this study, patients in the sample only displayed mental health concerns within a single context (i.e., home), and thus only one of the three informants (i.e., parents) primarily observed patients in this context. The results of this study may yield incremental validity evidence in support of parent reports and not the teacher and peer reports, but only because parents had access to observing the behaviors assessed, and not teachers and peers. Unfortunately, although some of the incremental validity studies in Table 5 included independent measures (i.e., beyond the informants' reports examined) of patients' mental health concerns, none assessed and reported the prevalence rate of context-specific displays of patients' mental health concerns. Stated another way, previous incremental validity studies have not reported estimates of which patients displayed their concerns within specific contexts (e.g., home or school) or across multiple contexts. Without the availability of these data, one cannot address the methodological confound that the study's findings in support of a given informant's report may have arisen because of a lack of variation in the contexts in which patients displayed concerns (cf. Figures 2a and 2b). Consequently, findings from such a study might not generalize to other samples of patients who display qualitatively distinct contextual variations in mental health concerns. In sum, study design issues preclude our ability to discern whether multi-informant clinical child assessments demonstrate incremental validity.

Construct Validity of Multi-Informant Mental Health Assessments

Theoretical and Methodological Approaches to Construct Validity

Our perspective on the construct validity of multi-informant assessments draws from the definitions provided by Borsboom and colleagues (2004). We also draw from core principles of convergent and divergent validity, and the multi-trait/multi-method (MTMM) matrix (e.g., Campbell & Fiske, 1959; Westen & Rosenthal, 2003, 2005). Specifically, in a multi-informant assessment of a given domain of patients' concerns (e.g., aggressive behavior), scores from any one informant's report may validly reflect levels of concerns on that domain. However, one informant's report alone cannot estimate whether patients display concerns across contexts (e.g., Kraemer et al., 2003); knowledge of multiple contexts requires multiple informants. If we were to use MTMM matrix terminology (see Campbell & Fiske, 1959), our focus is on assessments in which (a) multiple informants complete reports about the same domain (i.e., mono-domain), (b) each informant's report reflects observations within a single context (i.e., mono-context), and (c) patterns of scores across

multiple informants reflect contextual variations in patients' concerns (i.e., multi-context). By examining patterns of scores from these “mono-domain/multi-context” assessments, one can assess the construct validity of multi-informant assessments. Specifically, if an independent assessment of patients' concerns reveals that they contextually vary in where they display concerns, then one should observe large discrepancies among the reports of informants who observe patients' behavior in different contexts. Alternatively, if an independent assessment reveals that patients display concerns consistently across contexts, then one should observe high correspondence among informants' reports of patients' concerns. These patterns would, in effect, support the construct validity of interpreting multi-informant assessments as reflecting the extent to which patients contextually vary in displays of their concerns. To this end, we review research taking exactly this approach to understanding the construct validity of scores taken from multi-informant clinical child assessments.

Construct Validity in Multi-Informant Clinical Child Assessments

Until recently, the best evidence in support of the construct validity of multi-informant assessments came from the Achenbach et al. (1987) meta-analysis, as described previously. However, these findings provide an incomplete picture of construct validity. That is, similar to the incremental validity research reviewed previously, strong construct validity evidence would involve studies that compare the points of convergence and divergence between informants' reports to independent and contextually sensitive measures of patients' concerns (e.g., DB-DOS; see Wakschlag et al., 2010). With this independent assessment, one can examine whether an informant's report relates most strongly with behavioral observations of patients within the observational context that most closely matches their own context of observation (e.g., parent at home vs. teacher at school; see Figure 2). Interestingly, recent research has expanded upon work by Achenbach and colleagues (1987) by examining multi-informant assessments in reference to independent assessments of contextual variations in patients' mental health concerns.

Two studies support the construct validity of the multi-informant approach. First, a study of 327 preschool children aged 3-5 years found that children varied in the contexts in which they displayed disruptive behavior symptoms (De Los Reyes et al., 2009). Specifically, on the DB-DOS described previously, 29.4% of the children displayed disruptive behavior exclusively within interactions with parental adults, and 15% displayed disruptive behavior exclusively within interactions with non-parental adults (i.e., unfamiliar clinical examiners). Further, 8.8% of the children displayed disruptive behavior across interactions with parental and non-parental adults, and 46.8% did not display disruptive behavior within any of these interactions. Thus, De Los Reyes and colleagues (2009) examined a heterogeneous sample of children whose disruptive behavior was specific to home rather than non-home contexts (Figure 2a), specific to non-home rather than home contexts (Figure 2b), or non-specific or displayed across home and non-home contexts (Figure 2c), in addition to children who were unlikely to display disruptive behavior in any of the contexts. Further, in this study, parents' reports of children's disruptive behavior more closely “matched” how children behaved within interactions with parental adults than with non-parental adults, whereas teachers'

reports more closely matched how children behaved within interactions with non-parental adults than with parental adults.

Second, findings consistent with those of De Los Reyes and colleagues (2009) have been observed in a study of early adolescents ($N = 123$; mean age = 13.30), which focused on links between parents' and teachers' reports of aggressive behavior and the social experiences that tended to elicit these behaviors (Hartley et al., 2011). Specifically, Hartley et al. (2011) found that as parents and teachers provided increasingly similar reports of children's aggressive behavior, the similarities increased between the kinds of social experiences in which parents and teachers reported observing the children exhibiting aggressive behavior (e.g., peer interactions or receiving instructions from adult authority figures). Overall, when assessing children for displays of observable mental health concerns (i.e., aggressive and disruptive behavior), preliminary evidence indicates that informants who differ in where they observe children (i.e., parents and teachers) provide reports that map onto contextual variations in such concerns.

Discussion

Main Findings

In this paper, we reviewed research on the validity of the multi-informant approach to clinical child assessment. We conducted a quantitative review of 341 studies reporting estimates of cross-informant correspondence in reports about children's mental health (Table 1), representing the last quarter-century of research on the topic. We used the findings from this review as a backdrop for discussing recent conceptual work on how multi-informant assessments may reveal information about the specific contexts in which patients display mental health concerns. Guided by this conceptual work, we made specific predictions about the incremental and construct validity of multi-informant assessments. Based on these predictions, we evaluated whether the evidence supports using multi-informant assessments to measure context-specific displays of children's concerns. We sought to inform future research, as well as make recommendations for clinical practice on using and interpreting multi-informant assessments.

We made three findings. First, the strongest evidence in support of the multi-informant approach comes from studies of cross-informant correspondence. Our quantitative review indicated that multiple informants' reports share little variance with each other (Table 1). Further, moderators of cross-informant correspondence indicated that informants' reports tend to correspond at higher levels when informants (a) provide reports about concerns that are relatively easy to observe (e.g., externalizing vs. internalizing concerns), (b) observe children's concerns from the same setting (e.g., mother and father reports vs. parent and teacher reports), and (c) complete reports using dimensional measures (i.e., relative to categorical measures). Second, research on the incremental validity of multi-informant assessments is relatively underdeveloped relative to research on cross-informant correspondence. Importantly, our review should not be taken as an indication that multi-informant assessments *cannot* be incrementally valid. Rather, incremental validity studies generally have been constructed to address research questions such as, *Which specific informant(s) provide incrementally valid reports within assessments of mental health*

domain X and patient group Y? In fact, there is a dearth of literature systematically evaluating *which sets of informants* provide incrementally valid reports. Such work might take the form of studies constructed to address research questions such as, *For mental health domain X and patient group Y, do multi-informant assessments consisting of parent, teacher, and child reports provide incrementally valid information, relative to use of any one informant's report?* Third, our review of the construct validity of multi-informant assessments focused on studies examining the “match” between informant discrepancies in reports of patients' concerns and independent, context-sensitive assessments of patients' behavior. Two studies support the ability of multi-informant assessments to reflect contextual variations in patients' behavior for parent and teacher reports of (a) preschoolers' disruptive behavior and (b) early adolescents' aggressive behavior. Thus, studies support the construct validity of the multi-informant approach, and yet for a limited set of informants, mental health domains, and patients' developmental periods.

Recommendations for Clinical Practice

A key goal of this paper was to provide individuals conducting clinical work guidance on when the evidence supports interpreting the outcomes of multi-informant assessments as reflections of contextual variations in patients' mental health concerns. To this end, we identified evidence indicating that at times, multi-informant clinical child assessments pinpoint the specific contexts in which patients display mental health concerns. However, currently we do not have “best practices” for using and interpreting multi-informant assessments in clinical work. In line with these gaps in best practices, we can make three recommendations for applying the multi-informant approach to clinical work.

Using multi-informant assessments to personalize patient care—First, a key implication of the work reviewed above is that practitioners may observe significant within- and between-patient variations in terms of the specific contexts in which patients display concerns (Figure 2). This points to a key idea: Not all patients require or may be suited for the same mental health assessment procedures administered in a single way (e.g., NIMH, 2008). In line with this view, providing proper patient care may involve practitioners seeking to personalize assessments to optimally fit or address patients' unique needs (Figure 2).

In an effort to innovate personalized approaches to mental health care, the NIMH recently released a Strategic Plan for its funding priorities, which includes broadening outcome measurement so as to tailor assessments and interventions to patients' unique needs (Strategy 3.2; NIMH, 2008). Consequently, if a clinician wishes to administer a mental health assessment that is personalized to the unique needs of each patient, the work reviewed above indicates that this assessment should be able to capture information about the specific contexts in which the patient requires care. In this respect, future work may inform the development of contextually sensitive assessments that improve patient care.

For example, as mentioned previously social anxiety patients vary in that some experience social anxiety across social contexts and others within specific contexts (i.e., public speaking; Bögels et al., 2010). This creates the potential for patients to display anxiety in

some contexts but not others. Consequently, a practitioner ought to administer treatment planning assessments personalized to the unique needs of the patient. For instance, assessments can be tailored to classify different patients who display social anxiety mostly during public speaking (i.e., performance anxiety), or both preceding (i.e., anticipatory anxiety) and during public speaking. Based on these assessments, treatments can be tailored to patients experiencing predominantly performance anxiety (e.g., behavioral exposure) and/or anticipatory anxiety (e.g., emotion regulation strategies). Thus, practitioners might leverage assessment methods to identify and implement treatment techniques tailored to each patient. Such assessments may result in cost-effective symptom reduction and improved functioning, relative to existing treatments that may or may not effectively target the unique needs of each patient.

Developing new multi-informant assessment procedures—Second, new methods for using and interpreting multi-informant assessments might assist in developing the personalized assessment approaches we just described. To this end, recent theoretical work on interpreting multi-informant assessment outcomes in research (e.g., OTM; De Los Reyes, Thomas et al., 2013) may assist in future efforts toward developing evidence-based practices in interpreting multi-informant assessment outcomes. Specifically, a key feature of the OTM is based on the idea that practitioners have a great deal of evidence indicating that multiple informants' reports will often yield discrepant outcomes and in predictable ways (e.g., Achenbach et al., 1987; De Los Reyes & Kazdin, 2005; Kraemer et al., 2003). As a result, the OTM guides practitioners toward (a) hypothesizing patterns of convergence and divergence among informants' reports and then (b) constructing assessments that directly test these hypothesized patterns.

For example, a practitioner with a referral question involving an initial assessment for a child's hyperactivity might look to the research literature on typical rates of reporting discrepancies between parent and teacher hyperactivity reports. The practitioner could also conduct a short unstructured phone interview with the parent and teacher before conducting the assessment. Taking precautions not to ask leading questions that might bias informants' reports (Groth-Marnat, 2009), the practitioner could probe for similarities and differences between the home and school contexts, such as in how parent and teacher manage the child's behavior within and across contexts. Using this contextual information as a guide, the practitioner would predict the patterns of information to be gleaned from parent and teacher reports. The practitioner would then structure the assessment to make sense of the parent and teacher reports.

Specifically, based on the qualitative data the practitioner might expect the patient's hyperactivity to vary by contextual demands. These variations might reflect that adult authority figures within the home and school contexts (i.e., parents vs. teachers) differ in the strategies used to manage the patient's behavior, with more effective management strategies present in the school context. These variations might result in the practitioner predicting informants' reports across contexts to vary in levels of hyperactivity, with greater levels of hyperactivity reported in the home context, relative to the school context. The result may be collecting reports of children's hyperactivity from parents and teachers. To corroborate the meaning behind convergence or divergence between parent and teacher reports, the

practitioner might also administer independent observational assessments of the child's behavior. These observational assessments might include having the child complete a set of structured and unstructured tasks in the home and school contexts. These observational assessments might resemble the contextually sensitive observational assessments developed to assess disruptive behavior among preschool children (Wakschlag et al., 2010). Let us assume that the data confirmed the practitioner's hypotheses: observations of greater parent-reported levels of hyperactivity, relative to teacher report. If so, the independent assessment would allow the practitioner to make an informed decision as to whether the parent and teacher reports varied for meaningful reasons. For instance, the independent assessment might have revealed the presence of higher levels of hyperactivity and less effective behavior management strategies at home relative to school. The key point to be taken from this example is that assessment procedures guiding clinical decision-making would greatly benefit from a priori predictions of patterns of convergence and divergence between informants' reports. These predictions can inform the gathering of independent qualitative and quantitative data about the contexts in which patients display concerns. In turn, these data and predictions may result in greater certainty when interpreting the outcomes of multi-informant assessments, resulting in increased reliability and validity in clinical decision-making.

Integrating the OTM with approaches from evidence-based medicine—

Developing personalized assessment procedures that leverage multiple informants' reports could be further facilitated by incorporating knowledge from other literatures focused on patient-centered models of clinical decision-making. Specifically, assessment methods within evidence-based medicine (EBM) often involve forming likelihood judgments for an assessment case, based on whether one or more pieces of information collected by an assessor support or fail to support a given conclusion about that patient's clinical presentation (for a review, see Youngstrom, 2013). As an example, consider a practitioner who collects clinical information about a patient who may evidence concerns with oppositional defiant disorder (ODD). The difference between EBM assessments and traditional psychological assessment procedures is that in EBM, one assigns specific values to data derived from information sources, based on whether these data support one outcome, relative to an alternative outcome.

For instance, to assess a patient for the presence versus absence of an ODD diagnosis, a practitioner might begin by consulting prior research on risk and associated features of ODD. Based on prior research, the practitioner would pose a clinical question to inform gathering clinical information about the patient, such as, *What are the base rates of inconsistent parenting practices among the parents of patients diagnosed with ODD and among non-patients?* The practitioner would administer to the parent a short survey assessment of inconsistent parenting. The practitioner would compare the outcomes of this survey against known base rates of inconsistent parenting. Essentially, the practitioner examines whether the parent's inconsistent parenting outcomes “matched” the outcomes identified in the literature for parents of patients diagnosed with ODD, relative to an alternative outcome (i.e., outcomes of inconsistent parenting for parents of non-patients). The resulting comparisons, referred to as diagnostic likelihood ratios (DLRs) are essentially

estimates of the ratio of true positives-to-false positives (e.g., ratio of the base rate of inconsistent parenting practices among the parents of ODD patients/ratio for parents of non-patients). These DLRs carry with them interpretative conventions similar to odds ratios (OR) calculated within logistic regression (see Straus, Glasziou, Richardson, & Haynes, 2011). That is, DLRs at or below 2 carry little value in terms of guiding clinical decisions, DLRs of approximately 5 carry relatively greater value, and DLRs of 10 or above may ensure sound confidence in making the particular clinical decision in question (e.g., ruling in an ODD diagnosis; see also Youngstrom, 2013). With this ODD example, if the parenting information about the patient yielded a high DLR, the practitioner might conclude that the patient is quite likely to meet diagnostic criteria for ODD and begin treatment for the condition. The practitioner may conduct additional assessments for the presence of commonly comorbid conditions (e.g., ADHD; see Drabick & Kendall, 2010).

Within an EBM assessment, multiple informants' reports might prove particularly useful for treatment planning. To continue with our previous example, perhaps DLRs for the presence of an ODD diagnosis were based primarily on useful, easy-to-collect information about inconsistent parenting. However, the information that informed initial diagnosis did not yield sensitive data on whether the patient evidences ODD concerns specific to particular contexts (e.g., home only vs. home and school). Yet, knowledge about the context(s) within which the patient evidences concerns would greatly facilitate treatment planning. In this sense, a multi-informant assessment consisting of behavioral checklists administered to the patient's parents and teachers could yield great utility in revealing context-specificity in displays of the patient's ODD concerns. Here, the clinical question might be, *What is the likelihood that the patient's concerns are pervasive across contexts or specific to the home context?* To address this question, the practitioner would administer survey reports to the patient's parent and teacher. Additionally, the practitioner would consult prior work to identify estimates of patients exhibiting disruptive behavior in interactions with parental and non-parental adults, perhaps based on data from the DB-DOS (e.g., De Los Reyes et al., 2009). Within these estimates, the practitioner would identify the base rates of patients exhibiting cross-contextual displays of disruptive behavior on the DB-DOS who also evidenced disruptive behavior across parent and teacher reports versus only parent report (i.e., 29% vs. 10%; see Table 5 of De Los Reyes et al., 2009). These estimates would translate to a DLR of approximately 2.9. Based on this DLR figure, if the patient in fact evidenced disruptive behavior concerns across parent and teacher reports, the data would point to the patient displaying concerns across contexts. Treatment would then progress based on the implementation of techniques at both home and school.

One additional issue warrants comment. In this EBM example, the only data collected on behalf of the patient were the parent and teacher reports. That is, the practitioner relied on parent and teacher report data to make clinical decisions (i.e., determine whether patient displayed concerns across home and school contexts). However, the practitioner turned to independent behavioral assessment data collected in prior published work to construct the DLR that was used to interpret the parent and teacher reports (De Los Reyes et al., 2009). The value in this approach is that the practitioner consulted prior published work to *interpret* the survey reports they collected from the patient's parent and teacher, and gauge whether

the reports contained information that should guide decision-making. This process is quite similar to current practices in interpreting clinic data, such as when a mental health professional consults published clinical cutoff values for a symptom scale to determine whether a patient evidences clinically relevant mental health concerns. Thus, this illustration highlights that methods taken from EBM hold promise for improving the use and interpretation of multi-informant assessments.

Research and Theoretical Implications

Addressing methodological issues raised by incremental validity research—

This review has important research and theoretical implications. Specifically, future multi-informant assessment research would benefit greatly from addressing the methodological confounds we discussed previously with regard to prior incremental validity research. We have three recommendations for addressing particularly pressing methodological issues. First, future incremental validity studies should involve use of measures completed by multiple informants that hold item content, response labeling, and scaling constant, in order to rule out methodological differences among reports when drawing inferences about incremental validity. Relatedly, as in research on preschool disruptive behavior (i.e., DB-DOS; Wakschlag et al., 2010), future research should focus on developing measures that can assess contextual variations in patients' mental health concerns, independently from the informants completing reports about patients. Using these measures, researchers can gauge, for instance, whether findings supporting the incremental validity of parent but not teacher reports were driven by the propensity of patients in the sample to display mental health concerns only observable by parents (Figure 2a).

Second, a benchmark or standard does not exist for identifying an incrementally valid informant for assessing particular mental health concerns for particular patient groups. That is, mental health researchers have developed a number of guidelines for making evidence-based decisions, most notably with regard to identifying efficacious treatments (for a review, see Roth & Fonagy, 2005). However, there do not exist specific criteria for determining how much incremental validity (e.g., predictive power, represented in effect size metrics) is sufficient to conclude that a given informant's report should be included in an assessment, let alone whether multiple informants' reports ought to be collected versus only a single report. This is an important consideration, in that one tends to observe small effects for the incremental validity of a given informant's report in relation to other informants' reports (Hunsley & Meyer, 2003).

Third, incremental validity research would benefit from basic efficacy studies of whether informants can provide incrementally valid reports when predicting validity criterion measures constructed blind or independent of all of the informants' reports. This work would focus on asking the question, *All things being equal, are scores taken from multi-informant assessments incrementally valid, relative to scores taken from any one informant?* However, like treatment efficacy research, this type of incremental validity research might involve use of validity criterion measures that do not reflect clinical indices or decisions made in “real-world” settings. Thus, in tandem with basic efficacy studies, we recommend a focus on incremental validity effectiveness research. This form of incremental validity work would

focus on asking the question, *Are scores taken from multi-informant assessments incrementally valid in predicting “real-world” clinically meaningful criterion measures, relative to scores taken from any one informant?* Essentially, these incremental validity studies would share many elements with prior incremental validity work, most notably the focus on such validity criterion measures as consensus diagnoses and treatment response. However, the difference between prior research and the efficacy and effectiveness variants of incremental validity research we propose would be in systematically ruling out the effects of criterion contamination on the outcomes of incremental validity studies.

Using basic mental health research to inform incremental validity research—

Much of our discussion of the multi-informant approach makes a fundamental assumption that mental health concerns manifest such that patients display signs of these concerns within some contexts and not others (Figures 2a and 2b) or across contexts (Figure 2c). The importance of understanding these contextual variations may be relevant to some (e.g., social anxiety, attention and hyperactivity), but not all, mental health concerns. Indeed, displays of some concerns may rarely vary across contexts (e.g., schizophrenia).

Our findings highlight the importance of basic research on contextual variations in mental health for informing the use and interpretation of multiple informants' reports, particularly with regard to incremental validity. Indeed, we discussed previously how lack of knowledge on contextual variations in mental health concerns may hinder our ability to draw conclusions on the generalizability of findings from incremental validity studies. That is, the composition of contextual variations in mental health concerns in a sample may often dictate the extent to which any one informant provides incrementally valid reports (Figure 2). Consequently, future research must link our basic understanding of contextual variations in patients' mental health concerns with understanding and interpreting findings in incremental validity research.

We previously described observational measures used in multi-informant research on preschool disruptive behavior that have revealed individual differences among patients in the specific contexts in which they displayed disruptive behavior concerns (De Los Reyes et al., 2009). We recommend that future incremental validity studies implement independent, contextually sensitive measures such as these to estimate contextual variations in patients' concerns. Using these measures, researchers may identify, for example, whether the sample was predominantly represented by patients who displayed mental health concerns exclusively within the home context (Figure 2a). In assessments of preschoolers' disruptive behavior, such a patient sample would be likely to result in parent reports evidencing incrementally valid reports to a far greater extent than other informants' reports, such as teachers. Yet, we previously discussed research indicating that preschoolers may evidence mental health concerns within and across a variety of contexts (e.g., Dirks et al., 2012). Thus, samples of patients displaying heterogeneity in mental health concerns within and across contexts may yield distinct estimates of incremental validity for parent and teacher reports (see Figures 2a and 2b). Consequently, the availability of contextually sensitive measures independent of the informants providing reports about patients would provide researchers with important data to rule out important methodological confounds.

Few measures exist for assessing contextually sensitive variations in patients' mental health concerns in a way that is independent from the informants providing mental health reports (e.g., using trained laboratory observers or clinical evaluators). This lack of contextually sensitive behavioral assessments may pose challenges to future research. To this end, we have two recommendations on the development and testing of these independent measures. First, previous work illustrates measures for assessing contextual variations in patients' reactions to authority figures (e.g., DB-DOS; see De Los Reyes et al., 2009). Importantly, a great deal of work points to the relevance of assessing patients' interactions with authority figures such as parents and teachers for a number of highly prevalent mental health concerns, such as attention and hyperactivity, depressive mood, and disruptive behavior (APA, 2013). This indicates the potential for researchers to develop versions of the DB-DOS tailored to assessing contextual variations in concerns displayed by patients of varying clinical presentations. Further, deficits in communicating and interacting with others in social settings (i.e., *social competence deficits*) characterize multiple mental health domains assessed and diagnosed among patients across the lifespan, including attention and hyperactivity, schizophrenia, and social anxiety (APA, 2013). Along these lines, recent work with adult social anxiety patients indicates that the administration of laboratory tasks across varying social interaction contexts (e.g., one-on-one social interactions and public speaking) allow for the detection of patients who display social competence deficits across social interactions or only within specific interactions (Beidel, Rao, Scharfstein, Wong, & Alfano, 2010; De Los Reyes, Bunnell et al., 2013). Thus, we encourage future work on whether independent assessments tailored to assess contextual variations in social competence deficits improve interpretability of multiple informants' reports of patients of varying developmental periods and clinical presentations.

Second, much of our discussion about using contextually sensitive independent measures to test the validity of the multi-informant approach involved use of structured behavioral observations of patients' behavior. This limited view of independent assessments extends from the fact that tests reviewed previously of the construct validity of the multi-informant approach relied on these or similar approaches (De Los Reyes et al., 2009; Hartley et al., 2011). However, other modalities may prove as useful, if not more useful, than behavioral observations. For instance, rather than take behavioral ratings of patients' performance within clinically relevant tasks, one could assess patients' physiology using wireless, ambulatory devices (e.g., heart rate monitors; for a review, see De Los Reyes & Aldao, 2015). The principles underlying this approach would function quite similarly to that of previous examples. Specifically, one might assess changes in physiological arousal across different clinically meaningful contexts to test whether these contextual changes in physiology “match” the differences observed between multiple informants' reports. For example, one might assess a patient's physiological arousal when completing a frustrating task indicative of the home context (e.g., discussing a topic that tends to elicit conflict between parent and patient) and also of the school context (e.g., completing a difficult mathematics problem in a noisy room). Here, one might test whether patients who display relatively high levels of arousal during the “math problem” task and relatively low levels during the “conflict discussion” task also tend to be the patients whose teachers report mental health concerns that go uncorroborated by reports completed by parents.

Venturing outside use of tasks administered in traditional clinic or research settings, there exist a variety of modalities with great potential for testing the validity of the multi-informant approach. For instance, mental health professionals often seek to integrate multi-informant assessments of mental health with such modalities as estimates of environmental versus genetic contributions to mental health (Baker, Jacobson, Raine, Lozano, & Bezdjian, 2007; Bartels, Boomsma, Hudziak, van Beijsterveldt, & van den Oord, 2007; Derks, Hudziak, Van Beijsterveldt, Dolan, & Boomsma, 2006), neurobiological assessments (e.g., electroencephalography; Curtis & Cicchetti, 2007), salivary assays of the cortisol awakening response (Bitsika, Sharpley, Sweeney, & McFarlane, 2014), social network analyses (Neal, Cappella, Wagner, & Atkins, 2011), geographical assessments of neighborhood risks (e.g., high-crime and economically disadvantaged areas; Odgers, Caspi, Bates, Sampson, & Moffitt, 2012), and performance-based neuropsychological tests (e.g., tests of executive functioning; Silver, 2014). Future research might involve constructing batteries consisting of multiple independent assessments that vary in their propensity to reflect one informant's report to a greater extent than other informants' reports. For example, a research team interested in assessing externalizing concerns (e.g., disruptive behavior disorders and ADHD) might predict that geographic assessments of neighborhood disadvantage near the vicinity of patients' homes reflect parents' reports about patients' concerns to a greater extent than teachers' reports about patients. Conversely, performance-based assessments of patients' executive functioning may include use of structured tasks akin to school activities (e.g., completing in-class assignments), and thus these assessments may reflect teachers' reports about patients to a greater extent than parents' reports about patients. Armed with this battery, the research team could examine whether the discrepancies between parent and teacher reports represent ecologically valid reflections of contextual variations in patients' mental health concerns. For example, instances in which parent reports indicate disruptive behavior concerns that go uncorroborated based on teacher reports may reflect patients who evidence both a relatively high likelihood of living in an economically disadvantaged neighborhood and scores in the normative range on standardized tests of executive functioning. In sum, multi-modal independent assessments may result in a richer understanding of whether mental health professionals can use the outcomes from multi-informant assessments to draw inferences about the context(s) in which patients display mental health concerns.

Expanding construct validity research on multi-informant assessments to additional clinical domains and developmental periods—Research on the construct validity of multi-informant assessments has yielded some promising findings. Yet, the findings are limited in their relevance to assessments focused on a circumscribed set of informants, mental health domains, and patient developmental periods. Clearly, the mental health domains for which patients' concerns may contextually vary far outnumber the domains within which the available evidence supports the construct validity of the multi-informant approach. Thus, we recommend that in the child literature, future research should focus on testing the construct validity of multi-informant assessments: (a) of children's mood and anxiety concerns, (b) that incorporate informants other than parent and teacher reports (e.g., clinician, patient, peer reports), and (c) that focus on developmental periods other than preschool and early adolescence.

Importantly, mental health researchers have made inroads in recent years on understanding multi-informant approaches to assessing adult mental health (e.g., Oltmanns & Turkheimer, 2009; van der Ende, Verhulst, & Tiemeier, 2012). When assessing adult mental health, self-reports and clinician ratings may be the most often used sources, yet recently researchers also have focused on collecting reports from collateral informants such as spouses or adult caregivers, as well as trained raters (e.g., clinical interviewers and laboratory observers; Achenbach, 2006). Further, meta-analytic reviews of adult mental health assessments have revealed correspondence levels at low-to-moderate magnitudes, similar to the correspondence levels observed in assessments of children (Achenbach, Krukowski, Dumenci, & Ivanova, 2005).

As with clinical child assessments, commonly used informants in adult mental health assessments may vary in their unique perspectives on adult patients' mental health concerns. For example, in multi-informant assessments using patients' self-reports and clinicians' reports about patients, reports might also contextually vary. That is, patients may base their reports on how they behave in work and home contexts; clinicians may take this information into account, but their reports also may reflect their trained observations of patients' behaviors in clinic settings (Achenbach et al., 2005; Groth-Marnat, 2009). Therefore, when assessing adult patients, informants selected to provide reports may differ in their opportunities for observing concerns.

To this end, there exists preliminary support for the incremental and construct validity of multi-informant approaches to assessing adult mental health. For example, in the assessment of personality pathology, collateral informants' reports may yield incremental information over-and-above patients' self-reports (e.g., Galione & Oltmanns, 2013; Miller, Pilkonis, & Clifton, 2005; Ready, Watson, & Clark, 2002). Further, one study supports the construct validity of multi-informant approaches to assessing adult social anxiety. Specifically, in this study patients completed a series of social interaction tasks from which independent raters assessed patients' social competence deficits (i.e., one-on-one interactions and a public speaking task; De Los Reyes, Bunnell et al., 2013). Patients and clinicians also completed reports about patients' internalizing symptoms. Consistent with previous work with children (De Los Reyes et al., 2009; Hartley et al., 2011), patients and clinicians were more likely to agree than disagree that the patient experienced moderate-to-high internalizing symptoms when patients exhibited social competence deficits consistently (i.e., rather than inconsistently) across social tasks.

Despite promising findings about the use of multi-informant approaches in adult mental health assessments, this area of work is far under-developed relative to multi-informant approaches to clinical child assessments. Indeed, the quantitative review conducted by Achenbach and colleagues (2005) evaluated a sample of 108 studies that provided sufficient information to estimate levels of cross-informant correspondence, accounting for 0.2% of the 51,000 articles identified for the review. Additionally, little work exists to estimate the consistency of low-to-moderate levels of correspondence among different developmental periods in adulthood (e.g., emerging adulthood and mid- to late-adulthood; for an exception, see van der Ende et al., 2012). Thus, future research should focus on construct validity as it relates to (a) assessments of other concerns such as mood, (b) use of collateral informants

such as spouses and adult caregivers (i.e., in the case of elderly patients), and (c) specific developmental periods.

Concluding Comments

We provided a conceptual overview of the multi-informant assessment approach, and synthesized cross-informant correspondence, incremental validity, and construct validity research literatures to assess whether the extant data support taking this approach to assessing children's mental health. Findings from the last 25 years of research on cross-informant correspondence in reports of children's mental health (Table 1) essentially mirror findings from the preceding 25 years (Achenbach et al., 1987). These robust findings across 50 years of research have allowed us to develop theoretical principles for interpreting the outcomes of multi-informant assessments (De Los Reyes, Thomas et al., 2013). We used these theoretical principles to critically evaluate research on the incremental and construct validity of the multi-informant assessment approach. In so doing, we identified a series of limitations in the research designs of previous studies testing the incremental and construct validity of multi-informant approaches to assessing children's mental health. In turn, we must limit our conclusions to stating that the best evidence in support of the validity of the multi-informant approach exists largely for assessments of observable mental health concerns (e.g., aggressive and disruptive behavior). The issues we raised in our review merit further study in patients exhibiting varied mental health concerns (e.g., anxiety, attention problems and hyperactivity, depressive symptoms) and across a wider range of developmental periods (e.g., childhood, adolescence, emerging adulthood, older adulthood). Future research on multi-informant assessments may inform the development of contextually sensitive assessment paradigms that mental health professionals can use to personalize assessment techniques and interventions to fit patients' unique needs. Identifying when multi-informant assessments reveal information about the specific contexts in which patients display mental health concerns may improve how mental health professionals use this information to make clinical decisions, provide patient care, and evaluate the outcomes of such care.

Acknowledgments

This work was partially supported by a pre-doctoral National Research Service Award to Sarah A. Thomas from the National Institute on Drug Abuse (F31-DA033913).

References

- Achenbach TM. Definitely more than measurement error: But how should we understand and deal with informant discrepancies? *Journal of Clinical Child and Adolescent Psychology*. 2011; 40:80–86. [10.1080/15374416.2011.533416](https://doi.org/10.1080/15374416.2011.533416) [PubMed: 21229445]
- Achenbach TM. As others see us: Clinical and research implications of cross-informant correlations for psychopathology. *Current Directions in Psychological Science*. 2006; 15:94–98. [10.1111/j.0963-7214.2006.00414.x](https://doi.org/10.1111/j.0963-7214.2006.00414.x)
- Achenbach TM, Krukowski RA, Dumenci L, Ivanova MY. Assessment of adult psychopathology: Meta-analysis and implications of cross-informant correlations. *Psychological Bulletin*. 2005; 131:361–382. [10.1037/0033-2909.131.3.361](https://doi.org/10.1037/0033-2909.131.3.361) [PubMed: 15869333]

- Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*. 1987; 101:213–232.10.1037/0033-2909.101.2.213 [PubMed: 3562706]
- Achenbach, TM.; Rescorla, L. Manual for the ASEBA school-age forms and profiles. Burlington, VT: University of Vermont Research Center for Children, Youth, and Families; 2001.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders-text revision. 4th. Washington, DC: Author; 2000. DSM-IV-TR
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th. Washington, DC: Author; 2013. DSM-V
- Anderson ER, Hope DA. The relationship among social phobia, objective and perceived physiological reactivity, and anxiety sensitivity in an adolescent population. *Journal of Anxiety Disorders*. 2009; 23:18–26.10.1016/j.janxdis.2008.03.011 [PubMed: 18436426]
- Angold A, Costello J. The relative diagnostic utility of child and parent reports of oppositional defiant behaviors. *International Journal of Methods in Psychiatric Research*. 1996; 6:253–259.10.1002/(SICI)1234-988X(199612)6:4<253::AID-MPR170>3.3.CO;2-O
- Baldwin JS, Dadds MR. Reliability and validity of parent and child versions of the Multidimensional Anxiety Scale for Children in community samples. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2007; 46:252–260.10.1097/01.chi.0000246065.93200.a1 [PubMed: 17242629]
- Baker L, Jacobson K, Raine A, Lozano D, Bezdjian S. Genetic and environmental bases of childhood antisocial behavior: a multi-informant twin study. *Journal of Abnormal Psychology*. 2007; 116:219–235.10.1037/0021-843X.116.2.219 [PubMed: 17516756]
- Bartels M, Boomsma D, Hudziak J, van Beijsterveldt T, van den Oord E. Twins and the study of rater (dis)agreement. *Psychological Methods*. 2007; 12:451–466.10.1037/1082-989X.12.4.451 [PubMed: 18179355]
- Beidel DC, Rao PA, Scharfstein L, Wong N, Alfano CA. Social skills and social phobia: An investigation of DSM-IV subtypes. *Behaviour Research and Therapy*. 2010; 48:992–1001.10.1016/j.brat.2010.06.005 [PubMed: 20637452]
- Beidel DC, Turner SM, Young BJ, Ammerman RT, Sallee FR, Crosby L. Psychopathology of adolescent social phobia. *Journal of Psychopathology and Behavioral Assessment*. 2007; 29(1): 47–54.10.1007/s10862-006-9021-1
- Bird HR, Gould MS, Staghezza B. Aggregating data from multiple informants in child psychiatry epidemiological research. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1992; 31:78–85.10.1097/00004583-199201000-00012 [PubMed: 1537785]
- Bitsika V, Sharpley CF, Sweeney JA, McFarlane JR. HPA and SAM axis responses as correlates of self- vs parental ratings of anxiety in boys with an autistic disorder. *Physiology and Behavior*. 2014; 127:1–7.10.1016/j.physbeh.2013.12.011 [PubMed: 24412722]
- Blais MA, Hilsenroth MJ, Castlebury F, Fowler JC, Baity MR. Predicting DSM-IV Cluster B personality disorder criteria from MMPI-2 and Rorschach data: a test of incremental validity. *Journal of Personality Assessment*. 2001; 76:150–168.10.1207/S15327752JPA7601_9 [PubMed: 11206295]
- Bögels SM, Alden L, Beidel DC, Clark LA, Pine DS, Stein MB, Voncken M. Social anxiety disorder: Questions and answers for the DSM-V. *Depression and Anxiety*. 2010; 27:168–189.10.1002/da.20670 [PubMed: 20143427]
- Borenstein, M.; Hedges, LV.; Higgins, JPT.; Rothstein, HR. Introduction to meta-analysis. Chichester: Wiley; 2009.
- Borsboom D, Mellenbergh GJ, van Heerden J. The concept of validity. *Psychological Review*. 2004; 111:1061–1071.10.1037/0033-295X.111.4.1061 [PubMed: 15482073]
- Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Lijmer JG. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clinical Chemistry*. 2003; 49:7–18.10.7326/0003-4819-138-1-200301070-00012-w1 [PubMed: 12507954]

- Brookman-Frazee L, Haine RA, Gabayan EN, Garland AF. Predicting frequency of treatment visits in community-based youth psychotherapy. *Psychological Services*. 2008; 5:126–138.10.1037/1541-1559.5.2.126 [PubMed: 20396643]
- Brown-Jacobsen AM, Wallace DP, Whiteside SPH. Multimethod, multi-informant agreement, and positive predictive value in the identification of child anxiety disorders using the SCAS and ADIS-C. *Assessment*. 2011; 18:382–392.10.1177/1073191110375792 [PubMed: 20644080]
- Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review*. 2006; 26:17–31.10.1016/j.cpr.2005.07.003 [PubMed: 16199119]
- Campbell DT, Fiske DW. Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin*. 1959; 56:81–105.10.1037/h0046016 [PubMed: 13634291]
- Carey WB. Temperament and behavior problems in the classroom. *School Psychology Review*. 1998; 27:522–533. Stable URL: <http://psycnet.apa.org/psycinfo/1999-00154-005>.
- Carlson GA, Blader JC. Diagnostic implications of informant disagreement for manic symptoms. *Journal of Child and Adolescent Psychopharmacology*. 2011; 21:399–405.10.1089/cap.2011.0007 [PubMed: 22040185]
- Carlson GA, Youngstrom EA. Clinical implications of pervasive manic symptoms in children. *Biological Psychiatry*. 2003; 53:1050–1058.10.1016/S0006-3223(03)00068-4 [PubMed: 12788250]
- Cicchetti D. The emergence of developmental psychopathology. *Child Development*. 1984; 55:1–7.10.2307/1129830 [PubMed: 6705613]
- Cohen, J.; Cohen, P.; West, S.; Aiken, L. *Applied multiple regression/correlation analysis for the behavioral sciences*. 3rd. Hillsdale, NJ: Erlbaum; 2003.
- Comer JS, Kendall PC. A symptom-level examination of parent-child agreement in the diagnosis of anxious youth. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2004; 43:878–886.10.1097/01.chi.0000125092.35109.c5 [PubMed: 15213589]
- Conrad M, Hammen C. Role of maternal depression in perceptions of child maladjustment. *Journal of Consulting and Clinical Psychology*. 1989; 57:663–667.10.1037/0022-006X.57.5.663 [PubMed: 2794188]
- Crick, NR.; Werner, NE.; Casas, JF.; O'Brien, KM.; Nelson, DA.; Grotjeter, JK.; Markon, K. Childhood aggression and gender: A new look at an old problem. In: Bernstein, D., editor. *Nebraska symposium on motivation*. Lincoln, NE: University of Nebraska Press; 1998. p. 75-141.
- Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine*. 2003; 78:775–780.10.1097/00001888-200308000-00003 [PubMed: 12915363]
- Cuijpers P, Dekker J. Psychologische behandeling van depressie; een systematisch overzicht van meta-analysen [Psychological treatment for depression: A systematic review of meta-analyses]. *Nederlands Tijdschrift voor Geneeskunde*. 2005; 149:1892–1897. Retrieved from <http://hdl.handle.net/1871/17373>. [PubMed: 16136742]
- Curtis WJ, Cicchetti D. Emotion and resilience: A multilevel investigation of hemispheric electroencephalogram asymmetry and emotion regulation in maltreated and nonmaltreated children. *Development and Psychopathology*. 2007; 19:811–840.10.1017/S0954579407000405 [PubMed: 17705904]
- Dadds MR, Perrin S, Yule W. Social desirability and self-reported anxiety in children: An analysis of the RCMAS Lie Scale. *Journal of Abnormal Child Psychology*. 1998; 26:311–317.10.1023/A:1022610702439 [PubMed: 9700522]
- Dawes RM. Two methods for studying the incremental validity of a Rorschach variable. *Psychological Assessment*. 1999; 11:297–302.10.1037/1040-3590.11.3.297
- De Los Reyes A. Introduction to the special section. More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child and Adolescent Psychology*. 2011; 40:1–9.10.1080/15374416.2011.533405 [PubMed: 21229439]

- De Los Reyes A. Strategic objectives for improving understanding of informant discrepancies in developmental psychopathology research. *Development and Psychopathology*. 2013; 25:669–682.10.1017/S0954579413000096 [PubMed: 23880384]
- De Los Reyes A, Aldao A. Introduction to the special section. Toward implementing physiological measures in clinical child and adolescent assessments. *Journal of Clinical Child and Adolescent Psychology*. 2015; 44 Advance online publication. 10.1080/15374416.2014.891227
- De Los Reyes A, Aldao A, Thomas SA, Daruwala SE, Swan AJ, Van Wie M, et al. Lechner WV. Adolescent self-reports of social anxiety: Can they disagree with objective psychophysiological measures and still be valid? *Journal of Psychopathology and Behavioral Assessment*. 2012; 34:308–322.10.1007/s10862-012-9289-2
- De Los Reyes A, Alfano CA, Beidel DC. The relations among measurements of informant discrepancies within a multisite trial of treatments for childhood social phobia. *Journal of Abnormal Child Psychology*. 2010; 38:395–404.10.1007/s10802-009-9373-6 [PubMed: 20013046]
- De Los Reyes A, Alfano CA, Beidel DC. Are clinicians' assessments of improvements in children's functioning "global"? *Journal of Clinical Child and Adolescent Psychology*. 2011; 40:281–294.10.1080/15374416.2011.546043 [PubMed: 21391024]
- De Los Reyes A, Augenstein TM, Aldao A, Thomas SA, Daruwala SE, Kline K, Regan T. Implementing psychophysiology in clinical assessments of adolescent social anxiety: Use of rater judgments based on graphical representations of psychophysiology. *Journal of Clinical Child and Adolescent Psychology*. 2015; 44 Advance online publication. 10.1080/15374416.2013.859080
- De Los Reyes A, Bunnell BE, Beidel DC. Informant discrepancies in adult social anxiety disorder assessments: Links with contextual variations in observed behavior. *Journal of Abnormal Psychology*. 2013; 122:376–386.10.1037/a0031150 [PubMed: 23421526]
- De Los Reyes A, Ehrlich KB, Swan AJ, Luo T, Van Wie M, Pabón SC. An experimental test of whether informants can report about child and family behavior based on settings of behavioral expression. *Journal of Child and Family Studies*. 2013; 22:177–191.10.1007/s10826-012-9567-3
- De Los Reyes A, Goodman KL, Kliewer W, Reid-Quiñones KR. Whose depression relates to discrepancies? Testing relations between informant characteristics and informant discrepancies from both informants' perspectives. *Psychological Assessment*. 2008; 20:139–149.10.1037/1040-3590.20.2.139 [PubMed: 18557691]
- De Los Reyes A, Goodman KL, Kliewer W, Reid-Quiñones KR. The longitudinal consistency of mother-child reporting discrepancies of parental monitoring and their ability to predict child delinquent behaviors two years later. *Journal of Youth and Adolescence*. 2010; 39:1417–1430.10.1007/s10964-009-9496-7 [PubMed: 20020188]
- De Los Reyes A, Henry DB, Tolan PH, Wakschlag LS. Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology*. 2009; 37:637–652.10.1007/s10802-009-9307-3 [PubMed: 19247829]
- De Los Reyes A, Kazdin AE. Measuring informant discrepancies in clinical child research. *Psychological Assessment*. 2004; 16:330–334.10.1037/1040-3590.16.3.330 [PubMed: 15456389]
- De Los Reyes A, Kazdin AE. Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*. 2005; 131:483–509.10.1037/0033-2909.131.4.483 [PubMed: 16060799]
- De Los Reyes A, Kazdin AE. Conceptualizing changes in behavior in intervention research: The range of possible changes model. *Psychological Review*. 2006; 113:554–583.10.1037/0033-295X.113.3.554 [PubMed: 16802881]
- De Los Reyes A, Kundey SMA, Wang M. The end of the primary outcome measure: A research agenda for constructing its replacement. *Clinical Psychology Review*. 2011; 31:829–838.10.1016/j.cpr.2011.03.011 [PubMed: 21545781]
- De Los Reyes A, Marsh JK. Patients' contexts and their effects on clinicians' impressions of conduct disorder symptoms. *Journal of Clinical Child and Adolescent Psychology*. 2011; 40:479–485.10.1080/15374416.2011.563471 [PubMed: 21534058]

- De Los Reyes A, Salas S, Menzer MM, Daruwala SE. Criterion validity of interpreting scores from multi-informant statistical interactions as measures of informant discrepancies in psychological assessments of children and adolescents. *Psychological Assessment*. 2013; 25:509–519.10.1037/a0032081 [PubMed: 23544400]
- De Los Reyes A, Thomas SA, Goodman KL, Kundery SMA. Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology*. 2013; 9:123–149.10.1146/annurev-clinpsy-050212-185617
- De Los Reyes A, Youngstrom EA, Pabón SC, Youngstrom JK, Feeny NC, Findling RL. Internal consistency and associated characteristics of informant discrepancies in clinic referred youths age 11 to 17 years. *Journal of Clinical Child and Adolescent Psychology*. 2011a; 40:36–53.10.1080/15374416.2011.533402 [PubMed: 21229442]
- De Los Reyes A, Youngstrom EA, Swan AJ, Youngstrom JK, Feeny NC, Findling RL. Informant discrepancies in clinical reports of youths and interviewers' impressions of the reliability of informants. *Journal of Child and Adolescent Psychopharmacology*. 2011b; 21:417–424.10.1089/cap.2011.0011 [PubMed: 22040187]
- Derks E, Hudziak J, Van Beijsterveldt C, Dolan C, Boomsma D. Genetic analyses of maternal and teacher ratings on attention problems in 7-year-old Dutch twins. *Behavior Genetics*. 2006; 36:833–844.10.1007/s10519-006-9084-5 [PubMed: 16773450]
- DiBartolo PM, Albano AM, Barlow DH, Heimberg RG. Cross-informant agreement in the assessment of social phobia in youth. *Journal of Abnormal Child Psychology*. 1998; 26:213–220.10.1023/A:1022624318795 [PubMed: 9650627]
- Dirks MA, Boyle MH, Georgiades K. Psychological symptoms in youth and later socioeconomic functioning: Do associations vary by informant? *Journal of Clinical Child and Adolescent Psychology*. 2011; 40:10–22.10.1080/15374416.2011.533403 [PubMed: 21229440]
- Dirks MA, De Los Reyes A, Briggs-Gowan MJ, Cella D, Wakschlag LS. Embracing not erasing contextual variability in children's behavior – theory and utility in the selection and use of methods and informants in developmental psychopathology. *Journal of Child Psychology and Psychiatry*. 2012; 53:558–574.10.1111/j.1469-7610.2012.02537.x [PubMed: 22360546]
- Drabick DAG, Gadow KD, Loney J. Source-specific oppositional defiant disorder: Comorbidity and risk factors in referred elementary schoolboys. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2007; 46:92–101.10.1097/01.chi.0000242245.00174.90 [PubMed: 17195734]
- Drabick DAG, Gadow KD, Loney J. Co-occurring ODD and GAD symptom groups: Source-specific syndromes and cross-informant comorbidity. *Journal of Clinical Child and Adolescent Psychology*. 2008; 37:314–326.10.1080/15374410801955862 [PubMed: 18470769]
- Drabick DAG, Kendall PC. Developmental psychopathology and the diagnosis of mental health problems among youth. *Clinical Psychology: Science and Practice*. 2010; 17:272–280.10.1111/j.1468-2850.2010.01219.x [PubMed: 21243110]
- Duhig AM, Renk K, Epstein MK, Phares V. Interparental agreement on internalizing, externalizing, and total behavior problems: A meta-analysis. *Clinical Psychology: Science and Practice*. 2000; 7:435–453.10.1093/clipsy.7.4.435
- Frank SJ, Van Egeren LA, Fortier JL, Chase P. Structural, relative, and absolute agreement between parents' and adolescent inpatients' reports of adolescent functional impairment. *Journal of Abnormal Child Psychology*. 2000; 28:395–402.10.1023/A:1005125211187 [PubMed: 10949963]
- Galione JN, Oltmanns TF. Identifying personality pathology associated with major depressive episodes: Incremental validity of informant reports. *Journal of Personality Assessment*. 2013; 95:625–632.10.1080/00223891.2013.825624 [PubMed: 24004355]
- Garb, HN. *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association; 1998.
- Garb HN. Incremental validity and the assessment of psychopathology in adults. *Psychological Assessment*. 2003; 15:508–520.10.1037/1040-3590.11.3.297 [PubMed: 14692846]
- Garner WR, Hake HW, Eriksen CW. Operationism and the concept of perception. *Psychological Review*. 1956; 63:149–159.10.1037/h0042992 [PubMed: 13323170]

- Goodman KL, De Los Reyes A, Bradshaw CP. Understanding and using informants' reporting discrepancies of youth victimization: A conceptual model and recommendations for research. *Clinical Child and Family Psychology Review*. 2010; 13:366–383.10.1007/s10567-010-0076-x [PubMed: 20799062]
- Goodman SH, Gotlib IH. Risk for psychopathology in the children of depressed mothers: A developmental model for understanding mechanisms of transmission. *Psychological Review*. 1999; 106:458–490.10.1037/0033-295X.106.3.458 [PubMed: 10467895]
- Grills AE, Ollendick TH. Multiple informant agreement and the Anxiety Disorders Interview Schedule for Parents and Children. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2003; 42:30–40.10.1097/00004583-200301000-00008 [PubMed: 12500074]
- Groth-Marnat, G. *Handbook of psychological assessment*. 5th. Hoboken, NJ: John Wiley & Sons; 2009.
- Hartley AG, Zakriski AL, Wright JC. Probing the depths of informant discrepancies: Contextual influences on divergence and convergence. *Journal of Clinical Child and Adolescent Psychology*. 2011; 40:54–66.10.1080/15374416.2011.533404 [PubMed: 21229443]
- Hawley KM, Weisz JR. Child, parent, and therapist (dis)agreement on target problems in outpatient therapy: The therapist's dilemma and its implications. *Journal of Consulting and Clinical Psychology*. 2003; 71:62–70.10.1037/0022-006X.71.1.62 [PubMed: 12602426]
- Haynes SN, Lench HC. Incremental validity of new clinical assessment measures. *Psychological Assessment*. 2003; 15(4):456–466.10.1037/1040-3590.15.4.456 [PubMed: 14692842]
- Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analysis. *British Medical Journal*. 2003; 327:557–560.10.1136/bmj.327.7414.557 [PubMed: 12958120]
- Hunsley J. Introduction to the special section on incremental validity and utility in clinical assessment. *Psychological Assessment*. 2003; 15:443–445.10.1037/1040-3590.15.4.443 [PubMed: 14692840]
- Hunsley, J.; Lee, CM. *Introduction to clinical psychology*. Hoboken: Wiley & Sons; 2010.
- Hunsley J, Mash EJ. Developing guidelines for the evidence-based assessment (EBA) of adult disorders. *Psychological Assessment*. 2005; 17 Special section. 10.1037/1040-3590.17.3.251
- Hunsley J, Mash EJ. Evidence-based assessment. *Annual Review of Clinical Psychology*. 2007; 3:29–51.10.1146/annurev.clinpsy.3.022806.091419
- Hunsley J, Meyer GJ. The incremental validity of psychological testing and assessment: Conceptual, methodological, and statistical issues. *Psychological Assessment*. 2003; 15:446–455.10.1037/1040-3590.15.4.446 [PubMed: 14692841]
- Jensen-Doss A, Weisz JR. Diagnostic agreement predicts treatment process and outcomes in youth mental health clinics. *Journal of Consulting and Clinical Psychology*. 2008; 76:711–722.10.1037/0022-006X.76.5.711 [PubMed: 18837589]
- Johnston C, Murray C. Incremental validity in the psychological assessment of children and adolescents. *Psychological Assessment*. 2003; 15:496–507.10.1037/1040-3590.15.4.496 [PubMed: 14692845]
- Jouriles EN, Thompson SM. Effects of mood on mothers' evaluations of children's behavior. *Journal of Family Psychology*. 1993; 6:300–307.10.1037/0893-3200.6.3.300
- Kazdin AE, Kagan J. Models of dysfunction in developmental psychopathology. *Clinical Psychology: Science and Practice*. 1994; 1:35–52.10.1111/j.1468-2850.1994.tb00005.x
- Klein DN, Dougherty LR, Olino TM. Toward guidelines for evidence-based assessment of depression in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*. 2005; 34:412–432.10.1207/s15374424jccp3403_3 [PubMed: 16026212]
- Kraemer HC, Measelle JR, Ablow JC, Essex MJ, Boyce WT, Kupfer DJ. A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry*. 2003; 160:1566–1577.10.1176/appi.ajp.160.9.1566 [PubMed: 12944328]
- Kramer TL, Phillips SD, Hargis MB, Miller TL, Burns BJ, Robbins JM. Disagreement between parent and adolescent reports of functional impairment. *Journal of Child Psychology and Psychiatry*. 2004; 45:248–259.10.1111/j.1469-7610.2004.00217.x [PubMed: 14982239]

- Kwon K, Kim EM, Sheridan SM. A contextual approach to social skills assessment in the peer group: Who is the best judge? *School Psychology Quarterly*. 2012; 27:121–133.10.1037/a0028696 [PubMed: 22731871]
- Ladd GW, Kochenderfer-Ladd B. Identifying victims of peer aggression from early to middle childhood: Analysis of cross-informant data for concordance, estimation of relational adjustment, prevalence of victimization, and characteristics of identified victims. *Psychological Assessment*. 2002; 14:74–96.10.1037//1040-3590.14.1.74 [PubMed: 11911051]
- Lapouse R, Monk MA. An epidemiologic study of behavior characteristics of children. *American Journal of Public Health*. 1958; 48:1134–1144.10.2105/AJPH.48.9.1134 [PubMed: 13571479]
- Lipsey MW, Wilson DB. The efficacy of psychological, educational, and behavioral treatment. *American Psychologist*. 1993; 48:1181–1209.10.1037/0003-066X.48.12.1181 [PubMed: 8297057]
- Loeber R, Green SM, Lahey BB. Mental health professionals' perception of the utility of children, mothers, and teachers as informants of childhood psychopathology. *Journal of Clinical Child Psychology*. 1990; 19:136–143.10.1207/s15374424jccp1902_5
- Loeber R, Green SM, Lahey BB, Stouthamer-Loeber M. Optimal informants on childhood disruptive behaviors. *Development and Psychopathology*. 1989; 1:317–337.10.1017/S09545794000050X
- Luthar SS, Cicchetti D, Becker B. The construct of resilience: A critical evaluation and guidelines for future work. *Child Development*. 2000; 71:543–562.10.1111/1467-8624.00164 [PubMed: 10953923]
- Markon KE, Chmielewski M, Miller CJ. The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin*. 2011; 137:856–879.10.1037/a0023678 [PubMed: 21574681]
- Marsh JK, De Los Reyes A, Wallerstein A. The influence of contextual information on lay judgments of childhood mental health concerns. *Psychological Assessment*. 2014 Advance online publication. 10.1037/pas0000012
- Mash, E.J., Hunsley, J., editors. *Journal of Clinical Child and Adolescent Psychology*. 2005. Evidence-based assessment of child and adolescent disorders: issues and challenges. Special section
- McMahon RJ, Frick PJ. Evidence-based assessment of conduct problems in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*. 2005; 34:477–505.10.1207/s15374424jccp3403_6 [PubMed: 16026215]
- Meyer GJ, Finn SE, Eyde LD, Kay GG, Moreland KL, Dies RR, et al. Reed GM. Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*. 2001; 56:128–165.10.1037/0003-066X.56.2.128 [PubMed: 11279806]
- Miller JD, Pilkonis PA, Clifton A. Self- and other-reports of traits from the Five Factor Model: Relations to Personality Disorder. *Journal of Personality Disorders*. 2005; 19:400–419.10.1521/pepi.2005.19.4.400 [PubMed: 16178682]
- Mischel W, Shoda Y. A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*. 1995; 102:246–268.10.1037/0033-295X.102.2.246 [PubMed: 7740090]
- Møller AP, Jennions MD. How much variance can be explained by ecologists and evolutionary biologists? *Oecologia*. 2002; 132:492–500.10.1007/s00442-002-0952-2
- National Institute of Mental Health. The National Institute of Mental Health strategic plan (NIH Publication No 08-6368). 2008. Retrieved from <http://www.nimh.nih.gov/about/strategic-planning-reports/index.shtml>
- Neal JW, Cappella E, Wagner C, Atkins MS. Seeing eye to eye: Predicting teacher-student agreement on classroom social networks. *Social Development*. 2011; 20:376–393.10.1111/j.1467-9507.2010.00582.x [PubMed: 21666768]
- Nimon K. Regression commonality analysis: Demonstration of an SPSS solution. *Multiple Linear Regression Viewpoints*. 2010; 36:10–17. Stable URL: http://mlrv.ua.edu/2010/vol36_1/nimon_proof_2.pdf.
- Ogden CL, Caspi A, Bates CJ, Sampson RJ, Moffitt TE. Systematic social observation of children's neighborhoods using Google Street View: a reliable and cost-effective method. *Journal of Child Psychology and Psychiatry*. 2012; 53:1009–1017.10.1111/j.1469-7610.2012.02565.x [PubMed: 22676812]

- Offord DR, Boyle MH, Racine Y, Szatmari P, Fleming JE, Sanford M, Lipman EL. Integrating assessment data from multiple informants. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1996; 35:1078–1085.10.1097/00004583-199409000-00008 [PubMed: 8755805]
- Oltmanns TF, Turkheimer E. Person perception and personality pathology. *Current Directions in Psychological Science*. 2009; 18:32–36.10.1111/j.1467-8721.2009.01601.x [PubMed: 20539833]
- Owens J, Hoza B. Diagnostic utility of DSM-IV-TR symptoms in the prediction of DSM-IV-TR ADHD subtypes and ODD. *Journal of Attention Disorders*. 2003; 7:11–27.10.1177/108705470300700102 [PubMed: 14738178]
- Pelham WE Jr, Fabiano GA, Massetti GM. Evidence-based assessment of Attention-Deficit Hyperactivity Disorder in children and adolescents. *Journal of Clinical Child & Adolescent Psychology*. 2005; 34:449–476.10.1207/s15374424jccp3403_5 [PubMed: 16026214]
- Peterson RA. On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research*. 2001; 28:450–461. Retrieved September 8, 2014 from <http://www.jstor.org/stable/10.1086/323732>.
- Phares V, Compas BE. Adolescents' subjective distress over their emotional/behavioral problems. *Journal of Consulting and Clinical Psychology*. 1990; 58:596–603.10.1037/0022-006X.58.5.596 [PubMed: 2254506]
- Phares V, Danforth JS. Adolescents', parents', and teachers' distress over adolescents' behavior. *Journal of Abnormal Child Psychology*. 1994; 22:721–732.10.1007/BF02171998 [PubMed: 7876459]
- Piacentini JC, Cohen P, Cohen J. Combining discrepant diagnostic information from multiple sources: Are complex algorithms better than simple ones? *Journal of Abnormal Child Psychology*. 1992; 20:51–63.10.1007/BF00927116 [PubMed: 1548394]
- Power TJ, Andrews TJ, Eiraldi RB, Doherty BJ, Ikeda MJ, DuPaul GJ, Landau S. Evaluating Attention Deficit Hyperactivity Disorder using multiple informants: The incremental utility of combining teacher with parent reports. *Psychological Assessment*. 1998; 10:250–260.10.1037/1040-3590.10.3.250
- Rapee RM, Barrett PM, Dadds MR, Evans L. Reliability of the *DSM-III-R* childhood anxiety disorders using structured interview: Interrater and parent–child agreement. *Journal of the American Academy of Child and Adolescent Psychiatry*. 1994; 33:984–992.10.1097/00004583-199409000-00008 [PubMed: 7961354]
- Ready RE, Watson D, Clark LA. Psychiatric patient and informant-reported personality: Predicting concurrent and future behavior. *Assessment*. 2002; 9:361–372.10.1177/1073191102238157 [PubMed: 12462756]
- Renk K, Phares V. Cross-informant ratings of social competence in children and adolescents. *Clinical Psychology Review*. 2004; 24:239–254.10.1016/j.cpr.2004.01.004 [PubMed: 15081518]
- Rescorla LA, Ginzburg S, Achenbach TM, Ivanova MY, Almqvist F, Begovac I, et al. Verhulst FC. Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. *Journal of Clinical Child & Adolescent Psychology*. 2013; 42:262–273.10.1080/15374416.2012.717870 [PubMed: 23009025]
- Rescorla LA, Bochiccio LS, Achenbach TM, Ivanova MY, Almqvist F, Begovac I, et al. Verhulst FC. Parent-teacher agreement on children's problems in 21 societies. *Journal of Clinical Child & Adolescent Psychology*. 2014; 43:627–642.10.1080/15374416.2014.900719 [PubMed: 24787452]
- Richters JE. Depressed mothers as informants about their children: A critical review of the evidence for distortion. *Psychological Bulletin*. 1992; 112:485–499.10.1037/0033-2909.112.3.485 [PubMed: 1438639]
- Rogosa D, Brandt D, Zimowski M. A growth curve approach to the measurement of change. *Psychological Bulletin*. 1982; 92:726–748.10.1037/0033-2909.92.3.726
- Rogosa DR, Willett JB. Demonstrating the reliability of the difference score in the measurement of change. *Journal of Educational Measurement*. 1983; 20:335–343.10.1111/j.1745-3984.1983.tb00211.x
- Roth, A.; Fonagy, P. What works for whom?: A critical review of psychotherapy research. 2nd. New York: Guilford Press; 2005.

- Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence-based medicine: What it is and what it isn't. *British Medical Journal*. 1996; 312:71–72.10.1136/bmj.312.7023.71 [PubMed: 8555924]
- Sanislow CA, Pine DS, Quinn KJ, Kozak MJ, Garvey MA, Heinssen RK, Wang PS, Cuthbert BN. Developing constructs for psychopathology research: Research domain criteria. *Journal of Abnormal Psychology*. 2010; 119:631–639.10.1037/a0020909 [PubMed: 20939653]
- Schmidt FL, Hunter JE. Measurement error in psychological research: lessons from 26 research scenarios. *Psychological Methods*. 1996; 1:199–223.10.1037/1082-989X.1.2.199
- Schwarz N. Self-reports: How the questions shape the answers. *American Psychologist*. 1999; 54:93–105.10.1037/0003-066X.54.2.93
- Silver CH. Sources of data about children's executive functioning: review and commentary. *Child Neuropsychology*. 2014; 20:1–13.10.1080/09297049.2012.727793 [PubMed: 23030631]
- Silverman WK, Ollendick TH. Evidence-based assessment of anxiety and its disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*. 2005; 34:380–411.10.1207/s15374424jccp3403_2 [PubMed: 16026211]
- Silverman WK, Rabian B. Test-retest reliability of the DSM-III-R childhood anxiety disorders symptoms using the Anxiety Disorders Interview Schedule for children. *Journal of Anxiety Disorders*. 1995; 9:139–150.10.1016/0887-6185(94)00032-8
- Smetana JG. "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*. 2008; 2:19–25.10.1111/j.1750-8606.2008.00036.x
- Smetana JG, Campione-Barr N, Metzger A. Adolescent development in interpersonal and societal contexts. *Annual Review of Psychology*. 2006; 57:255–284.10.1146/annurev.psych.57.102904.190124
- Smith BH, Pelham WE Jr, Gnagy E, Molina B, Evans S. The reliability, validity, and unique contributions of self-report by adolescents receiving treatment for Attention-Deficit/Hyperactivity Disorder. *Journal of Consulting and Clinical Psychology*. 2000; 68:489–499.10.1037//0022-006X.68.3.489 [PubMed: 10883565]
- Smith GT, Fischer S, Fister SM. Incremental validity principles in test construction. *Psychological Assessment*. 2003; 15:467–477.10.1037/1040-3590.15.4.467 [PubMed: 14692843]
- Spear LP. The adolescent brain and age-related behavioral manifestations. *Neuroscience and Behavioral Reviews*. 2000; 24:417–463.10.1016/S0149-7634(00)00014-2
- Straus, SE.; Glasziou, P.; Richardson, WS.; Haynes, RB. Evidence-based medicine: How to practice and teach EBM. 4th. New York, NY: Churchill Livingstone; 2011.
- Swets JA. The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist*. 1992; 47:522–532.10.1037/0003-066X.47.4.522 [PubMed: 1595983]
- Swets JA, Dawes RM, Monahan J. Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*. 2000; 1:1–26.10.1111/1529-1006.001
- Tamim RM, Bernard RM, Borokhovski E, Abrami PC, Schmid RF. What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. *Review of Educational Research*. 2011; 81:4–28.10.3102/0034654310393361
- Thomas SA, Aldao A, De Los Reyes A. Implementing clinically feasible psychophysiological measures in evidence-based assessments of adolescent social anxiety. *Professional Psychology: Research and Practice*. 2012; 43:510–519.10.1037/a0029183
- Thuppal M, Carlson GA, Sprafkin J, Gadow KD. Correspondence between adolescent report, parent report, and teacher report of manic symptoms. *Journal of Child and Adolescent Psychopharmacology*. 2002; 12:27–35.10.1089/10445460252943542 [PubMed: 12014592]
- Tulbure BT, Szentagotai A, Dobrean A, David D. Evidence based clinical assessment of child and adolescent social phobia: a critical review of rating scales. *Child Psychiatry and Human Development*. 2012; 43:795–820.10.1007/s10578-012-0297-y [PubMed: 22438106]
- van der Ende J, Verhulst FC, Tiemeier H. Agreement of informants on emotional and behavioral problems from childhood to adulthood. *Psychological Assessment*. 2012; 24:293–300.10.1037/a0025500 [PubMed: 21928910]

- Villabø M, Gere M, Torgersen S, March JS, Kendall PC. Diagnostic efficiency of the child and parent versions of the Multidimensional Anxiety Scale for Children. *Journal of Clinical Child and Adolescent Psychology*. 2012; 41:75–85.10.1080/15374416.2012.632350 [PubMed: 22233247]
- Wakschlag LS, Tolan PH, Leventhal BL. Research review: 'Ain't misbehavin': Towards a developmentally-specified nosology for preschool disruptive behavior. *Journal of Child Psychology and Psychiatry*. 2010; 51:3–22.10.1111/j.1469-7610.2009.02184.x [PubMed: 19874427]
- Weissman MM, Wickramaratne P, Warner V, John K, Prusoff BA, Merikangas KR, Gammon GD. Assessing psychiatric disorders in children: discrepancies between mothers' and children's reports. *Archives of General Psychiatry*. 1987; 44:747–753.10.1001/archpsyc.1987.01800200075011 [PubMed: 3632247]
- Weisz JR, Jensen Doss A, Hawley KM. Youth psychotherapy outcome research: A review and critique of the evidence base. *Annual Review of Psychology*. 2005; 56:337–363.10.1146/annurev.psych.55.090902.141449
- Westen D, Rosenthal R. Quantifying construct validity: two simple measures. *Journal of Personality and Social Psychology*. 2003; 84:608–618.10.1037/0022-3514.84.3.608 [PubMed: 12635920]
- Westen D, Rosenthal R. Improving construct validity: Cronbach, Meehl, and Neurath's ship. *Psychological Assessment*. 2005; 17:409–412.10.1037/1040-3590.17.4.409 [PubMed: 16393006]
- Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. Bossuyt PM. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*. 2011; 155:529–536.10.1059/0003-4819-155-8-201110180-00009 [PubMed: 22007046]
- Yates BT, Taub J. Assessing the costs, benefits, cost-effectiveness, and cost-benefit of psychological assessment: We should, we can, and here's how. *Psychological Assessment*. 2003; 15:478–495.10.1037/1040-3590.15.4.478 [PubMed: 14692844]
- Yeh M, Weisz JR. Why are we here at the clinic? Parent–child (dis)agreement on referral problems at outpatient treatment entry. *Journal of Consulting and Clinical Psychology*. 2001; 69:1018–1025.10.1037/0022-006X.69.6.1018 [PubMed: 11777105]
- Youngstrom, E. Evidence-based strategies for the assessment of developmental psychopathology: Measuring prediction, prescription, and process. In: Craighead, WE.; Milkowitz, DJ., editors. *Psychopathology History, diagnosis, and empirical foundations*. Hoboken: Wiley; 2008. p. 34-77.
- Youngstrom EA. Future directions in psychological assessment: combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *Journal of Clinical Child and Adolescent Psychology*. 2013; 42:139–59.10.1080/15374416.2012.736358 [PubMed: 23153181]
- Youngstrom E, Findling RL, Calabrese JR. Who are the comorbid adolescents? Agreement between psychiatric diagnosis, youth, parent, and teacher report. *Journal of Abnormal Child Psychology*. 2003; 31:231–245.10.1023/A:1023244512119 [PubMed: 12774858]
- Youngstrom EA, Findling RL, Calabrese JR, Gracious BL, Demeter C, Bedoya DD, Price M. Comparing the diagnostic accuracy of six potential screening instruments for Bipolar Disorder in youth aged 5 to 17 years. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2004a; 43:847–858.10.1097/01.chi.0000125091.35109.1e [PubMed: 15213586]
- Youngstrom EA, Findling RL, Calabrese JR. Effects of adolescent manic symptoms on agreement between youth, parent, and teacher ratings of behavior problems. *Journal of Affective Disorders*. 2004b; 82S:S5–S16.10.1016/j.jad.2004.05.016 [PubMed: 15571790]
- Youngstrom, EA.; Kahana, SY.; Starr, M.; Schwartz, C. Emotion states and bias in clinical ratings: Causal but complicated. In: De Los Reyes, A.; Saavedra, LM., editors. *Informant discrepancies in clinical child research: Processes of informant discrepancies and distortions and impact on clinical decision-making*; Symposium conducted at the annual meeting of the Association for the Advancement of Behavior Therapy; New Orleans. 2004 Nov. Chairs
- Youngstrom E, Izard C, Ackerman B. Dysphoria-related bias in maternal ratings of children. *Journal of Consulting and Clinical Psychology*. 1999; 67:905–916.10.1037//0022-006X.67.6.905 [PubMed: 10596512]

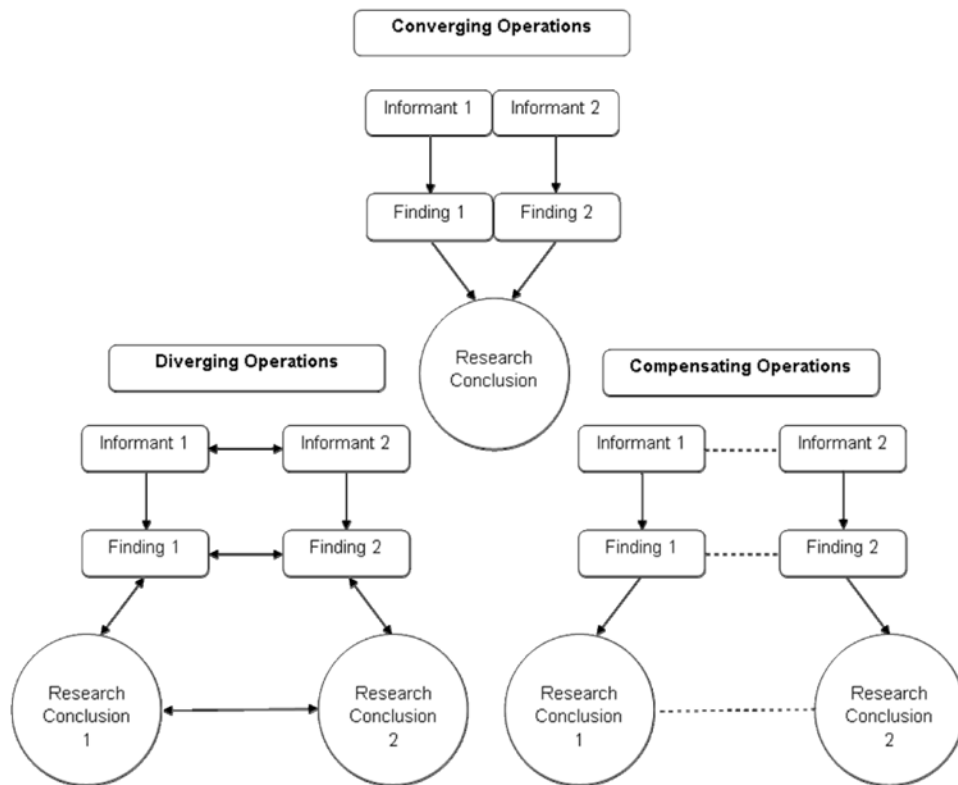


Figure 1.

Graphical representation of the research concepts that comprise the Operations Triad Model. The top half (Figure 1a) represents Converging Operations: a set of measurement conditions for interpreting patterns of findings based on the consistency within which findings yield similar conclusions. The bottom half denotes two circumstances within which researchers identify discrepancies across empirical findings derived from multiple informants' reports and thus discrepancies in the research conclusions drawn from these reports. On the left (Figure 1b) is a graphical representation of Diverging Operations: a set of measurement conditions for interpreting patterns of inconsistent findings based on hypotheses about variations in the behavior(s) assessed. The solid lines linking informants' reports, empirical findings derived from these reports, and conclusions based on empirical findings denote the systematic relations among these three study components. Further, the presence of dual arrowheads in the figure representing Diverging Operations conveys the idea that one ties meaning to the discrepancies among empirical findings and research conclusions and thus how one interprets informants' reports to vary as a function of variation in the behaviors being assessed. Lastly, on the right (Figure 1c) is a graphical representation of Compensating Operations: a set of measurement conditions for interpreting patterns of inconsistent findings based on methodological features of the study's measures or informants. The dashed lines denote the lack of systematic relations among informants' reports, empirical findings, and research conclusions. Originally published in De Los Reyes, Thomas, et al. (2013). © Annual Review of Clinical Psychology. Copyright 2012 Annual Reviews. All rights reserved. The Annual Reviews logo, and other Annual Reviews products

referenced herein are either registered trademarks or trademarks of Annual Reviews. All other marks are the property of their respective owner and/or licensor.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

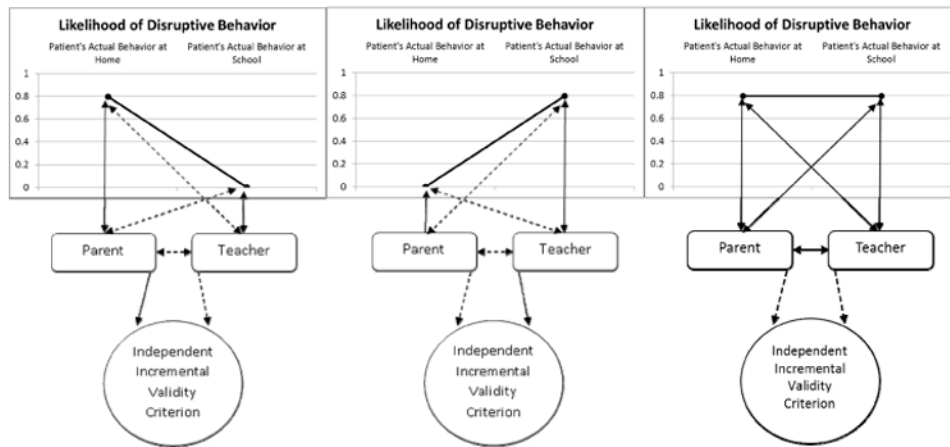


Figure 2. Graphical depiction of incremental and construct validity predictions that follow from the OTM and conceptual rationale for taking a multi-informant approach to mental health assessment.

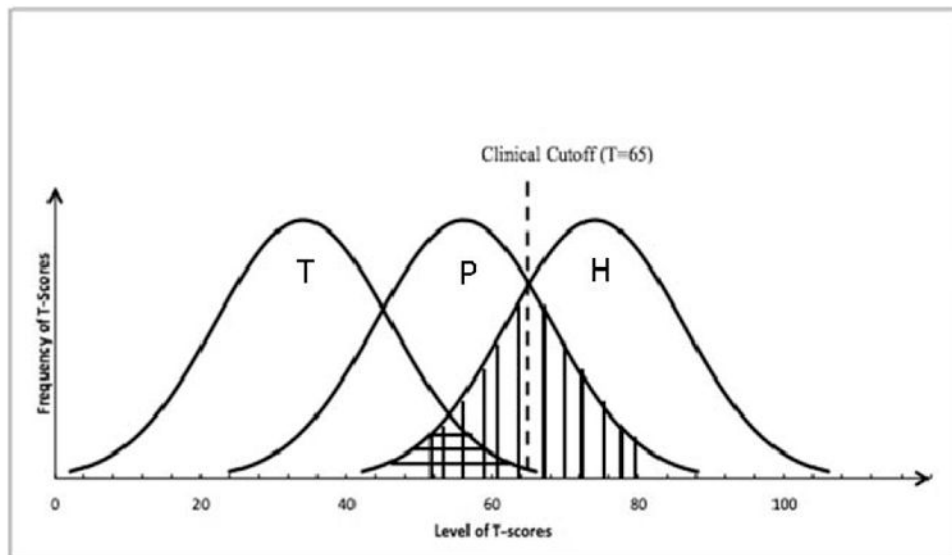


Figure 3. Graphical depiction of T-score distributions of teacher reports (left, labeled “T”), parent reports (center, labeled “P”), and independent home-based assessments of disruptive behavior (right, labeled “H”), in a hypothetical sample of young children.

Table 1
Meta-Analytic Findings Regarding Cross-Informant Correspondence in Reports of Children's Mental Health

Mean Cross-Informant Correlations						
Symptom Type	# of Data Points	Overall Effect Size (95% CI)	P-C	P-T	T-C	M-F
Internalizing	697	.25 (.24-.25) ($p < .001$)	.26 ($n = 372$)	.21 ($n = 155$)	.20 ($n = 62$)	.48 ($n = 109$)
Externalizing	521	.30 (.30-.31) ($p < .001$)	.32 ($n = 198$)	.28 ($n = 162$)	.29 ($n = 53$)	.58 ($n = 109$)

Note. Correlations are in the Pearson r metric. P-C = Parent-Child; P-T = Parent-Teacher; T-C = Teacher-Child; M-F = Mother-Father.

Table 2
Summary of Exemplary Findings Regarding the Incremental Validity of Cross-Informant Reports When Assessing Children's Psychosocial Functioning

Symptom/Disorder	Author Group (year)	Informants	Description
ADHD	Power et al. (1998)	P, T	Researchers recommended use of a single parent report or teacher report to <i>rule out</i> ADHD in children, but to combine parent report and teacher report to identify children meeting criteria for an ADHD diagnosis.
	Smith, Pelham, Gnagy, Molina, & Evans (2000)	C, T	On average, adolescent self-reports contributed less than a 1% unique variance above teacher ratings in the assessment of ADHD. In contrast, teacher reports contributed, on average, 8% unique variance relative to self-report. However, teacher reports evidenced domain-specific incremental validity (i.e., higher unique variance contributed by teachers for behaviors displayed in the classroom vs. outside the classroom).
ADHD subtypes and ODD	Owens & Hoza (2003)	P, T	When assessing Inattentive or Hyperactive/Impulsive subtypes of ADHD, combining parent and teacher reports did not provide any additional information above relying on parent report or teacher report alone. In contrast, when assessing ADHD, Combined Type or ODD, teacher reports alone and the combination of teacher reports and parent reports proved superior to parent reports alone in predicting a diagnosis.
Anxiety	Villabø, Gere, Torgersen, March, & Kendall (2012)	C, P	Adding parent reports to child self-reports (i.e., on the MASC) increased prediction of children's anxiety disorder diagnoses. Specifically, parent reports incrementally predicted children's Separation Anxiety Disorder diagnoses, whereas child self-reports incrementally predicted children's Social Phobia diagnoses.
Bipolar Disorder	Youngstrom et al. (2004a)	C, P, T	Across estimates of diagnostic efficiency for six different rating scales, parent reports evidenced superior prediction of a pediatric bipolar disorder diagnosis relative to both self-reports and teacher-reports. Additionally, combining self-reports or teacher reports with parent reports did not significantly enhance the prediction of a diagnosis of pediatric bipolar disorder.
ODD	Angold & Costello (1996)	C, P	Children tended to self-report fewer oppositional symptoms relative to parents' reports about children. However, children self-reported certain oppositional symptoms (i.e., frequent loss of temper) at greater levels relative to parents' reports about children. Researchers recommended both self-reports and parent reports whenever possible when considering an ODD diagnosis.
Relational Adjustment	Ladd & Kochenderfer-Ladd (2002)	C, P, T, Peer	No single informant's report of victimization consistently predicted relational adjustment across grades 2-4. The informant whose report explained the most variance in relational adjustment varied as a function of grade: (a) teacher reports in grade 2, (b) teacher reports and self-reports in grade 3, and (c) a combination of self-reports and teacher reports in grade 4. Overall, parent reports explained little unique variance.
Social Skills and Academic Competence	Kwon, Kim, & Sheridan (2012)	T, Peer	Teacher-reported social skills for children in grades 3-5 did not uniquely or significantly predict social status indicators (e.g., likeability and popularity) above information obtained from peer reports. In contrast, both peer-reported and teacher-reported social skills demonstrated incremental validity in predicting academic competence and children's positive school functioning.

Note. C = Child; P = Parent; T = Teacher. ADHD = Attention-Deficit/Hyperactivity Disorder; MASC = Multidimensional Anxiety Scale for Children; ODD = Oppositional Defiant Disorder; MDD = Major Depressive Disorder; CD = Conduct Disorder.

Table 3
Cross-Informant Correspondence and Relations between Referral Status and Informants' Reports, Based on Parent, Teacher, and Youth Forms of the Achenbach System of Empirically Based Assessments (Achenbach & Rescorla, 2001)

Correlation Estimates of Cross-Informant Correspondence			
Psychopathology Domain	<i>CBCL-YSR</i>	<i>CBCL-TRF</i>	<i>YSR-TRF</i>
Internalizing Problems	.48	.21	.17
Externalizing Problems	.56	.36	.28
Total Problems	.54	.35	.21

Estimated Common Variance Attributable to Multiple Informants ^a	
Psychopathology Domain	
Internalizing Problems	1.3%
Externalizing Problems	3.0%
Total Problems	0.8%

Estimated Variance in Informant Accounted for by Referral Status, Converted from Point Biserial Pearson <i>r</i> Correlations			
Psychopathology Domain	<i>CBCL</i>	<i>YSR</i>	<i>TRF</i>
Internalizing Problems	26%	11%	14%
Externalizing Problems	33%	17%	19%
Total Problems	36%	15%	26%

Estimated Correlation with Referral Status for a Single Informant			
Psychopathology Domain	<i>CBCL</i>	<i>YSR</i>	<i>TRF</i>
Internalizing Problems	.51	.33	.37
Externalizing Problems	.57	.41	.44
Total Problems	.60	.39	.51

Estimated Maximum Possible Correlation with Referral Status for Multiple Informants ^b	
Psychopathology Domain	
Internalizing Problems	.72
Externalizing Problems	.85
Total Problems	.88

Note. CBCL = Child Behavior Checklist (i.e., Parent Report Form); YSR = Youth Self-Report; TRF = Teacher Report Form.

^aEstimates based on regression commonality analyses illustrated by Cohen et al. (2003) and Nimon (2010), using cross-informant correlations reported above.

^bBy "maximum," we mean that we estimated an upper limit correlation between informants' reports and referral status, assuming unique variance among the informants' reports.

Table 4
Summary of Studies on the Relation Between Informant Discrepancies in Parent and Child Reports of Children's Mental Health and Clinical Decision-Making

Author Group (year)	Mean Patient Age in Years (age range)	Description	Did Clinicians' Decisions Correspond to a Greater Extent With One Informant Relative to Other Informants?
Brown-Jacobsen, Wallace, & Whiteside (2011)	12.3 (7-18)	When assessing children's anxiety symptoms, both parent- and child self-reported symptoms were significantly related to the clinician's diagnostic impression. For parent and child questionnaire reports, considered simultaneously, only parent reports uniquely related to the clinician's diagnostic impression of the child. Conversely, when only the parent or the child endorsed an anxiety disorder diagnosis, the clinician made a diagnosis approximately 30% of the time, with the prevalence based on parent-reported symptoms (28.6%) slightly higher than child-reported symptoms (26.8%).	Yes, but clinicians' decisions corresponded more with parent reports relative to child self-reports, depending on the measurement methodology
De Los Reyes, Alfano, & Beidel (2011)	11.6 (7-16)	When assessing pre- and post-treatment <i>social functioning</i> in a sample of children, clinician ratings of treatment improvement more strongly agreed with parent-reported post-treatment functioning, relative to children's self-reported functioning.	Yes (Parent > Child)
DiBartolo, Albano, Barlow, & Heimberg (1998)	13.9 (8-17)	Overall, clinicians weighted parent reports more strongly than they did child self-reports when determining the severity of child social phobia.	Yes (Parent > Child)
Grills & Ollendick (2003)	10.7 (7-17)	When assessing children's anxiety symptoms, a clinicians' consensus diagnoses more highly correlated with parent reports, relative to child self-reports.	Yes (Parent > Child)
Hawley & Weisz (2003)	11.5 (7-17)	When asked to identify target problems in outpatient treatment, low agreement was found within the child-parent-therapist triad. More than ¾ of the participating triads disagreed on target problems at the beginning of treatment. Therapists exhibited significantly higher agreement with parents versus children on the majority of target problems except for family and environmental problems where the therapists agreed more with child self-report than parent report.	Yes, but clinicians' decisions corresponded more with parent reports or child self-reports, depending on the problem domain
Kramer et al. (2004)	13.7 (11-18)	Parent-adolescent agreement on reports of functional impairment was low, and clinician agreement was variable. Overall, clinicians rated impairment as more severe when both the parent and adolescent endorsed the problem; however, when a problem was reported a single informant, the weight given to the report by the clinician varied by problem type. For example, for problems endorsed by a parent alone, clinicians generally weighted the severity as equal to a problem endorsed by both the parent and adolescent. In contrast, some areas of impairment such as sibling issues or school work were rated as less severe when only reported by the adolescent versus by both the parent and adolescent, whereas adolescent-only reports of illegal behavior were rated more serious than parent-only reports.	Yes, but clinicians' decisions corresponded more with parent reports or adolescent self-reports, depending on the problem domain
Youngstrom, Findling, & Calabrese (2004b)	14.1 (11-17)	Clinician's ratings of adolescent manic symptoms were more strongly correlated with parent reports relative to adolescent self-reports.	Yes (Parent > Child)

Table 5
Summary of Characteristics in Studies Examining the Incremental Validity of a Multi-Informant Approach to the Assessment of Mental Health

Mental Health Domain	Descriptive Information		Informants Examined for Incremental Validity			Criterion Measures			Contextual Variations in Assessed Behavior	
	Author Group (year)	Informants	Use of the Same Measure Format to Gather Informants' Reports	Information Sources Used for Criterion Measures	Criterion Independent from Informants Examined for Incremental Validity	Objective Criterion Used ^a	Independent Assessment of Prevalence of Contextual Variations in Participants' Behavior			
ADHD	Power et al. (1998)	P, T	✓	Combined diagnostic interview and T report	X	X	X			
	Smith, Pelham, Gnagy, Molina, & Evans (2000)	C, T	/	Independent observers	✓	✓	X			
ADHD subtypes and ODD	Owens & Hoza (2003)	P, T	✓	P, T	X	X	X			
Anxiety	Villabø, Gere, Torgersen, March, & Kendall (2012)	C, P	✓	Diagnostic interview	X	X	X			
Bipolar Disorder	Youngstrom et al. (2004a)	C, P, T	/	Diagnostic interview	X	X	X			
ODD	Angold & Costello (1996)	C, P	✓	Diagnostic interview; P and C interview reports of impairment, parental burden, and service use	X	X	X			
Relational Adjustment	Ladd & Kochenderfer-Ladd (2002)	C, P, T, Peer	/	C, P, T, Peer	X	X	X			
Social Skills and Academic Competence	Kwon, Kim, & Sheridan (2012)	T, Peer	X	C, T, Peer	/	X	X			

Note. ✓ = study utilized characteristic described by column header; / = study utilized characteristic described by column header for some but not all measures; X = study did not utilize characteristic described by column header. C = Child; P = Parent; T = Teacher. ADHD = Attention-Deficit/Hyperactivity Disorder; ODD = oppositional defiant behaviors.

^a criterion based on objective measure (e.g., record reviews, behavioral observations, or behavioral measures).