Research Review: Multi-informant integration in child and adolescent psychopathology diagnosis

Michelle M. Martel[1], **Kristian Markon**[2], and **Gregory T. Smith**[1]

[1]Department of Psychology, University of Kentucky, KY, USA

[2]Department of Psychological and Brain Sciences, University of Iowa, IA, USA

Abstract

Background—An empirically based, clinically usable approach to cross-informant integration in clinical assessment is needed. Although the importance of this ongoing issue is becoming increasingly recognized, little in the way of solid recommendations is currently provided to researchers and clinicians seeking to incorporate multiple informant reports in diagnosis of child psychopathology. The issue is timely because recent developments have created new opportunities for improved handling of this problem. For example, advanced theories of psychopathology and normal and abnormal child development provide theoretical guidance for how integration of multiple informants should be handled for specific disorders and at particular ages. In addition, more sophisticated data analytic approaches are now available, including advanced latent variable models, which allow for complex measurement approaches with consideration of measurement invariance.

Findings—The increasing availability and mobility of computing devices suggests that it will be increasingly feasible for clinicians to implement more advanced methods rather than being confined to the easily-memorized algorithms of the DSM system.

Conclusions—Development of models of cross-informant integration for individual disorders based on theory and tests of the incremental validity of more sophisticated cross-informant integration approaches in comparison to external validation criteria (e.g., longitudinal trajectories and outcomes; treatment response; behavior genetic etiology) should be a focus of future work.

Introduction

Cross-informant integration in developmental psychopathological diagnosis is a longstanding problem because of the modest agreement among informants. Influential reviews date back to 1987 when Achenbach and colleagues conducted a meta-analysis of agreement between different informant reports of child and adolescent behavioral and emotional problems. Achenbach and colleagues' (1987) meta-analytic review indicated that the highest agreement was found between more similar informants (e.g., parents; r=.60) followed by different types of informants (e.g., parents and teachers; r=.28) with the lowest

Correspondence: Michelle M. Martel, Department of Psychology, University of Kentucky, 207C Kastle Hall, Lexington, Kentucky, 40506, United States; michelle.martel@uky.edu.

agreement found between the self and an other (i.e., child/adolescent and other; r=.22). Furthermore, higher agreement was found for externalizing (i.e., undercontrolled) versus internalizing (i.e., overcontrolled) problems (Achenbach et al., 1987). A subsequent meta-analysis conducted on inter-parental agreement in ratings of child internalizing and externalizing problems also found that agreement between parents was moderate with high(est) agreement for ratings of child externalizing behavior (Duhig et al., 2000). Finally, De Los Reyes and colleagues' (2015) very recent meta-analysis suggested similar moderate correspondence (r=.28) between informant reports with slightly higher correspondence for externalizing (vs. internalizing) problems. Given the modest agreement across informants, it remains unclear how to integrate different informant reports in research and clinical settings with important implications for diagnostic reliability and validity.

Importantly, there have been recent promising advances in addressing this problem both from a theoretical perspective (De Los Reyes, 2013; De Los Reyes & Kazdin, 2005) and in terms of emerging data analytic strategies (Horton & Fitzmaurice, 2004). This paper will review these advances, highlight emerging analytic strategies, and provide an empirical example in which we utilize such an approach to assess the viability of a novel multiple informant approach to assessment of a common childhood disorder plagued by this problem: Attention-Deficit/Hyperactivity Disorder (ADHD).

Theoretical advances concerning informant discrepancies

Based on a synthesis of existing work to date, there appear to be a number of different explanations for discrepancies in informant ratings of child behavior including (1) cross-situational variability in child behavior, (2) differential ABC (i.e., antecedents and consequences) contextual demands across settings, (3) the different perspectives of raters, (4) different attributions regarding the causes of child behavior, (5) parent characteristics (e.g., psychopathology) influencing ratings, (6) child characteristics (e.g., age, gender) influencing ratings, (7) the visibility of attributes rated (e.g, internalizing versus externalizing), and (8) externally-generated rating bias (e.g., social desirability; negative or positive rater main effects) with various factors potentially interacting with one another (Connelly & Ones, 2010). Empirical work by De Los Reyes and others likewise suggests informant discrepancies often have high internal reliability and relate meaningfully to outcomes such as self-report youth mood problems (De Los Reyes et al., 2011). For example, observations of disruptive behavior problems in laboratory observational paradigms conducted with trained examiners versus parents relate to discrepancies in ratings made by teachers versus parents, consistent with the idea that there is some aspect of situational specificity of child behavior that differentially influence informant ratings made on the basis of different situational contexts (De Los Reyes et al., 2009). An emerging consensus in the field is that such discrepancies should be modelled explicitly rather than just ignored as error (Kraemer et al., 2003).

De Los Reyes (2005; 2013) has suggested that discrepancies (vs. agreement) in informant ratings of child psychopathology should be 'embrace[d], not erase[d]' and has recommended use of a theoretically-informed perspective, the Attribution Bias Context Model, for understanding these discrepancies. In brief, this model suggests that differences in informant

attributions regarding the cause of the child's problems and differences in informant perspectives interacting with the clinical assessment process lead to informant discrepancies in clinical practice. De Los Reyes and colleagues (2013) have subsequently extended this idea by advancing an updated Operations Triad Model which provides a framework for how to evaluate informant discrepancies and integrate across informant reports. They provide a decision tree for interpreting informant reports and deciding whether meaningful information can be extracted from informant discrepancies. In short, the authors suggest that theoretical considerations should dictate whether informant correspondence or discrepancies are hypothesized and then tests of converging (i.e., agreement between reporters) or diverging (disagreement between reporters with reliable, valid, and strong methodological ratings) operations can be tested. If converging or diverging operations are not supported, then posthoc tests of compensating operations should be tested; that is, researchers should evaluate whether unreliability, lack of validity, or methodological factors can explain lack of convergence or divergence.

The core implication of this work is that discrepancies are themselves substantively important, and this has moved the field forward, as well as providing opportunities for further work in this area. Recognizing the value in the discrepancies further highlights the importance of obtaining multiple informant reports, which is of course essential for some disorders (e.g., ADHD; APA, 2013). However, a salient limitation of work in this area for researchers and clinicians alike is that there are no current recommendations for how to integrate multiple informant reports in child and adolescent psychopathology for the purpose of making clinical or research diagnoses at the level of specific disorders.

Existing approaches for managing multiple informants

In practice, clinicians often end up needing to make a judgment call when informants disagree, and there are little or no empirically-based guidelines about how to go about making this decision. For research, a number of approaches have been suggested. Various pooling strategies for combining information from multiple sources into a single number have been popular (Horton & Fitzmaurice, 2004). Notable examples of these types of approaches include the OR and AND algorithms in which if either informant (for OR algorithm) or both informants (for AND algorithm) endorse a symptom as present, it is counted in the sum score total (e.g., see Lahey et al., 1994). Most agree these algorithms are less than ideal given that they perform worse as the number of informants increase and do not take measurement error into account (e.g., Solanto & Alvir, 2009; Valo & Tannock, 2010). A differential weighting approach has also been advocated, based on either principal components, confirmatory factor analysis, or preference given to the informant who knows the child best; yet, these weighting approaches have historically not outperformed equal ratings and are more complex, as well as being dependent on specific sample weights and/or subjective clinician judgement (Piacentini, Cohen, & Cohen, 1992). Yet, they may merit more attention in the future as better psychometric approaches are available to yield better weighting estimates. Another commonly-used example of a pooling approach is 'best estimate' diagnosis in which a clinical team gets together and reviews all available information before deciding together on diagnostic or criterion status based on consensus opinion. Although this has the advantage of representing a clinical consensus, it is costly,

does not directly address the problem of how to combine different informant reports, and relies on clinical judgment rather than statistical decision making procedures (Grove & Meehl, 1996). It is also unclear how it might be utilized in clinical settings. A final example is an average approach to symptom rating integration (which entails averaging across different informant ratings) which is possibly the strongest option based on psychometric criteria (reviewed by Horton & Fitzmaurice, 2004), although it should be noted that this approach makes the assumption that cross-situational generalizability (vs. discrepancies) of behavior should be emphasized. By focusing on a single measurement, however, none of the pooling approaches take advantage of findings that discrepancies actually provide useful diagnostic information.

There are also a number of other approaches available when differential informant effects are of interest in data analysis. However, all of these approaches are limited in their potential clinical utility, at least barring further technological developments. For example, one can examine different informant reports separately to identify differential associations with external criteria, but this approach increases Type I error and does not allow for direct comparison of results. Regression with multiple outcomes, a special case of GLM, is an alternative strategy that allows for more direct comparison and control of Type I error (see Horton & Fitzmaurice, 2004). This strategy, while retaining information about individual informant report effects does not allow for explicit modeling of convergence and divergence between the two and does not allow for integration of the two in any way.

There are several ways to explicitly model informant discrepancies. A manifest difference score between different informant reports can be calculated to provide a measure of the magnitude of divergence between reports. However, this approach is entirely redundant with the original reports (Laird & Weems, 2011). In addition, it suffers from interpretation problems, including how to interpret low difference scores in that they could refer to agreement about high OR low symptomatology. Finally, this approach is vulnerable to low reliability because it includes error.

In regression analyses, discrepancies in informant reports can be represented using interaction terms predictive of external criteria (De Los Reyes, 2013; Laird & De Los Reyes, 2013; Laird & LaFleur, 2016; Tackett et al., 2013). However, these interaction terms do not characterize overarching patterns of informant discrepancies (e.g., parents reporting higher levels of symptoms than teachers); they only help to describe idiosyncratic differences in informant reports.

Another approach to examination of convergence and discrepancies in multiple informant reports is principal components analysis (PCA; De Los Reyes, 2013; van Dulmen & Egeland, 2011). Kraemer and colleagues (2003), for example, suggest that modeling the orthogonal variance in informant reports in PCA allows for trait versus perspective versus context to be isolated. Yet, orthogonal reports would seem difficult, if not impossible, to obtain in practice. Further, since error will be conflated with contextual effects, this seems less than ideal.

More advanced latent factor approaches, such as latent profile/class analysis, allows these effects to be parsed from one another and would seem preferable since both error and contextual effects are likely to be important. Different types of discrepancy and agreement patterns, such as discrepancies in opposite directions versus informant agreement, can be captured by latent profile/class analysis (De Los Reyes et al., 2011; De Los Reyes, 2013). Although this is a helpful descriptive method for summarizing different profiles, particularly in research settings, it is also limited in that it may oversimplify actual patterns of discrepancies by categorizing them.

Notable limitations of existing approaches

It is imperative that currently-proposed approaches to multiple informant integration utilize theoretical models as frameworks for exploring informant convergence and discrepancies in relation to individual disorders (as suggested by De Los Reyes et al., 2013; Garner, Hake, & Eriksen, 1956). That is, specific disorders may require different methods for handling informant discrepancies depending on the theoretical conceptualization of the disorder. For example, ADHD is currently conceptualized in the DSM as a trait manifest across settings, and this is made explicit in the diagnostic criteria by the requirement that several symptoms be present in two or more settings (APA, 2013; Loeber et al., 1989). Therefore, for ADHD, current diagnostic criteria requires that some symptoms must be present in 2 or more settings, as rated by different individuals, thus suggesting the importance of some convergence in ratings (i.e., Converging Operations; De Los Reyes et al., 2013).

However, this would not necessarily be the best approach for other disorders. Research on self and informant reports of psychopathology and other related constructs such as personality suggest that informants may vary in their ability to accurately report on particular traits. For example, those traits that are less observable including neuroticism, negative affect, and depression, may be most accurately reported on by the self. However, those traits that are more observable, including anxiety, extraversion, agreeableness, relationship patterns, as well as those that are highly evaluative in nature (e.g., narcissism) may be best reported on by others (Connelly & Ones, 2010; Klein, 2003; Klonsky & Oltmanns, 2002; Markon et al., 2013; Vazire, 2010). In line with this idea, others' ratings of observable or evaluative aspects of personality seem to add incremental validity in relation to prediction of external criteria (e.g., academic performance, quality of relationships; Connelly & Ones, 2010; Miller, Pilkonis, & Clifton, 2005; Vazire & Mehl, 2008). Thus, while multi-informant assessment of all of these problems is likely beneficial, self ratings may require heavier weighting for depressive symptoms (Jensen, 1999; Klein, Dougherty, & Olino, 2005), while other ratings may require heavier weighting for relational problems or more observable behaviors such as anxiety and bipolar disorder (Silverman & Ollendick, 2005; Youngstrom, Findling, Youngstrom, & Calabrese, 2005).

In addition to the importance of considering the possible utility of disorder-specific theories to explain discrepancies, it is important to consider development, and in particular developmental differences in (a) reporter knowledge of target behaviors and (b) reporting ability. For example, when Conduct Disorder (i.e., 'a persistent pattern of disregarding the basic rights of others or violation of major age-appropriate societal norms or rules') is

assessed in adolescents, more weight may need to be given to adolescent self-report than when it is assessed in children, because parents and teachers of adolescents may have less first-hand knowledge of an adolescent's behavior than of a child's behavior (Loeber et al., 1989). Another consideration is that, at least for externalizing disorders, younger children appear to be less valid reporters than older children and adolescents (Cantwell et al., 1997; Verhulst & Ende, 1992). Therefore, other reporters are likely to be more accurate reporters of child externalizing behavior problems.

Thus, important and promising advances in modeling informant discrepancies include (a) the emergence and use of theoretical models specifying the importance and likelihood of informant agreement and informant accuracy at the level of individual disorders, and (b) recognition of the importance of psychological development in influencing informant access to the constructs to be rated, which in turn influences the degree of informant agreement. For example, based on the current state of knowledge, parent and teacher convergence in ADHD ratings of child behavior appear to be an important goal, with respect to the validity of diagnostic conclusions. Parental reports of child Oppositional Defiant Disorder symptoms may deserve to be weighted heavily, just as do self-ratings of adolescent Conduct Disorder symptoms; recognizing, of course, that other informants provide valuable additional information. In fact, for Conduct Disorder in particular discrepancies between parents and adolescents may provide information important for treatment such as lack of adequate parental monitoring (Fergusson, Boden, & Horwood, 2009; Laird & LeFluer, 2016). Self-ratings appear particularly important and may deserve heavy weighting for childhood and adolescent depressive problems, although – here again—others may provide important supplementary information; yet, the more readily observable anxiety and bipolar symptoms may be readily reported on by multiple informants who can all provide important information. With respect to personality disorders, self-ratings may prove most important for negative affect components, whereas ratings by others may prove most important for assessing less flattering traits like narcissism, relational problems, and psychopathy. These and other possibilities merit empirical examination.

Promising new or underutilized approaches to testing informant convergence and discrepancies: multitrait-multimethod matrix and structural equation modeling

There are two particularly promising approaches to testing informant convergence and discrepancies in ratings of child and adolescent psychopathology that are relatively new or under-utilized: (1) Multitrait-multimethod matrix approach and (2) structural equation modeling. These should be developed and tested in research settings and imported into the clinic as vetted. Campbell and Fiske's (1959) multitrait-multimethod (MTMM) matrix approach to establishing convergent and discriminant validity for measures of psychological constructs remains highly influential in the field and can provide a means for the integration of multiple informant ratings of psychological constructs (Kenny & Kashy, 1992; Schmitt & Stults, 1986; Widaman, 1992). Whereas traditional MTMM models are utilized with different methods of assessment, in the current case, one can instead examine different

informants of the target traits. Therefore, we encourage researchers to use MTMM as a preliminary step in examination of cross-informant agreement and discrepancies.

A second promising step in evaluating multiple informant agreement and discrepancies is to utilize formal latent variable models such as confirmatory factor analysis (CFA) or structural equation modeling (SEM). With these approaches, it is possible to systematically model multi-method data such as ratings by multiple informants (Lance, Noble, & Scullen, 2002; van Dulmen & Egeland, 2011). When modeling multiple informant data, it is vitally important to choose the appropriate model for a given disorder, as discussed above. Is a disorder best represented by one, two, or three factors? Alternatively, is a disorder best understood in terms of higher-order and lower-order dimensions? If so, then second-order or bifactor (also known as hierarchical) models provide alternatives to simple factor models by enabling simultaneous estimation of general and specific factors. However, higher-order and hierarchical models still suggest quite different assumptions about underlying structure. A second-order factor model allows individual symptom domains to be modeled separately with symptom domains being entirely encompassed by a higher-order factor. A bifactor model is conceptually distinct. It allows for individual symptoms to simultaneously load onto an overall, or 'general' ('g'), factor along with completely or partially distinct ('specific;' 's') latent components. This model would be in line with a multiple-pathway, or multiple component process, conception of disorder.

As noted above, the choice of a particular model requires a theoretically-driven conceptualization of the disorder and this will likely vary, at least somewhat, from disorder to disorder. ADHD, for example, with its two symptom domains of inattention and hyperactivity-impulsivity, appears best characterized by a bifactor structure of a general ADHD factor and specific inattention and hyperactivity-impulsivity factors (Martel et al., 2010; Toplak et al., 2009). Other disorders will be best conceptualized and represented by other factor structures. Once a model is chosen, there are a series of steps researchers can take to evaluate the presence of, and understand the meaning of, informant discrepancies. For example, in a bifactor model, items from different informants may differ in their average loadings on a general factor. Such a finding might speak to the relative agreement of the different reports.

One could also fit Bauer and colleague's (2013) tri-factor model to informant data based on theoretical considerations. This model provides a nice example of a viable measurement model with particular relevance to critical examination of multiple informant ratings of a construct. In this model, use of specific items rated by multiple informants allows estimation of common (consensus) views of target behaviors (or trait aspects of behavior), the unique perspectives of each informant, and specific variance associated with each item (Bauer et al., 2013). By decomposing ratings into components associated with consensus ratings of the target, specific informant perspectives of the target, and specific symptom domain, or item, ratings, this model helps quantify how much each indicator reflects the many different influences on ratings. That is, the model allows one to examine discrepancies in informant perspectives after accounting for consensus, or trait, views, as well as accounting for specific item or symptom domain components, based on theory of disorder conceptualization. Since the model helps to identify items that tend to reflect shared perspectives of informants, as

well as those that better reflect unique perspectives of a given type of informant, it helps us to better understand informant convergence and discrepancy.

Further, in all models, including the trifactor model, tests of measurement invariance can also be useful to test whether the items function the same for different reporters (Geiser, Burns, & Servera, 2014; Olino & Klein, 2015; Woehr, Sheehan & Bennett, 2005). Item response theory (IRT) can provide information about differential item functioning (DIF) and thus utility of potential weighting of individual symptom items. IRT allows both item and person characteristics (e.g., item severity; person trait level) to be explicitly modeled. Two-parameter IRT models, which have been applied to diagnoses and multiple informant ratings, provide discrimination and difficulty (i.e., severity) estimates for each item. A discrimination parameter is akin to a loading in factor analysis and reflects the amount of information provided by an item at that point in the trait continuum where the item provides the most information; in other words, this parameter reflects the strength of an item's relationship with the construct. Higher numbers suggest the item is more strongly related to the overall construct. A difficulty or severity parameter, in contrast, reflects the location on the trait continuum where the item provides the most information; for a binary indicator (e.g., a diagnostic criterion), this parameter indicates the trait level required for there to be a 50% probability of endorsing the particular item. In a commonly used version of the two-parameter logistic model, higher parameter estimates suggest higher levels are required for item endorsement.

Sophisticated tests of differential item functioning (DIF; a form of measurement noninvariance with discrete items) can be conducted in IRT. DIF in IRT allows for the possibility that one informant might be a better rater for part of the continuum, but not the other part of the continuum (de Ayala, 2009). There are many advantages to having this information. For example, suppose informants tend to agree on high levels of symptomology. If disagreement only occurs at low or moderate levels of symptoms, then this may not a problem in measuring clinically significant levels of psychopathology (Jin & Wang, 2014; Markon et al., 2013; Reise et al., 2011).

Random item IRT, or random-effects IRT with random item parameters, allows one to represent situations where the measurement properties of items have some random variability, as might be expected with informants. Typically, the specific informants that are drawn on represent some random sample of a possible pool of informants; random item IRT provides a mechanism for representing expected variability in measurement properties (and scores or trait estimates) due to the randomness of selecting informants. Random item parameters in random item IRT can also be modeled as a function of covariates (e.g., type of informant), which promises not only other, more sophisticated ways of examining DIF, but also ways to explain DIF when present (Jin & Wang, 2014; Putka et al., 2011).

Summary

Overall, MTMM, SEM-based CFA, and DIF can be used to help determine empirically (a) whether there are meaningful informant discrepancies; (b) which reporter provides more information about particular symptom items or sets of items in relation to a trait; and (c) whether informant discrepancies are limited to one range of dysfunction, such as the mild

range. In addition, items could be prioritized or symptom lists shortened through consideration of individual item weights in SEM or discrimination parameters in IRT, leading to novel algorithm approaches or the development of short screening measures. Furthermore, finding that particular informants provide better information about particular symptoms might suggest the merit of increasing their report's weight when making diagnostic decisions. To date, these approaches have not been comprehensively utilized in research or clinical settings, but they merit serious consideration.

Upcoming approaches: diagnostic classification, machine learning, and recursive partitioning

There are also several novel and innovative approaches to incorporation of information from multiple informants that merit consideration as their methods develop further. Diagnostic classification models, also known as cognitive diagnosis models and multiple classification latent class models, allow for multivariate classifications of respondents based on multiple postulated latent skills (Rupp & Templin, 2008; Templin & Henson, 2006). They are a special subtype of the latent class model, where items are assumed to reflect certain abilities that are either present or not (e.g., a test comprised of items testing whether or not a respondent has mastered a sequence of increasingly complex rules). In these models, the quality of individual items is proportional to their separation in probability of endorsement between two latent ability and nonability classes. Diagnostic classification models are more restrictive in their assumptions than latent trait and traditional latent class models and do not allow for graded, continuous variability in the abilities they represent. However, in cases where it is reasonable to assume that items or diagnostic criteria reflect sets of skills that have or have not been mastered, diagnostic classification models might streamline the process of identifying presence or absence of a skill or behavior.

With the advance of 'big data,' machine learning is another potential promising avenue that allows for sophisticated computer algorithms to be developed on the basis of the input of large datasets. Such inductive learning systems can incorporate Bayesian priors calculated from large datasets into calculations through the naïve Bayesian classifier, for example (Kononenko, 2001).

Finally, recursive partitioning (e.g., CART, random forest) methods provide a promising way to integrate multi-informant data in actuarial prediction settings (reviewed by Strobl, Malley, & Tutz, 2009). Recursive partitioning methods, loosely speaking, nonparametrically identify combinations of ranges of different variables that optimally predict a given outcome. Applied to informant ratings, they afford a way to actuarially predict outcomes using not only linear combinations of different informant ratings, but interactions and nonlinear combinations of ratings as well. For example, a recursive partitioning method might identify which ranges of self-rated externalizing and parent ratings of externalizing in combination best predict risk of legal involvement among adolescents. Although these methods hold substantial promise, more research is needed to identify how to best develop and select recursive partitioning models, as they can be significantly affected by overfitting.

All of these approaches, while speculative, appear promising and deserve further consideration, but are not yet quite ready for mainstream use, in our opinion.

Example: Use of proposed methods in relation to ADHD

We next provide an example of how some of the more well-validated of these approaches can be applied, using ADHD as an example. As noted above, the current conceptualization of ADHD in the DSM-5 is that a diagnosis requires convergence across raters. That is, children must manifest at least six symptoms in one of the two symptom domains and substantial interference in functioning occurring in two or more settings (e.g., at home and at school; APA, 2013). We report on data drawn from 725 children (55.3% male), 6 to 17 years old (M=10.82; SD=2.33), and their primary caregivers and teachers. The data set includes maternal report on ADHD symptoms via diagnostic interview with a clinician-generated impairment rating based on parent report of impairment, and maternal, paternal, and teacher report on ADHD symptoms available via report on the ADHD-Rating Scale (ADHD-RS). In addition, children completed common laboratory tests of executive function. Families were recruited from the community and then evaluated for study eligibility. Children with ADHD-related problems were over-recruited. 28.5% were ethnic minority, and family income ranged from 0 to $600,000 per year ($M$=67,550.83, SD=47,323.34), as reported by parents. Children came from 426 families; 299 families had two children in the study. All families completed informed consent. Primary analyses utilized maternal, paternal, and teacher report on the ADHD-RS which is rated using a 0 (not at all) to 3 (very much) rating scale, similar to the type of scales that are typically used in clinical situations.

Using this data, a MTMM analysis was conducted using bivariate correlations between maternal, paternal, and teacher ratings of inattention and hyperactivity-impulsivity, as shown in Table 1. First, it should be noted that all ratings of ADHD symptoms, across raters, were significantly associated with one another, consistent with the idea of a common core ADHD construct. Correlations between parent ratings of inattention and hyperactivity-impulsivity were high (r=.67–.70, p<.01). Correlations between parent and teacher ratings of hyperactivity-impulsivity were moderate (r=.43–.47, p<.01), while correlations between parent and teacher ratings of inattention were somewhat lower (r=.28–.33, p<.01). Overall, this suggests moderate convergent validity across raters at the ADHD symptom domain level with increased convergent validity between raters in similar settings (e.g., at home) and for more easily-observable behaviors (e.g., hyperactivity-impulsivity). However, correlations between same-informant ratings of inattention and hyperactivity-impulsivity (r=.63–.66, p<.01) reflected large effects, suggesting substantial shared source variance, a shared evaluative bias across symptom domains, and/or possibly a general ADHD factor (Campbell & Fiske, 1959). The correlation between different-informant ratings of inattention and hyperactivity-impulsivity (e.g., parent-rated inattention and teacher-rated hyperactivity-impulsivity and vice versa) were significant and of moderate to large effect size (r=.48–.55, p<.01), suggesting only limited discriminant validity of the symptom domains. Finally, within-informant correlations of different domains across settings (i.e., parent-rated inattention and parent-rated hyperactivity; r=.63–.66) were higher than between-informant correlations of the same domain (i.e., parent- vs. teacher-rated inattention; r=.28–.43), although the opposite was the case for informants rating in the same setting (i.e., for mothers and fathers; r=.67–.

7). This could very well be consistent with situational differences in child behavior across settings.

Next, structural equation models (SEM) were conducted in Mplus (version 7) using Delta parameterization and weighted least squares means and variance adjusted (WLSMV) estimation (Muthen & Muthen, 2013). In addition, the presence of siblings was accounted for using the clustering feature of Mplus which takes into account the nonindependence of the data when computing test statistics and significance tests. Utilizing mother, father, and teacher ratings of ADHD symptoms, a trifactor CFA model was fit. This model, shown in simplified form in Figure 1, separates general ADHD trait variance from inattentive and hyperactive-impulsive symptom domain variance and informant (i.e., mother, father, and teacher) perspective variance. The initial model exhibited good model fit to the data (Hu & Bentler, 1999; McDonald & Ho, 2002) with an RMSEA of .04 (90% confidence interval=. 034–.038) and CFI of .98, suggesting that -- once the unique perspectives of each informant and variance unique to each individual symptom domain are accounted for--all individual informant reports on individual items contribute significantly to the general ADHD construct; these factor loadings are shown in Table 2. This is consistent with theory of cross-situational convergence of ADHD-related behaviors.

Factor loadings across raters were uniformly high (.3–.8) on the ADHD general factor with parent-rated item loadings slightly higher (.4–.8) than teacher-rated item loadings (.3–.7), suggesting parents may be somewhat better at identifying convergence, or cross-situational manifestation of symptoms across domains. On the inattentive specific factor, factor loadings for the symptoms 'Difficulty sustaining attention' and 'Does not listen' were lower (.11–.33) than other symptom factor loadings across all three raters. For the hyperactive-impulsive specific factor, father-rated symptom items exhibited lower factor loadings than mother- and teacher-rated symptom items. Finally, for the rater-specific factors, while all teacher-rated symptom items exhibited relatively high loadings over .51, maternal and paternal ratings of the hyperactive items 'Leaves seat,' 'Runs or climbs,' and 'On the go' exhibited relatively low loadings of .26 or below.

Threshold parameter estimates for the full model are shown in Table 3. These are akin to difficulty or severity parameters in IRT, in that they represent differences between items in the probability of a given response option being endorsed, independent of the latent trait. There is one threshold parameter for each boundary between one response option and the next (i.e., one for the boundary between responding '0' and '1' [not at all], another for the boundary between '1' and '2', and another for the boundary between '2' and '3' [very much]). The thresholds shown in Table 3 suggest that, for most symptoms, fathers were less likely to rate children high on symptoms given an equivalent value of the latent ADHD trait. Or, in other words, children had to exhibit more extreme levels of problem behavior for fathers to rate them highly on most symptoms, compared to mothers and teachers. An exception to this general rule was that teachers required higher levels of ADHD-related problems, compared to mothers and fathers, to rate children highly on the symptoms: 'Does not listen,' 'Driven by a motor,' and 'Interrupts/intrudes.' Further, all raters required high levels of ADHD problems to rate children highly on the hyperactive symptom of 'Runs/climbs.'

We next tested differential item functioning at the general factor level by constraining the general factor loadings for the same item across informants to be the same (i.e., constrained the loading of inattentive symptom #1 [and all symptoms] on the general factor to be equivalent across mothers, fathers, and teachers). This model fit even better (RMSEA=.027; 90% confidence interval=.024–.029; CFI=.99) than the first model where all loadings were freely estimated. This suggests that all informants are equivalent in their ability to rate the overall ADHD factor after controlling for specific informant perspectives and specific ADHD symptom domains.

We also tested differential item functioning at the specific inattention and hyperactivity-impulsivity factor level by constraining the inattentive and hyperactive-impulsive factor loadings for the same item across informants to be the same (i.e., constrained the loading of inattentive symptom #1 [and all symptoms] on the *inattentive* factor [and hyperactive-impulsive factor] to be equivalent across mothers, fathers, and teachers). This model also fit fairly well (RMSEA=.034; 90% confidence interval=.032–.036; CFI=.98), similar to the first model where all loadings were freely estimated, suggesting that mothers, fathers, and teachers tended to agree in their report of individual items in the specific ADHD symptom domains of inattention and hyperactivity-impulsivity, although perhaps not as much as they agreed in their ratings of the general ADHD construct. Therefore, most informant agreement seems to be at the general latent ADHD factor level.

Then, we tested differential item functioning at the level of thresholds, or informant differences in ratings of severity of individual symptoms. This model also fit fairly well with an RMSEA of .042 (90% confidence interval=.04–.044) and a CFI of .97, suggesting that different raters similarly endorsed levels of symptoms across children.

Finally, we provide preliminary evaluation of the external validity of the factors through examination of associations with executive function and clinician ratings of impairment based on a clinical interview administered to the parent. A common laboratory measure of inhibition, the aspect most commonly associated with ADHD, was significantly correlated the general ADHD factor (r=.34, p<.01), specific inattention (r=.17, p<.01), and specific hyperactivity-impulsivity (r=.14, p<.05), but not with the mother, father, or teacher perspective factors (r=−.01–.09, p>.1). A similar pattern of results was noted for other aspects of executive function. Clinician ratings of impairment were significantly associated with all factors, including the general ADHD factor (r=.61, p<.01), specific inattention factor (r=.31, p<.01), specific hyperactivity-impulsivity factor (r=.10, p<.01), mother perspective factor (r=.32, p<.01), father perspective factor (r=.15, p<.01), and teacher perspective factor (r=.10, p<.01) with most variance accounted for by the general ADHD factor.

Thus, these results suggest that there was substantial convergence across parent and teacher ratings of ADHD based on factor loadings in the trifactor model and the 'g' factor in relation to external validation criteria. This might suggest the utility of an average approach to symptom integration for ADHD, particularly in determination of overall diagnosis, consistent with extant theory of the cross-situational manifestation of ADHD symptoms and recommendations to require cross-situational convergence in informant report of symptoms

for diagnosis. Further, in cases where mothers and teachers endorse severe symptoms and fathers endorse subthreshold symptoms, current results suggest that mother and teacher report may be effectively emphasized. Further validation of such results in relation to other external criteria and in regard to prediction of longitudinal functional outcomes are an important next step in this line of work.

Guidelines for future work

We suggest that future research in this area take a series of structured steps to provide guidelines for integration of multiple informant reports of psychopathology: 1) Develop a clear theoretical model of specific disorders and informant perspectives, 2) Utilize cutting-edge MTMM and SEM models to test such models, using DIF tests, 3) Provide tests of external validation of resulting algorithms, and 4) Make research and/or clinical recommendations. We have stress the importance of advances in development of disorder-specific theories for understanding the presence and nature of informant discrepancies. Once such theories are in place, researchers will be able to use new, sophisticated data analytic approaches to better understand this phenomenon like those modeled here and advocated above. We hope that the approaches used here will inspire researchers with expertise in theory of other disorders (e.g., depression, anxiety, bipolar disorder, conduct disorder, etc.) to conduct their own theory-driven tests of how best to utilize multiple informant reports in relation to specific disorders. A crucial next step is to include independent external validators in models in order to judge the performance of specific informant integration approaches and, perhaps, in order to help anchor or weight various algorithm solutions. Work of the type done by Burt et al. (2005) is illustrative for showing how external criteria could be utilized for validation. She and her colleagues showed differential associations of genetic and both shared and nonshared environmental influences with externalizing behaviors, dependent on the way informant observations are integrated. Her work suggests that shared variance between mother and child report appears most influenced by shared environmental factors, while the variance unique to each informant is influenced by genetic or nonshared environmental factors (Burt et al., 2005). Here again, a well-developed theoretical model of how such informants should be viewed in conceptualization of individual disorders, or group of disorders, is critical. A very important applied step is to develop an approach by which clinicians can take advantage of findings concerning multiple informants. It may well be that technological advances (e.g., statistical algorithms that can be run in the moment on a hand-held device) will facilitate translation from complex data analytic findings to information that is usable and useful for clinicians.

Acknowledgments

References

American Psychiatric Association. Diagnostic and statistical manual of mental disorders—Fifth ed. Washington, D.C: American Psychiatric Association; 2013.

Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. Psychological bulletin. 1987; 101(2):213. [PubMed: 3562706]

Bauer DJ, Howard AL, Baldasaro RE, Curran PJ, Hussong AM, Chassin L, Zucker RA. A trifactor model for integrating ratings across multiple informants. Psychological Methods. 2013; 18(4):475–493. http://doi.org/10.1037/a0032475. [PubMed: 24079932]

Burt AS, McGue MATT, Krueger RF, Iacono WG. Sources of covariation among the child-externalizing disorders: informant effects and the shared environment. Psychological medicine. 2005; 35(08):1133–1144. [PubMed: 16116939]

Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological bulletin. 1959; 56(2):81. [PubMed: 13634291]

Cantwell DP, Lewinsohn PM, Rohde P, Seeley JR. Correspondence between adolescent report and parent report of psychiatric diagnostic data. Journal of the American Academy of Child & Adolescent Psychiatry. 1997; 36(5):610–619. [PubMed: 9136495]

Comer JS, Kendall PC. A symptom-level examination of parent–child agreement in the diagnosis of anxious youths. Journal of the American Academy of Child & Adolescent Psychiatry. 2004; 43(7): 878–886. [PubMed: 15213589]

Connelly BS, Ones DS. An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. Psychological Bulletin. 2010; 136(6):1092–1122. DOI: 10.1037/a0021212 [PubMed: 21038940]

De Ayala, RJ. The Theory and Practice of Item Response Theory. New York: Guilford Press; 2009.

De Los Reyes A. Strategic objectives for improving understanding of informant discrepancies in developmental psychopathology research. Development & Psychopathology. 2013; 25(3):669–682. [PubMed: 23880384]

De Los Reyes A, Augenstein TM, Wang M, Thomas SA, Drabrick DAG, Burgers DE, Rabinowitz J. The validity of the multi-informant approach to assessing child and adolescent mental health. Psychological Bulletin. 2015; 141(4):858–900. [PubMed: 25915035]

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(5):637–652. [PubMed: 19247829]

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. [PubMed: 16060799]

De Los Reyes A, Thomas SA, Goodman KL, Kundey SM. Principles underlying the use of multiple informants' reports. Annual Review of Clinical Psychology. 2013; 9:123–149.

De Los Reyes A, Youngstrom EA, Pabon SC, Youngstrom JK, Feeny NC, Findling RL. Internal consistency and associated characteristics of informant discrepancies in clinical referred youths age 11 to 17 years. Journal of Clinical Child & Adolescent Psychology. 2011; 40(1):36–53. [PubMed: 21229442]

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(4):435–453.

Fergusson DM, Boden JM, Horwood LJ. Situational and generalized conduct problems and later life outcomes: Evidence from a New Zealand birth cohort. Journal of Child Psychology and Psychiatry. 2009; 50(9):1084–1092. [PubMed: 19298467]

Garner WR, Hake HW, Eriksen CW. Operationism and the concept of perception. Psychological Review. 1956; 63:149–159. [PubMed: 13323170]

Geiser C, Burns GL, Servera M. Testing for measurement invariance and latent mean differences across methods: interesting incremental information from multitrait-multimethod studies. Frontiers in Psychology. 2014; 5doi: 10.3389/fpsyg.2014.01216

Grove WM, Meehl PE. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. Psychology, public policy, and law. 1996; 2(2):293–323.

Horton NJ, Fitzmaurice GM. TUTORIAL IN BIOSTATISTICS Regression analysis of multiple source and multiple informant data from complex survey samples. Statistics in medicine. 2004; 23(18): 2911–2933. [PubMed: 15344194]

Hu L, Bentler PM. Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling. 1999; 6(1):1–55.

Jensen PS, Rubio-Stipec M, Canino G, Bird HR, Dulcan MK, Schwab-Stone ME, Lahey BB. Parent and child contributions to diagnosis of mental disorder: are both informants always necessary? Journal of the American Academy of Child & Adolescent Psychiatry. 1999; 38(12):1569–1579. [PubMed: 10596258]

Jensen PS, Salzberg AD, Richters JE, Watanabe HK. Scales, diagnoses, and child psychopathology: I. CBCL and DISC relationships. Journal of the American Academy of Child & Adolescent Psychiatry. 1993; 32(2):397–406. [PubMed: 8444770]

Jin KY, Wang WC. Generalized IRT Models for Extreme Response Style. Educational and Psychological Measurement. 2014; 74(1):116–138. http://doi.org/10.1177/0013164413498876.

Kenny DA, Kashy DA. Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. Psychological Bulletin. 1992; 112(1):165.

Klein DN. Patients' versus informants' reports of personality disorders in predicting 7 1/2-year outcome in outpatients with depressive disorders. Psychological Assessment. 2003; 15(2):216. [PubMed: 12847782]

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(3): 412–432. [PubMed: 16026212]

Klonsky ED, Oltmanns TF. Informant-reports of personality disorder: Relation to self-reports and future research directions. Clinical Psychology: Science and Practice. 2002; 9(3):300–311.

Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artificial Intelligence in medicine. 2001; 23(1):89–109. [PubMed: 11470218]

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. [PubMed: 12944328]

Lahey BB, Applegate B, McBurnett K, Biederman J, Greenhill L, Hynd GW, et al. DSM-IV field trials for Attention Deficit Hyperactivity Disorder in children and adolescents. American Journal of Psychiatry. 1994; 151:1673–1685. [PubMed: 7943460]

Laird RD, De Los Reyes A. Testing informant discrepancies as predictors of early adolescent psychopathology: Why difference scores cannot tell you what you want to know and how polynomial regression may. Journal of abnormal child psychology. 2013; 41(1):1–14. [PubMed: 22773360]

Laird RD, LaFleur LK. Disclosure and monitoring as predictors of mother-adolescent agreement in reports of early adolescent rule-breaking behavior. Journal of Clinical Child and Adolescent Psychology. 2016; 45:188–200. [PubMed: 25470114]

Laird RD, Weems CF. The equivalence of regression models using difference scores and models using separate scores for each informant: Implications for the study of informant discrepancies. Psychological Assessment. 2011; 23:388–397. [PubMed: 21319905]

Lance CE, Noble CL, Scullen SE. A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. Psychological Methods. 2002; 7(2):228–244. DOI: 10.1037/1082-989X.7.2.228 [PubMed: 12090412]

Loeber R, Green SM, Lahey BB, Stouthamer-Loeber M. Optimal informants on childhood disruptive behaviors. Development and Psychopathology. 1989; 1(04):317–337.

Markon KE, Quilty LC, Bagby RM, Krueger RF. The development and psychometric properties of an informant-report form of the Personality Inventory for DSM-5 (PID-5). Assessment. 2013 1073191113486513.

Martel MM, Von Eye A, Nigg JT. Revisiting the latent structure of ADHD: is there a 'g'factor? Journal of Child Psychology and Psychiatry. 2010; 51(8):905–914. [PubMed: 20331490]

McDonald RP, Ho MR. Principles and practice in reporting structural equation analyses. Psychological Methods. 2002; 7(1):64–82. [PubMed: 11928891]

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(4):400–419. [PubMed: 16178682]

Muthen, LK., Muthen, BO. Mplus User's Guide. 4. Los Angeles, CA: Author; 2013.

Olino, TM., Klein, DN. Psychometric Comparison of Self- and Informant-Reports of Personality. Assessment. 2015. http://doi.org/10.1177/1073191114567942

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(1):51–63. [PubMed: 1548394]

Putka DJ, Lance CE, Le H, McCloy RA. A Cautionary Note on Modeling Multitrait-Multirater Data Arising From Ill-Structured Measurement Designs. Organizational Research Methods. 2011; 14(3):503–529. DOI: 10.1177/1094428110362107

Reise SP, Ventura J, Keefe RSE, Baade LE, Gold JM, Green MF, … Bilder R. Bifactor and item response theory analyses of interviewer report scales of cognitive impairment in schizophrenia. Psychological Assessment. 2011; 23(1):245–261. DOI: 10.1037/a0021501 [PubMed: 21381848]

Rupp AA, Templin JL. Unique Characteristics of Diagnostic Classification Models: A Comprehensive Review of the Current State-of-the-Art. Measurement. Interdisciplinary Research & Perspective. 2008; 6(4):219–262. DOI: 10.1080/15366360802490866

Schmitt N, Stults DM. Methodology review: Analysis of multitrait-multimethod matrices. Applied Psychological Measurement. 1986; 10(1):1–22.

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(3):380–411. [PubMed: 16026211]

Solanto MV, Alvir J. Reliability of DSM-IV symptom ratings of ADHD: Implications for DSM-V. Journal of Attention Disorders. 2009; 13(2):107–116. [PubMed: 19372494]

Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests. Psychological Methods. 2009; 14(4):323–348. [PubMed: 19968396]

Tackett JL, Herzhoff K, Reardon KW, Smack AJ. The relevance of informant discrepancies for the assessment of adolescent personality pathology. Clinical Psychology: Science and Practice. 20(4): 378–392.

Templin JL, Henson RA. Measurement of psychological disorders using cognitive diagnosis models. Psychological Methods. 2006; 11(3):287–305. DOI: 10.1037/1082-989X.11.3.287 [PubMed: 16953706]

Toplak ME, Pitch A, Flora DB, Iwenofu L, Ghelani K, Jain U, Tannock R. The unity and diversity of inattention and hyperactivity/impulsivity in ADHD: evidence for a general factor with separable dimensions. Journal of Abnormal Child Psychology. 2009; 37(8):1137–1150. [PubMed: 19562477]

Valo S, Tannock R. Diagnostic instability of DSM-IV ADHD subtypes: Effects of informant source, instrumentation, and methods for combining symptom reports. Journal of Clinical Child & Adolescent Psychology. 2010; 39(6):749–760. [PubMed: 21058123]

van Dulmen MH, Egeland B. Analyzing multiple informant data on child and adolescent behavior problems: Predictive validity and comparison of aggregation procedures. International Journal of Behavioral Development. 2011; 35(1):84–92.

Vazire S. Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. Journal of personality and social psychology. 2010; 98(2):281–300. [PubMed: 20085401]

Vazire S, Mehl MR. Knowing me, knowing you: the accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. Journal of personality and social psychology. 2008; 95(5):1202–1216. [PubMed: 18954202]

Verhulst FC, Ende J. Agreement between parents' reports and adolescents' self-reports of problem behavior. Journal of Child Psychology and Psychiatry. 1992; 33(6):1011–1023. [PubMed: 1400684]

Youngstrom EA, Findling RL, Kogos Youngstrom J, Calabrese JR. Toward an evidence-based assessment of pediatric bipolar disorder. Journal of Clinical Child and Adolescent Psychology. 2005; 34(3):433–448. [PubMed: 16026213]

Widaman KF. Multitrait-multimethod models in aging research. Experimental aging research. 1992; 18(4):185–201. [PubMed: 1459163]

Woehr DJ, Sheehan MK, Bennett W Jr. Assessing Measurement Equivalence Across Rating Sources: A Multitrait-Multirater Approach. Journal of Applied Psychology. 2005; 90(3):592–600. DOI: 10.1037/0021-9010.90.3.592 [PubMed: 15910153]

Key points

- An empirical approach to cross-informant integration in clinical assessment is needed

- Little in the way of solid recommendations is currently provided

- Theoretically-based SEM models can provide useful tests

- ADHD is provided as an applied example of this method

- Important future directions include additional external validation

- New methods of cross-informant integration will advance clinical and research assessment practices
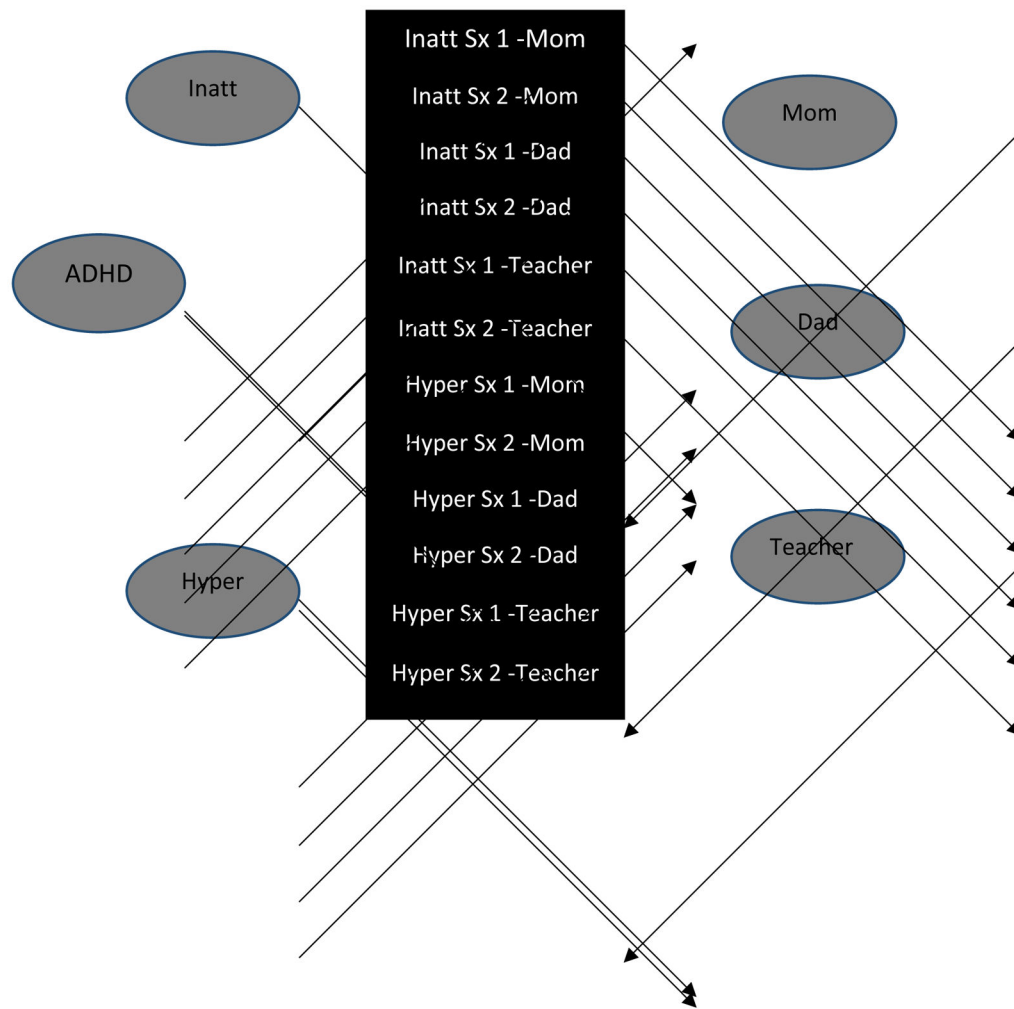
Figure 1.
Simplified Trifactor ADHD Model: General ADHD, Inattention and Hyperactivity Symptom
Domain, and Mother, Father, and Teacher Perspective Factors

Table 1

Multitrait—Multimethod Matrix

Method Measure	Mother-rated inattention	Mother-rated hyperactivity-impulsivity	Father-rated inattention	Father-rated hyperactivity-impulsivity	Teacher-rated inattention	Teacher-rated hyperactivity-impulsivity
Mother-rated inattention	1					
Mother-rated hyperactivity-impulsivity	.66**	1				
Father-rated inattention	**.70****	.48**	1			
Father-rated hyperactivity-impulsivity	.48**	**.67****	.66**	1		
Teacher-rated inattention	**.33****	.47**	**.28****	.48**	1	
Teacher-rated hyperactivity-impulsivity	.54**	**.43****	.55**	**.47****	.63**	1

Note.

**
$p<.01$.

Table 2

ADHD Symptom Loadings in Trifactor Model

	ADHD General Factor	Inattention Factor	Hyperactivity-impulsivity Factor	Mother Factor	Father Factor	Teacher Factor
Maternal report						
Close attention	.50	.49		.50		
Sustain attention	.75	.23		.47		
Listens	.72	.14		.40		
Follow through	.63	.50		.47		
Organizing	.56	.47		.50		
Mental effort	.58	.53		.44		
Loses things	.49	.45		.57		
Distracted	.78	.30		.43		
Forgetful	.61	.42		.50		
Fidgets	.69		.42	.42		
Leaves seat	.77		.43	.20		
Runs or climbs	.68		.42	.19		
Plays quietly	.58		.39	.37		
'On the go'	.60		.46	.26		
Talks	.50		.45	.31		
Blurts out	.47		.45	.53		
Waiting turn	.68		.35	.36		
Interrupts	.58		.42	.47		
Paternal report						
Close attention	.58	.57			.44	
Sustain attention	.66	.33			.31	
Listens	.77	.11			.43	
Follow through	.75	.55			.34	
Organizing	.67	.49			.47	
Mental effort	.49	.57			.25	
Loses things	.48	.48			.48	
Distracted	.55	.38			.35	

	ADHD General Factor	Inattention Factor	Hyperactivity-impulsivity Factor	Mother Factor	Father Factor	Teacher Factor
Forgetful	.68	.46			.42	
Fidgets	.66		.15		.52	
Leaves seat	.77		.26		.23	
Runs or climbs	.75		.36		.18	
Plays quietly	.67		.28		.33	
'On the go'	.49		.40		.25	
Talks	.48		.45		.35	
Blurts out	.55		.28		.47	
Waiting turn	.68		.22		.39	
Interrupts	.57		.37		.50	
Teacher report						
Close attention	.48	.51				.51
Sustain attention	.63	.23				.71
Listens	.62	.15				.55
Follow through	.46	.57				.58
Organizing	.55	.49				.54
Mental effort	.47	.54				.60
Loses things	.56	.47				.54
Distracted	.68	.18				.67
Forgetful	.53	.47				.57
Fidgets	.61		.26			.53
Leaves seat	.50		.30			.67
Runs or climbs	.38		.41			.73
Plays quietly	.47		.38			.66
'On the go'	.39		.45			.68
Talks	.34		.49			.61
Blurts out	.38		.59			.57
Waiting turn	.41		.56			.63
Interrupts	.46		.61			.58

Note. All loadings significant at $p < .05$.

Table 3

ADHD Threshold Parameters in Trifactor Model

	Maternal			Paternal			Teacher		
	b1	b2	b3	b1	b2	b3	b1	b2	b3
Close Attention	−.60	.20	.91	−.58	.56	1.55	−.60	.44	1.05
Sustained Attention	−.20	.55	1.41	−.15	.90	1.68	−.22	.52	1.20
Listens	−.29	.73	1.47	−.24	.85	1.62	.17	1.12	1.75
Follow Through	−.41	.35	.99	−.49	.62	1.35	−.26	.50	1.05
Organization	−.44	.27	.97	−.30	.67	1.42	−.26	.45	1.06
Sustained Mental Effort	−.36	.22	.74	−.30	.46	1.12	−.16	.52	.95
Loses Things	−.37	.37	1.08	−.29	.75	1.43	−.06	.67	1.33
Easily Distracted	−.47	.17	.81	−.38	.38	1.15	−.44	.34	.88
Forgetful	−.31	.40	1.15	−.33	.61	1.39	−.10	.66	1.26
Fidgets	−.01	.58	1.13	.15	.92	1.60	.08	.74	1.24
Leaves Seat	.29	.97	1.50	.40	1.28	1.85	.30	1.03	1.47
Runs/Climbs	.63	1.19	1.81	.60	1.35	2.18	1.10	1.54	2.01
Plays Quietly	.47	1.20	1.73	.30	1.24	1.99	.51	1.16	1.90
'Driven by a Motor'	.28	.77	1.18	.10	.87	1.43	.53	1.02	1.55
Talks a Lot	−.16	.60	1.13	−.21	.73	1.48	−.02	.80	1.41
Blurts	.13	.91	1.45	.08	1.15	1.80	.38	1.13	1.50
Waiting Turn	.19	.92	1.49	.08	.97	1.57	.38	1.02	1.54
Interrupts/Intrudes	−.25	.59	1.26	−.23	.78	1.47	.26	.94	1.48

Note. All significant at *p*<.05.