

Integrating Multi-Informant Reports of Youth Mental Health: A Construct Validation Test of Kraemer and Colleagues' (2003) Satellite Model

Natalie R. Charamut¹, Sarah J. Racz¹, Mo Wang², Andres De Los Reyes^{1*}

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

²Department of Management, University of Florida, Gainesville, FL, USA

*** Correspondence:**

Andres De Los Reyes (adlr@umd.edu)

CITATION

Charamut, N.R., Racz, S.J., Wang, Mo, & De Los Reyes, A. (2022). Integrating multi-informant reports of youth mental health: A construct validation test of Kraemer and Colleagues' (2003) Satellite Model. *Frontiers in Psychology*. Advance online publication. <https://doi.org/10.3389/fpsyg.2022.911629>

FUNDING INFORMATION

Efforts by the fourth author were supported by a grant from the Fulbright U.S. Scholars Program (Fulbright Canada Research Chair in Mental Health). Efforts by the third and fourth authors were supported by a grant from the Institute of Education Sciences (R324A180032; National Center for Special Education Research).

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contribution statement

NC wrote the first draft of the paper and edited the paper. SJR and MW contributed to the data analysis and edited the paper. ADLR conceptualized the problem, designed the study, acquired study funding, organized and executed the data analyses, edited the paper, and supervised NC in writing the first draft of the paper. All authors approved the manuscript and agreed to be accountable for all aspects of the work.

Keywords: Converging Operations, Diverging Operations, Informant Discrepancies, Operations Triad Model, Validity, Satellite Model

Abstract

Accurately assessing youth mental health involves obtaining reports from multiple informants who typically display low levels of correspondence. This low correspondence may reflect *situational specificity*. That is, youth vary as to where they display mental health concerns and informants vary as to where and from what perspective they observe youth. Despite the frequent need to understand and interpret these *informant discrepancies*, no consensus guidelines exist for integrating informants' reports. The path to building these guidelines starts with identifying factors that reliably predict the level and form of these informant discrepancies, and do so for theoretically and empirically relevant reasons. Yet, despite the knowledge of situational specificity, few approaches to integrating multi-informant data are well-equipped to account for these factors in measurement, and those that claim to be well-positioned to do so have undergone little empirical scrutiny. One promising approach was developed roughly 20 years ago by Kraemer and colleagues (2003). Their Satellite Model leverages principal components analysis (PCA) and strategic selection of informants to instantiate situational specificity in measurement, namely components reflecting variance attributable to the *context* in which informants observe behavior (e.g., home/non-home), the *perspective* from which they observe behavior (e.g., self/other), and behavior that manifests across contexts and perspectives (i.e., *trait*). The current study represents the first *construct validation* test of the Satellite Model. A mixed-clinical/community sample of 134 adolescents and their parents completed six parallel surveys of adolescent mental health. Adolescents also participated in a series of simulated social interactions with research personnel trained to act as same-age, unfamiliar peers. A third informant (*unfamiliar untrained observer*) viewed these interactions and completed the same surveys as parents and adolescents. We applied the Satellite Model to each set of surveys and observed high internal consistency estimates for each of the six-item *trait* ($\alpha=.90$), *context* ($\alpha=.84$), and *perspective* ($\alpha=.83$) components. Scores reflecting the *trait*, *context*, and *perspective* components displayed distinct patterns of relations to a battery of criterion variables that varied in the context, perspective, and source of measurement. The Satellite Model instantiates situational specificity in measurement and facilitates unifying conceptual and measurement models of youth mental health.

Contribution to the field

The study represents the first test of the construct validity of a promising approach to integrating multi-informant data in youth mental health. The findings reveal that indices from this approach validly reflect clinically meaningful variations among informants in the contexts and perspectives from which they observe youth mental health.

Funding statement

Efforts by the third and fourth authors were supported by a grant from the Institute of Education Sciences (R324A180032).

Integrating Multi-Informant Reports of Youth Mental Health: A Construct Validation Test of Kraemer and Colleagues' (2003) Satellite Model

Introduction

Like a journalist writing a news story, assessing child and adolescent (i.e., youth) mental health involves attaining a holistic view of the youth undergoing evaluation. Thus, best practices in youth mental health assessment involve soliciting reports from multiple informants (De Los Reyes, 2011, 2013; Hunsley & Mash, 2007). Several decades ago, Achenbach and colleagues (1987) advanced the notion of *situational specificity* to characterize two key components of assessing youth mental health: (a) youth vary as to where they display mental health concerns and (b) the informants from whom assessors solicit reports (e.g., parents, teachers, youth, and their peers) vary in the contexts and perspectives from which they observe the youth. Consistent with this notion, informants' reports of the same youth's mental health tend to display relatively low levels of correspondence (i.e., mean $r = .28$; De Los Reyes et al., 2015). Evidence of these *informant discrepancies* traces back to the 1950s (e.g., Lapouse & Monk, 1958), manifests in assessments conducted globally (De Los Reyes et al., 2019a), and currently represents a literature that totals over 400 published studies (De Los Reyes & Makol, 2022).

Despite the frequent need to understand and interpret informant discrepancies, a persistent problem in measurement of youth mental health is that no consensus guidelines exist for integrating informants' reports (Beidas et al., 2015). Without clarity in these areas, researchers and clinicians lack guidance on how to leverage multiple informants' reports to make clinical decisions (De Los Reyes et al., 2019b; Marsh et al., 2018). *The central thesis of this paper is that the path to building these guidelines starts with identifying factors that reliably predict the level and form of these informant discrepancies, and do so for theoretically and empirically relevant reasons.* Yet, despite the knowledge of situational specificity, few approaches to integrating multi-informant data are well-equipped to account for these factors in measurement, and those that claim to be well-positioned to do so have undergone little empirical scrutiny.

In this paper, we address four aims. First, we review measurement and conceptual models for understanding informant discrepancies and integrating multi-informant data. Second, we use this review as a backdrop for considering an approach to integrating multi-informant data based on the notion of situational specificity (Kraemer et al., 2003). This approach holds promise, but has undergone surprisingly little validation testing. Third, we report findings of the first construct validation test of this integrative approach. Fourth, we describe the research, theoretical, and clinical implications of our construct validation study, and highlight directions for future work.

A key barrier to developing consensus guidelines lies with the most widely used approaches to integrating multi-informant data. As we have articulated elsewhere (for a recent review, see De Los Reyes et al., 2022a), the last 15 years of research very clearly point to a reality of multi-informant assessments conducted in youth mental health: The informant discrepancies produced by this approach often contain information relevant to understanding youth mental health (i.e., *domain-relevant information*). Yet, with few exceptions, available approaches do not account for this reality of multi-informant assessments.

In fact, the most commonly leveraged approaches to integrating multi-informant data stem from conceptual models (i.e., Converging Operations and the Multi-Trait Multi-Method

Matrix [MTMM]; Campbell & Fiske, 1956; Garner et al., 1956) that take an extreme view of informant discrepancies, namely that they carry no value. Given the historically wide use of these models, in the past informant discrepancies were thought to largely reflect *measurement confounds* such as random error and rater biases (De Los Reyes, 2011; Dirks et al., 2012).

Consider examples of the most commonly implemented integrative approaches. For instance, the composite score approach involves computing the sum or average of the informants' scores (see Martel et al., 2021). Its key underlying assumption is that greater agreement among informants' scores signals greater severity of the mental health concerns about which informants provide reports (e.g., De Los Reyes et al., 2016). Conversely, variance unique to any one source (i.e., as reflected by informant discrepancies) reflects error, and thus a core assumption underlying use of composite scoring is that estimates of psychological phenomena should emphasize shared or *common variance* among reports (see Borsboom, 2005; Edgeworth, 1888). This same rationale underlies applications of structural equation modeling (SEM) to integrating or modeling multi-informant data (see also Eid et al., 2008; Watts et al., 2021), combinational algorithms used to integrate data collected within diagnostic interviews (i.e., AND/OR rules; see Offord et al., 1996; Rubio-Stipec et al., 2003; Valo & Tannock, 2010; Youngstrom et al., 2003), and recent applications of measurement invariance techniques to detect informant discrepancies (e.g., Florean et al., 2022; Murray et al., 2021; Olinio et al., 2018; Russell et al., 2016).

To be clear, there is nothing inherently wrong about these integrative approaches. Indeed, all statistical approaches have assumptions underlying their use; the key is understanding whether these usage assumptions “fit” the data conditions to which they will be applied. In this respect, if the most widely used approaches to integrating multi-informant data emphasize common variance, then they are optimized for application to data conditions in which estimates from informants' reports agree to such an extent that they point to the same finding or conclusion in a study (i.e., Converging Operations). Consistent with this notion, when one uses approaches like composite scoring, SEM models that focus on estimating common variance, combinational algorithms, and measurement invariance techniques, one essentially adheres to key assumptions underlying Converging Operations (i.e., unique variance = measurement confounds). Yet, might assumptions underlying use of these integrative approaches be violated if valid data lies not only with instances in which estimates from informants' reports agree, but also when they disagree?

The dominant procedures for integrating multi-informant data treat *all* informant discrepancies as measurement confounds, which begs the question: *Are all informant discrepancies created equally?* The Operations Triad Model (OTM; De Los Reyes et al., 2013) helps us build an evidence base to address this question. The OTM posits that the answer to this question is “no” and as such, provides users with a means for distinguishing at least two forms of informant discrepancies. Within the OTM, *Diverging Operations* denotes scenarios in which informant discrepancies reveal a facet of domain-relevant information that, by definition, cannot be considered a measurement confound. An example of such a scenario might be if parent and teacher reports of a youth's hyperactivity disagree because the youth being assessed behaves differently at home and school, and the informants vary in where they observe the youth (i.e., situational specificity). Conversely, *Compensating Operations* denotes scenarios consistent with the integrative approaches described previously, namely that measurement confounds explain the discrepancies between informants' reports. Examples might include different psychometric properties between informants' reports (e.g., differences in internal consistency or score validity) or the presence of rater biases in one informant's report to a greater degree than in the other informant's report.

Taken together, the OTM delineates concepts (i.e., Converging, Diverging, and Compensating Operations) that researchers can leverage to pose hypotheses about what patterns of agreement and discrepancy between informants' reports of youth mental health might reflect. Further, researchers have leveraged the OTM's concepts to guide controlled tests of these hypotheses. For instance, consistent with the notion of situational specificity, when a parent reports elevated disruptive behavior in their child and the teacher does not, that child is highly likely to display disruptive behavior within parent-child interactions but not interactions between children and non-parental adults (De Los Reyes et al., 2009). In this case, the domain-relevant information revealed by the discrepancies (i.e., contextual variations in disruptive behavior) would be lost if the reports were integrated with a method that emphasizes common variance (e.g., composite scoring or SEM). In fact, a number of investigations point to informant discrepancies containing data that predict such domain-relevant criteria as pulmonary functioning (Al Ghriwati et al., 2018), treatment outcomes (Becker-Haimes et al. 2018; Humphreys et al., 2017; Makol et al., 2019; Zilcha-Mano et al., 2021), suicidal ideation (Augenstein et al., 2022), substance use risk (Lippold et al., 2014), and psychosocial impairments (De Los Reyes et al., 2022b). In sum, OTM-informed studies overwhelmingly indicate that at least some informant discrepancies reflect domain-relevant information.

The findings of OTM-informed studies beg another question: *What are the consequences of applying integrative approaches that only emphasize common variance to data conditions where domain-relevant variance resides in both common and unique variance?* When applied to data conditions that violate the assumption that only common variance matters—that unique variance cannot be domain-relevant—it is logical to hypothesize that using integrative approaches that carry the “only common variance matters” assumption has the consequence of depressing measurement validity (see also De Los Reyes et al., 2022c). Recent work supports such a hypothesis. For instance, integrative approaches that emphasize both common variance and domain-relevant unique variance outperform composite scores in terms of magnitudes of relations to criterion variables (De Los Reyes et al., 2022b) and direct tests of incremental validity (Makol et al., 2020). Further, the most commonly used MTMM-informed structural models cannot distinguish between informant discrepancies that reflect measurement confounds from those that reflect domain-relevant information (Watts et al., 2021). This work points to the need for guidance on approaches to integrating multi-informant data in youth mental health.

OTM-informed research supports the development of guidelines for integrating multi-informant assessments of youth mental health. In particular, it appears that for many mental health domains, integrative approaches ought to optimize use of both common variance (i.e., Converging Operations) *and* domain-relevant unique variance (i.e., Diverging Operations), while minimizing the impact of measurement confounds (i.e., Compensating Operations). This notion begs yet another question: *Which integrative approaches account for both common variance and domain-relevant unique variance?* Roughly 20 years ago, Kraemer and colleagues (2003) proposed the “Satellite Model,” an approach to integrating multi-informant data that addresses the question of optimizing both common variance and domain-relevant unique variance. The authors explain their approach using methods underlying global positioning systems (GPS). Within a GPS, each satellite provides vital information when locating a target in space (e.g., building or a person), but only insofar as their location in space optimizes use of the information they provide. That is, one cannot obtain accurate location data using one well-placed satellite, or even multiple satellites placed at the same set of latitudes and longitudes. Rather, accurate location data comes from *triangulation*, such that each satellite's position varies from the others

based on predictable coordinates. For example, satellites 1 and 2 might be placed at the same latitude but disparate longitudes, whereas satellite 3 might be placed at a disparate latitude relative to 1 and 2, and at a longitude that “sits in between” the other satellites.

Kraemer and colleagues (2003) use satellite placement to explain a key idea: When normatively observing the presence of discrepant information, making sense of patterns of information necessitates *forcing* discrepancies to occur, using factors that reliably predict the discrepancies, and are based on domain-relevant aspects of the phenomena undergoing evaluation. When applied to multi-informant data, the integrative approach involves detecting factors analogous to the latitudes and longitudes used in satellite placement in GPS. In order to implement this approach, Kraemer and colleagues applied principal components analysis (PCA) to synthesize multi-informant data into orthogonal, domain-relevant factors that predictably result in informant discrepancies. Supported by several decades of research on moderators of correspondence between reports from multiple informants (Achenbach et al., 1987, De Los Reyes et al., 2015), the “latitudes and longitudes” in the Satellite Model consist of (a) the *context* in which an informant observes the youth undergoing evaluation and (b) the *perspective* (e.g., self vs. other) through which an informant observes the youth. Users of this approach select informants who vary in their contexts and perspectives, thus allowing for a third component (i.e., *trait*) to reflect *common variance*, namely aspects of youth mental health that generalize across informants’ contexts and perspectives. As such, the Satellite Model optimizes use of both *common variance* (i.e., *trait*) and domain-relevant *unique variance* (i.e., *context* and *perspective*).

As a tool for detecting patterns of common variance (i.e., *trait* component) and domain-relevant unique variance (i.e., *context* and *perspective* components), the Satellite Model optimizes use of both common variance and domain-relevant unique variance among informants’ reports to predict domain-relevant outcomes. Yet, this approach has rarely been subjected to validation testing, and thus two questions warrant consideration. First, to what degree does the approach reveal patterns of common variance and domain-relevant unique variance that manifest across mental health domains (e.g., depressive symptoms, anxiety, psychosocial impairments)? The meta-analytic work on informant discrepancies points to these discrepancies manifesting regardless of domain (Achenbach et al., 1987; De Los Reyes et al., 2015). If so, then perhaps the path to increasing precision in estimating the “latitudes and longitudes” underlying this approach may be facilitated by applying the Satellite Model to a *battery* of multi-informant surveys. A battery of surveys would allow us to create multi-item scales of the *trait*, *context*, and *perspective* components derived from the Satellite Model, and to test their precision using well-established procedures (e.g., estimates of internal consistency; Nunnally & Bernstein, 1994).

Second, even if we discover that we can develop *precise* multi-item estimates of the *trait*, *context*, and *perspective* components derived from the Satellite Model, is it *accurate* to interpret these estimates as reflecting variations in informants’ contexts and perspectives? Indeed, this is a common issue in psychometrics and factor analysis. That is, using procedures like PCA to identify the factor structure of items only marks the first step in understanding variations among scores and the constructs they reflect (DiStefano & Hess, 2005; Messick, 1995; Pett et al., 2003; Strauss & Smith, 2009). Yet, researchers using the Satellite Model have subjectively interpreted scores reflecting the *trait*, *context*, and *perspective* components as if they directly reflect these components, without the underlying *construct validation* tests to support these interpretations. Essentially, users of the Satellite Model have engaged in the *naming fallacy*: “Just because a factor is named does not mean that the hypothetical construct is understood or even correctly labeled” (Kline, 2016, p. 300). In fact, we know of no previous study that has leveraged

construct validation strategies to understand scores taken from the Satellite Model in relation to external criterion variables.

To address the two aforementioned gaps in the literature, we conducted a construct validation test of the Satellite Model with a battery of multi-informant assessments of adolescent social anxiety, consistent with recent work (Makol et al., 2020). Evidence-based assessments of internalizing concerns like social anxiety typically involve collecting reports from both parents and adolescents (Hunsley & Mash, 2007). In fact, evaluations of adolescent social anxiety often rely *exclusively* on parent and adolescent reports (Cannon et al., 2020; De Los Reyes & Makol, 2019). Parents and adolescents often disagree in their reports of adolescent social anxiety, and historically these discrepancies have been interpreted as the adolescent underreporting or downplaying their social anxiety symptoms, perhaps due to a social desirability bias (De Los Reyes et al., 2012a, 2015). Yet, more parsimonious explanations might account for these discrepancies. For example, adolescent social anxiety often manifests in peer interactions, which makes this context an important part of assessment and treatment (Glenn et al., 2019; Hofmann et al., 1999). Importantly, relative to earlier developmental periods, parents have reduced opportunities to observe their adolescent outside of the home, particularly in terms of peer interactions (Smetana, 2008). Indeed, recent work finds that parent reports of adolescent social anxiety often fail to predict adolescents' self-reported reactions to peer interactions (Deros et al., 2018). In this sense, parents' lack of observation of their adolescent's interactions with peers likely contributes to the discrepancies between parent and adolescent reports. When viewed through the lens of the Satellite Model, these findings indicate that, although parent and adolescent reports provide some "coverage" of adolescent social anxiety, effective triangulation requires an additional "satellite": A third informant who simulates how a peer might observe the adolescent in a non-home context.

In an effort to collect survey data from this non-home informant, we recently developed the Unfamiliar Peer Paradigm (Cannon et al., 2020), a counterbalanced set of social interaction tasks designed to simulate adolescents' social interactions with same-age, unfamiliar peers. Within this paradigm, adolescents interact with trained research personnel who display a youthful appearance consistent with the adolescent's age (i.e., *peer confederates*). Adolescents interact with these peer confederates over a 20-minute period. Following this period, unfamiliar untrained observers (UOs) review video recordings of these interactions, and make reports using parallel versions of the measures completed by parents and adolescents. In essence, these UOs comprise the third satellite involved in implementing the Satellite Model. In fact, recent work finds that UOs' reports of adolescents' covert avoidance behaviors (i.e., safety behaviors) relate to both adolescents' self-reports of these behaviors and adolescents' social skills as displayed within the Unfamiliar Peer Paradigm (Rezeppa et al., 2021).

Building off of recent work (Makol et al., 2020; Rezeppa et al., 2021), the current study provides the first construct validation test of the Satellite Model for integrating scores taken from a battery of three informants' reports (i.e., parents, adolescents, and UOs). We leveraged a battery of six survey measures of adolescent mental health, and ran six independent PCAs in line with the Satellite Model. On these six sets of survey measures, adolescents, parents, and UOs made reports about domains relevant to understanding adolescent social anxiety. Specifically, when adolescents experience social anxiety, they often also experience concerns with other internalizing domains, such as depression (see Epkins & Heckler, 2011). Further, adolescents who experience social anxiety often also experience psychosocial impairments stemming from their anxiety, along with several anxiety-relevant processes including fears of negative

evaluation, and as mentioned previously, avoidance behaviors (see also De Los Reyes & Makol, 2019; De Los Reyes et al., 2019c; Karp et al., 2018; Qasmieh et al., 2018). Capitalizing on these common elements of adolescent social anxiety may optimize the precision of components assessed via the Satellite Model. That is, prior validation tests of the Satellite Model estimated components (i.e., *trait*, *context*, *perspective*) using only a single set of multi-informant surveys (i.e., all reports about one domain; see Makol et al., 2020). In essence, this approach results in single-item measures of the component scores derived from the PCA. Importantly, a long line of research indicates that single-item measures display relatively weak psychometric properties, compared to multi-item measures (see Nunnally & Bernstein, 1994).

What if applications of the Satellite Model capitalized on a key observation made in informant discrepancies research? Specifically, within any two informants' reports of the same youth, the *structure* of the discrepancies operates quite similarly across domains. That is, when examining informant discrepancies across domains rated within a given set of informants (i.e., the parent, teacher, and youth reports of a youth client's mental health), the patterns of reports (e.g., parent > teacher; youth < parent) tend to generalize, even across distinct domains (e.g., internalizing vs. externalizing; De Los Reyes et al., 2016; Lerner et al., 2017; Makol et al., 2019, 2021). In these respects, we expect *trait*, *context*, and *perspective* scores from multi-informant assessments to display high levels of internal consistency across rated domains. We also expect that aggregating multi-domain *trait*, *context*, and *perspective* scores will optimize measurement precision.

A construct validation test of the Satellite Model also involves selecting a battery of *criterion variables* to test the domain-relevance of scores reflecting the model's components. To return to the GPS analogy, if we treat informants' reports as satellites that are strategically positioned in disparate points in space, then we must also think of the criterion variables in this way. The criterion variables must also display "satellite positions" that systematically vary. They ought to vary both from each other and in line with the *trait*, *context*, and *perspective* components used to synthesize the adolescent, parent, and UO reports. In principle, this approach involves interpreting relations between the *trait*, *context*, and *perspective* components and criterion variables, based on the contexts and perspectives used to create the criterion variables. In practice, this approach involves leveraging criterion variables that tap multiple social contexts, perspectives, and measurement modalities. These criterion variables must also be domain-relevant or pertinent to understanding adolescent social anxiety.

Consequently, our construct validation test involved selecting three sets of domains and corresponding information sources that traversed home and non-home contexts, as well as the perspective of the rater. First, we leveraged a set of trained independent raters who observed adolescents within the Unfamiliar Peer Paradigm and rated their anxiety and social skills during the paradigm. In this way, we relied on a rater whose *perspective* was "neutral" or independent from that of the informants used to estimate the Satellite Model components (i.e., adolescents, parents, and UOs). Further, the trained observers made ratings of adolescents' behavior as it manifests in a non-home context. Second, family conflict is both ubiquitous to the adolescent period (see Adams & Laursen, 2001; Smetana & Gaines, 1999), and commonly manifests among adolescents who experience internalizing concerns (for a review, see Epkins & Heckler, 2001). Thus, we relied on parents' reports of adolescent-parent conflict to collect information about a home-based psychosocial domain, rated from an observer perspective. Third, adolescents who experience social anxiety often also experience relatively high resting arousal (e.g., Monk et al., 2001; Thomas et al., 2012a). By construction, resting arousal is both an internal experience and

occurs in a “neutral” setting, absent the environmental stimuli that typify adolescents’ daily home and non-home contexts. Thus, we relied on adolescents’ self-reports of resting arousal to collect information about a domain that is untethered from social context, and rated from a self-perspective. Collectively, this battery of criterion variables allowed us to test scores designed to reflect each of the Satellite Model’s *trait*, *context*, and *perspective* components.

We applied the Satellite Model to social anxiety assessments in a mixed-clinical/community sample of adolescents. Specifically, we extended the findings of Makol and colleagues (2020), who tested the approach using a single set of multi-informant reports of adolescent social anxiety. We addressed four aims. First, we examined levels of correspondence among parent, adolescent, and UO reports. As with prior work (Achenbach et al., 1987; De Los Reyes et al., 2015), we expected to observe low-to-moderate levels of correspondence across domains. Second, we scored parent-adolescent-UO triads derived from six sets of multi-informant assessments using the Satellite Model. Third, we tested the internal consistencies of the aggregated or “total scores” for both the common variance component (i.e., total *trait* score) as well as the unique variance components (i.e., total *context* and *perspective* scores). We expected these total scores to display acceptable levels of internal consistency.

Fourth, we tested the bivariate and unique relations of the aggregated *trait*, *context*, and *perspective* component scores and a series of criterion variables that, collectively, comprised a construct validation test of these scores. In prior work (Makol et al., 2020), a *trait* score based solely on three reports of adolescent social anxiety demonstrated criterion-related validity when predicting observed anxiety. Theoretically, the Satellite Model optimizes both common variance (i.e., *trait* score) and domain-relevant unique variance in prediction, with two components that capture distinct facets of this unique variance (i.e., *context* vs. *perspective*). As such, we expected the Satellite Model scores to differentially relate to our battery of criterion variables, depending on the context and perspective the criterion variable was designed to reflect.

Specifically, we expected independent observers’ ratings of adolescent social anxiety and social skills to uniquely relate to scores reflecting the *trait* and *context* components. The basis of this prediction lies in two aspects of our independent observers’ ratings. First, prior work indicates that all three informants’ reports (i.e., parents, adolescents, and UOs) each relate to independent observers’ ratings (see Glenn et al., 2019; Makol et al., 2020; Rezeppa et al., 2021). In this respect, relations between these reports and ratings of observed behavior generalize across informants’ contexts and perspectives (i.e., *trait* score). However, the Unfamiliar Peer Paradigm from which independent observers base their ratings nonetheless is based exclusively on behaviors displayed in non-home contexts (i.e., unfamiliar peer interactions). Thus, we would also expect the *context* component to uniquely relate to observers’ ratings. Conversely, we would not expect a link between the *perspective* component and independent observers’ ratings, given that independent observers’ perspectives, by construction, are not represented among the perspectives of the three informants.

In contrast to our first set of predictions regarding independent observers’ ratings, we expected two different patterns of relations between the *trait*, *context*, and *perspective* scores and criterion variables designed to capture aspects of the *context* and *perspective* components. Specifically, we expected parents’ reports of adolescent-parent conflict to uniquely relate to scores reflecting the *context* and *perspective* components, because (a) like behaviors displayed in the Unfamiliar Peer Paradigm, adolescent-parent conflict manifests in a specific context (i.e., the home) and (b) we relied on parent reports (i.e., a unique, observer perspective) to provide reports about conflict. Further, we expected adolescents’ self-reported resting arousal to uniquely relate

to scores reflecting the *perspective* component. This is because (a) resting arousal reflects a process that is untethered to contextual factors (i.e., no relation to the *context* component) and (b) like parent-reported conflict, we relied on a single unique perspective (i.e., adolescent self-report) to assess resting arousal.

Materials and Methods

Participants were 134 adolescents aged 14-15 years old ($M = 14.5$ years; $SD = 0.5$) and their parents who were recruited as part of a larger study (e.g., Cannon et al., 2020; Deros et al., 2018; Glenn et al., 2019). In order to participate in the study, parent-child dyads had to meet the following inclusion criteria: (a) speak English, (b) have a 14–15-year-old adolescent living at home, and (c) understand the consent and assent process. Within this sample, 89 adolescents identified as female and 45 identified as male. Parents identified their adolescent's race/ethnicity (African American or Black: 53%; White, Caucasian American, or European: 34%; Asian American or Asian: 5%; Hispanic or Latino/a (Spanish): 10%; American Indian: 0.7%; or "Other": 7%). The parents could select multiple racial/ethnic backgrounds, leading to these rates totaling over 100%. Parents reported weekly household income in increments of \$100 (e.g., \$101-\$200 per week) with the following breakdown: 26% of parents earned \$500 or less per week, 22% earned between \$501 and \$900 per week, and 51% earned more than \$901 weekly. Parents reported marital status with 50% currently married, 21% never married, 16% divorced, 8% separated, and 0.7% widowed. Parents also reported their highest level of education, with 3% less than high school or equivalent, 14% a high school diploma or equivalent (i.e., GED;), 17% some college, 10% an associate's or vocational degree, 19% a bachelor's degree, 23% a master's degree, and 13% an advanced degree.

The Institutional Review Board of the large mid-Atlantic university where the study was conducted approved the procedures of the study prior to administration. Participants were recruited from Maryland, Washington D.C., and Northern Virginia via public advertisements. We recruited participants through a variety of methods including advertisements online (e.g., Craigslist, the laboratory website, Facebook, and Google Ads), on public transportation (e.g., buses, Metro rail, and Metro stations), and in local spaces (e.g., flyers posted in the community including bulletin boards and community listservs, cards handed out during campus events). Recruitment also took place at the offices of local health professionals (i.e., doctor's offices, clinics, and hospitals) who serve the targeted population.

Two different advertisements were used to recruit participants, with one depicting a no-cost screening clinical assessment for evaluation of adolescent social anxiety (i.e., clinic-referred adolescents; $n = 45$) and the other depicting a study assessing parent-child interactions (i.e., community control adolescents; $n = 89$). Both groups completed the same assessments and tasks described below during an in-person laboratory visit. Following the visit, parents in the clinic-referred group received feedback on their adolescent's functioning and referrals for treatment, whereas those in the community control group did not receive feedback/referrals.

For this study, we used an analytic approach that pooled the two groups into one sample. By combining these two groups, we capitalized on key features of adolescent mental health concerns, namely that they dimensionally vary in the general population (i.e., fewer numbers of adolescents displaying concerns relative to those not displaying concerns, and scores ranging from relatively low concerns to relatively high concerns). This is an approach we have taken in multiple studies leveraging this same sample (e.g., Botkin et al., 2021; Greenberg & De Los Reyes, 2022; Okuno et al., 2021), including the study our current study seeks to extend (Makol et

al., 2020). Further, this approach is consistent with both current initiatives focused on dimensional models of psychopathology (e.g., Insel et al., 2010), and prior work indicating enhanced reliability and validity for dimensional approaches to measuring and examining psychopathology, relative to discrete approaches (e.g., testing aims separately within subgroups; Markon et al., 2011). Importantly, prior work indicates this approach results in clinic-referred and community control groups that display comparable demographic characteristics, thus further justifying use of this approach (see Cannon et al., 2020; Makol et al., 2020). Demographic data for these groups are available upon request from the corresponding author.

Parents completed an initial phone screen with the laboratory staff to assess if they and their adolescent met our inclusion criteria. If they met the criteria, we scheduled them to complete assessments in the laboratory. For the in-person assessment, research personnel described the study and provided parental consent and adolescent assent forms to review and sign. Following consent/assent, adolescents and parents completed parallel sets of survey measures independently on computers in counterbalanced order using the Qualtrics survey platform. Additionally, adolescents completed three counterbalanced social interactions tasks with study personnel trained to interact with the adolescents as unfamiliar peer confederates (see Cannon et al., 2020). Upon completing the study tasks, families received a total of \$100 in monetary compensation (\$50 for the parent and \$50 for the adolescent).

Instruments

Unfamiliar Peer Paradigm

Adolescents interacted with unfamiliar peer confederates in the Unfamiliar Peer Paradigm (Cannon et al., 2020). The Unfamiliar Peer Paradigm is a series of counterbalanced social interaction tasks with trained, gender-matched research assistants designed to be reflective of interactions with same-age, unfamiliar peers. Peer confederates had no prior contact with the adolescent with whom they interacted, and we masked peer confederates to adolescents' referral status and all other clinical information related to adolescent participants. The tasks included a series of structured, dyadic role-plays between adolescents and peer confederates (i.e., Simulated Social Interaction Test [SSIT]), an unstructured dyadic conversation between adolescents and peer confederates designed to simulate the first day of class (i.e., Unstructured Conversation Task [UCT]), and a public speaking task in which adolescents spoke about a series of predetermined social issues (i.e., Impromptu Speech Task [IST]). Each of these tasks have been described at length in prior published work (e.g., Deros et al., 2018; Glenn et al., 2019; Makol et al., 2020), and Cannon and colleagues (2020) provided an overview of the overall paradigm and empirical support for its use. Additionally, scripted procedures for the Unfamiliar Peer Paradigm's tasks exist on the Open Science Framework Platform (De Los Reyes, 2020).

UUOs' Reports of Adolescent Mental Health. Using archival videos of the adolescent's participation in the Unfamiliar Peer Paradigm, we randomly assigned UUOs to view up to five recordings of the social interaction tasks. After viewing the recordings, UUOs made survey reports about each adolescent using the battery of six survey measures described below. Importantly, UUOs received no training on how to make these survey reports. In this respect, they received measure instructions akin to the parents and adolescents involved in the study. We masked UUOs to adolescents' referral status and all other clinical information. In online supplementary material, we report the demographic characteristics of the UUOs who completed reports.

Satellite Model Survey Battery

Adolescents, parents, and UOs completed survey measures across a battery of psychosocial domains. All informants completed the measures from their own perspective on reporting about the adolescent. That is, all survey measures included parallel item content for all informants, with only minor word modifications to fit their perspective (e.g., “I” for adolescent vs. “My child” for parent vs. “The participant” for UO). Adolescents and parents completed survey measures immediately following completion of consent/assent forms and before administration of the Unfamiliar Peer Paradigm. UOs completed these same measures based on observations of the adolescents’ behavior during video recordings of the Unfamiliar Peer Paradigm. Extensive descriptions of each of the measures in our battery (i.e., psychometric properties, example items, response options) can be found in prior work (see Botkin et al., 2021; De Los Reyes et al., 2019c; Deros et al., 2018; Karp et al., 2018; Qasmieh et al., 2018; Rausch et al., 2017; Rezeppa et al., 2021).

Social Interaction Anxiety Scale. The Social Interaction Anxiety Scale (SIAS; Mattick & Clarke, 1998) is a 20-item measure of social anxiety displayed during direct social interaction. Informants made responses on a 5-point Likert scale ranging from “0” to “4” with higher scores indicating higher levels of social anxiety. Parents’ and adolescents’ SIAS reports display high levels of internal consistency ($\alpha > .90$) and distinguish adolescents on referral status (Deros et al., 2018).

Social Phobia Scale. The Social Phobia Scale (SPS; Mattick & Clarke, 1998) is a 20-item scale measuring concerns related to social anxiety regarding everyday behaviors. Each item is rated on a Likert scale that ranges from “0” to “4” with higher scores indicating higher levels of social anxiety. Similar to the SIAS, parents’ and adolescents’ SPS reports display high levels of internal consistency ($\alpha > .90$) and distinguish adolescents on referral status (Deros et al., 2018).

Brief Fear of Negative Evaluation Scale. The Brief Fear of Negative Evaluation Scale (BFNE; Leary, 1983) measures fears related to negative evaluation from other individuals. The BFNE is a 12-item measure rated on a 5-point Likert scale ranging from “1” to “5” with higher scores indicating higher evaluative fears. Parents’ and adolescents’ BFNE reports display high levels of internal consistency ($\alpha > .80$) and distinguish adolescents on referral status (Karp et al., 2018; Szollos et al., 2019).

Subtle Avoidance Frequency Examination. The Subtle Avoidance Frequency Examination (SAFE; Cuming et al., 2009) is a 32-item measure where each item describes a safety behavior that could be employed during a social interaction. Informants indicate the frequency the different safety behaviors on a 5-point scale ranging from “1” to “5” with higher scores indicating higher levels of safety behaviors. Informants’ reports of adolescents on the SAFE display high levels of internal consistency ($\alpha > .80$) and distinguish adolescents’ on referral status (Qasmieh et al., 2018; Rezeppa et al., 2021; Thomas et al., 2012b).

Work and Social Adjustment Scales for Youth. The Work and Social Adjustment Scales for Youth (WSASY; De Los Reyes et al., 2019c) assesses adolescents’ psychosocial impairments. It contains 5 items assessing the adolescent’s behavior without mention of mental health concerns or status (e.g., “*Because of the ways I think, feel or behave, my ability to do well in school is impaired.*”). Severity of impairment is indicated using a Likert scale from “0” to “8” with higher scores indicating greater levels of impairment. Parents’ and adolescents’ WSASY reports display high levels of internal consistency ($\alpha > .80$) and distinguish adolescents on the number of peer-related impairments (De Los Reyes et al., 2019c).

Beck Depression Inventory-II. The Beck Depression Inventory-II (BDI-II; Beck et al., 1996) is a widely used measure of depressive symptoms. The measure was originally designed for use with participants aged 13 years and older, and recent work supports its psychometric properties when administered to adolescents (e.g., Glenn et al., 2019; Qasmieh et al., 2018; Rausch et al., 2017). Respondents rate items describing depressive symptoms (e.g., sadness, loss of interest, feelings of guilt) on a 4-point scale with higher scores indicating greater depressive symptoms. As in prior work (e.g., Deros et al., 2018; Rausch et al., 2017), we excluded two items (9 and 21) which assess for suicidality and loss of interest in sex, as parents often decline to consent to their children reporting on these items due to their mature content. Despite excluding these items, sample internal consistency estimates remained high (Table 1).

Criterion Measures for Construct Validation

Independent Observers' Ratings of Adolescents' Social Anxiety and Social Skills.

We leveraged behavioral ratings from trained independent observers to assess adolescents' social anxiety and social skills within the Unfamiliar Peer Paradigm. By construction, these trained independent observers differed from each of the survey informants in two key ways. First, unlike parents, our trained independent observers made ratings that were based specifically on behaviors displayed in a non-home context. Second, although our trained independent observers displayed some overlap in their context of observation with adolescents, and complete contextual overlap with UOs, they made ratings with the benefit of training. In this respect, the perspective from which our trained independent observers made ratings differed from all three informants. Thus, these characteristics of trained independent observers comprised one facet of our larger construct validation test of the Satellite Model.

The trained independent observers consisted of undergraduate and post-baccalaureate research assistants who did not participate in any of the social interaction tasks as a peer confederate and did not complete survey reports as a UO. As with UOs, we masked independent observers to adolescents' referral status and all other clinical information. We provided extensive information on coder training and characteristics in online supplementary material, as well as in prior work (Botkin et al., 2021; Cannon et al., 2020; Glenn et al., 2019).

Independent observers made global ratings of each adolescent's social anxiety and social skills using an extensively validated behavioral coding scheme (e.g., Glenn et al. 2019). For each domain, independent observers based their ratings on observations of the SSIT (five ratings), UCT (one rating), and IST (one rating). Independent observers made social anxiety ratings on a 5-point scale ranging from 1 (Animated) to 5 (Severe anxiety), where higher scores indicated greater social anxiety. Further, independent observers made social skill ratings on a 5-point scale ranging from 1 (*Not effective at all*) to 5 (*Very effective*), where higher scores indicated greater social skills. For each adolescent, a pair of coders rated their social anxiety and social skills, with ratings displaying ICCs of .75 and .81, respectively ("excellent" range per Cicchetti, 1994).

For each adolescent, we computed composite scores for each of the seven task ratings (5 SSIT, 1 UCT, 1 IST) by taking an average of the pair of the independent observers' ratings. Although we computed composite scores for all 134 adolescents, some were missing data on one task rating (e.g., one of the five SSIT ratings), whereas three adolescents declined to give a speech for the IST. Consistent with prior work (Makol et al., 2020) and to reduce Type 1 Error, we created composite mean scores for all seven social anxiety ratings ($M = 3.09$, $SD = 0.82$) and seven social skills ratings ($M = 3.51$, $SD = 0.89$). For adolescents for whom we were missing data on these tasks, we computed their composite scores based on the six ratings we had

available for them. In terms of psychometric support, independent observers' ratings relate to well-established survey measures of adolescent social anxiety and related processes (e.g., safety behaviors, fears of evaluation, psychosocial impairments) and distinguish adolescents on referral status (e.g., Botkin et al., 2021; Cannon et al., 2020; Glenn et al., 2019; Rezeppa et al., 2021).

Parent-Reported Adolescent-Parent Conflict. To assess parent-adolescent conflict, parents completed the Issues Checklist (IC; Prinz et al., 1979). For the purposes of our study, parent-reported conflict data using the IC allowed us to measure a psychosocial domain that (a) is based specifically in the home context, (b) reflects external events (i.e., conflict), and (c) uses the lived experience of an informant whose report comes from an observer perspective. Thus, these characteristics of parent reports on the IC comprised the second facet of our construct validation test of the Satellite Model. On the IC, parents report on topics of disagreement within the past 4 weeks. We modified the IC for the purposes of time (i.e., reduce participant burden) and to assess ranges of conflict related to topics about which parents and adolescents typically encounter at home (e.g., chores, homework, and friends), as consistent with prior work (e.g., Adams & Laursen, 2001; De Los Reyes et al., 2012b; Ehrlich et al., 2011; Smetana & Gaines, 1999; Treutler & Epkins, 2003). Specifically, our modified checklist included 16 of the 44 topics listed on the original IC. A list of the 16 IC topics we assessed is available from the corresponding author. We also modified the response format so that parents could rate conflict about each topic using a 5-point Likert-type scale ranging from 1 (we do not disagree) to 5 (we disagree much). Parents completed the checklist with regard to conflicts between themselves and the adolescent with whom they participated in the study, and vice versa for the adolescent. For this study, we calculated total scores by summing the scores across the 16 items, with possible total scores ranging from 16 to 80 ($M = 33.66$, $SD = 12.01$). The psychometric properties of the IC used in this study and evidence of its reliability and validity have previously been reported (De Los Reyes et al., 2012b; Ehrlich et al., 2011; Rausch et al., 2017).

Adolescent Self-Reported Resting Arousal. Adolescents reported self-perceived levels of internal arousal using the Self-Assessment Manikin (SAM; Lang, 1980). The SAM is a 5-level pictorial scale of affect ranging from 1 (close-eyed/relaxed image) to 5 (wide-eyed/nervous image). Adolescents completed a rating of their *resting* arousal at a baseline period before administration of the Unfamiliar Peer Paradigm ($M = 1.55$; $SD = 0.62$). In this way, we could collect an arousal rating based on a psychological process (resting state) that is (a) stripped of all contextual information, (b) based on an internal process and thus (c) based on the lived experiences of an informant whose report comes from a self-perspective. These characteristics of adolescent self-reports on the SAM comprised the third facet of our construct validation test of the Satellite Model.

Data Analyses

Preliminary Analyses

We followed a multi-step plan for addressing our aims. First, each of our measures consisted of either multi-item surveys of unidimensional constructs or ratings of adolescent behavior for which we calculated composite scores (i.e., of two independent observers' ratings for each adolescent). Thus, consistent with prior work using these measures (e.g., Deros et al., 2018; Glenn et al., 2019; Qasmieh et al., 2018; Thomas et al., 2012b), and to produce estimates to compare against prior work, we assessed the reliability of scores taken from these measures by calculating estimates of either internal consistency (Cronbach's α for survey measures) or inter-rater reliability ($ICCs$ for independent observers' ratings). We interpreted these calculations

relative to conventions for α (e.g., Nunnally & Bernstein, 1994) and ICCs (e.g., Cicchetti, 1994). We then computed means and standard deviations for all continuous measures, and calculated statistics for skewness and kurtosis to determine if our data met assumptions for our planned parametric analyses (i.e., skewness/kurtosis in range of ± 2.0 ; Tabachnick & Fidell, 2001).

Cross-Informant Correspondence

Second, to estimate cross-informant correspondence on survey reports of adolescent mental health, we computed Pearson r correlations among adolescent, parent, and UUO reports on parallel measures (e.g., adolescent-parent, adolescent-UUO, and parent-UUO correlations of SIAS reports).

Scoring Multi-Informant Assessment Battery of Adolescent Mental Health

Third, to prepare multi-informant assessments for the PCA of the Satellite Model components, we scored adolescent, parent, and UUO reports on parallel measures (e.g., the three reports collected on the SPS) using the approach described by Kraemer and colleagues (2003). As mentioned previously, we used a set of informants who collectively varied in their contexts and perspectives, with (a) informants observing from a home-based, observer perspective (parents); (b) informants observing from a non-home-based, observer perspective (UUOs); and (c) informants observing from a self-perspective based on a mix of home and non-home contexts. As such, we expected our PCA to include a *trait* score component in which all informants' reports load strongly and in the same direction. We also expected our PCA to reveal a *context* score (i.e., informants from different contexts load in opposite directions) as well as a *perspective* score (i.e., self-reports load in the opposite direction of observer informants' reports). Consistent with Kraemer and colleagues (2003) and recent work by Makol and colleagues (2020), we conducted six unrotated PCAs, one for each of the parallel measures in our multi-informant battery (i.e., SIAS, SPS, BFNE, SAFE, WSASY, and BDI-II). Each of these PCAs essentially consisted of three "items," namely the adolescent, parent, and UUO reports on the same survey measure. In this respect, for each PCA, our subject-to-item ratio (i.e., $134/3 = 44.67:1$) was well above the typical subject-to-item ratios deemed "large" within PCA modeling contexts (e.g., 20:1; see Osborne & Costello, 2004). Within these six unrotated PCAs, we set the number of components to be extracted to three. We examined principal component weights for each informant's report to determine whether we identified the *trait*, *context*, and *perspective* scores described previously.

Internal Consistency for the Satellite Model Components and Creation of Total Scores

Fourth, we computed α statistics for each six-item set of component scores (i.e., one α per component domain). In preparation for our construct validation test of the Satellite Model, we computed three total scores, one for each Satellite Model component (e.g., total summation of the six *Context* score items).

Relations between the Satellite Model Components and Criterion Variables

Fifth, to test the criterion-related validity of scores taken from the Satellite Model, we computed a series of Pearson r correlations. The correlations estimated relations between the Satellite Model scores and criterion variables. For any one set of bivariate tests, if two or more components bore a relation to the criterion variable, we wanted to ensure that each component *uniquely* related to the criterion variable. Thus, we constructed linear regression models to

estimate unique relations between Satellite Model scores and that criterion variable. In these models, we entered the Satellite Model scores in separate steps, whereby we entered the common variance estimate (i.e., *trait* score) in the first step. We entered any unique variance estimates (i.e., *context* and/or *perspective* scores) separately, in subsequent steps.

Results

Preliminary Analyses

To determine if any of our study variables deviated significantly from normality (i.e., skewness and kurtosis), we conducted a descriptive analysis of adolescent, parent, and UUO responses to all surveys, as well as scores for all of our criterion variables. With one exception, the data met basic assumptions of parametric statistical tests (skewness/kurtosis in range of ± 2.0). Specifically, all three informants' reports on the BDI-II displayed significant skewness and/or kurtosis. We addressed these concerns by applying a square root transformation to all BDI-II reports, which brought them all underneath the thresholds reported previously. All analyses reported below use these transformed scores. Table 1 displays the means and standard deviations for informants' reports on all survey measures. Table 1 also displays α estimates for all informants' survey reports, which all displayed acceptable levels ($\alpha > .08$; Nunnally & Bernstein, 1994).

Cross-Informant Correspondence

In Table 2, we report bivariate correlations among adolescent, parent, and UUO reports on the six survey measures. Supporting previous work, we observed low-to-moderate correlations among informants' reports (Achenbach et al., 1987; De Los Reyes, et al., 2015).

Scoring Multi-Informant Assessment Battery of Adolescent Mental Health

In Table 3, we report the eigenvalues and component loadings for the PCAs of the six sets of adolescent, parent, and UUO survey reports. Consistent with prior work (e.g., Kraemer et al., 2003; Makol et al., 2020), these PCA models each revealed loadings consistent with the *trait*, *context*, and *perspective* components as described previously. Specifically, all informants' reports loaded positively onto the *trait* component, informants' reports from different contexts (i.e., parent vs. UUO) loaded onto the *context* component in opposite directions, and adolescent self-reports loaded onto the *perspective* component in a direction opposite of the loadings observed from the two observer informants (i.e., parent and UUO).

Internal Consistency for the Satellite Model Components and Creation of Total Scores

We computed α estimates for the six-item *trait*, *context*, and *perspective* scales. The six-item *trait* scale displayed an α of .90 and a mean inter-item correlation of .61. The six-item *context* scale displayed an α of .84 and a mean inter-item correlation of .47. The six-item *perspective* scale displayed an α of .83 and a mean inter-item correlation of .44. Thus, each scale displayed high internal consistency (i.e., $\alpha > .80$; see Nunnally & Bernstein, 1994), particularly for short, six-item scales (see also Youngstrom et al., 2019). Based on these findings, we computed total scores for each of the Satellite Model component scales by summing up each of the six-item scales for *trait* ($M = 0$; $SD = 4.92$; minimum value = -8.17; maximum value = 15.55), *context* ($M = 0$; $SD = 4.50$; minimum value = -13.11; maximum value = 11.16), and *perspective* ($M = 0$; $SD = 4.40$; minimum value = -9.33; maximum value = 10.34). All means for components derived from PCA are zero because of their standardization.

Demonstrating the Construct Validity of Scores Reflecting the Satellite Model Components

Using the *trait*, *context*, and *perspective* total scores described previously, we conducted our construct validation test of the Satellite Model. In Table 4, we report a summary of the findings and specify the direction of significant effects. Below, we describe each set of findings separately, by criterion variable, first in terms of the bivariate relations tested using correlations, and for those significant bivariate relations, we report findings from hierarchical multiple regressions testing for unique relations.

Construct Validity of Scores Reflecting the Trait and Context Components

Bivariate Relations. Bivariate correlations revealed significant relations between independent observers' ratings of adolescent social anxiety and scores reflecting the *trait* ($r = .51; p < .001$) and *context* ($r = .24; p < .01$) Satellite Model components, but not the *perspective* component ($r = .02; p = .81$). Similarly, bivariate correlations revealed significant relations between independent observers' ratings of adolescent social skills and scores reflecting the *trait* ($r = -.48; p < .001$) and *context* ($r = -.28; p < .001$) Satellite Model components, but not the *perspective* component ($r = -.11; p = .21$). These findings informed our tests of the unique relations of scores reflecting Satellite Model components and independent observers' ratings of adolescent social anxiety and social skills within the Unfamiliar Peer Paradigm.

Unique Relations. As a follow-up to our tests of bivariate relations, we constructed hierarchical regression models to test the unique effects of scores reflecting the *trait* and *context* Satellite Model components, using the analytic plan described previously. In Step 1 of the regression model testing unique effects in relation to independent observers' ratings of adolescent social anxiety ($\beta = .51; \Delta R^2 = .26; p < .001$), and social skills ($\beta = -.48; \Delta R^2 = .23; p < .001$), there was a significant effect of scores reflecting the *trait* Satellite Model component. Over-and-above effects observed in Step 1, scores reflecting the *context* component incrementally contributed a significant and moderate-magnitude effect in Step 2 for both social anxiety ($\beta = .24; \Delta R^2 = .06; p < .001$) and social skills ($\beta = -.29; \Delta R^2 = .08; p < .001$). In Step 2, scores reflecting the *trait* component continued to demonstrate a significant, unique effect for both social anxiety ($\beta = .51; p < .001$) and social skills ($\beta = -.48; p < .001$). Thus, in both of these regression models, scores reflecting the *trait* and *context* components demonstrated unique relations with trained independent observers' ratings about adolescents' social anxiety and social skills.

Construct Validity of Scores Reflecting the Context and Perspective Components

Bivariate Relations. Bivariate correlations revealed significant relations between parent reports of adolescent-parent conflict and scores reflecting the *context* ($r = -.27; p < .01$) and *perspective* ($r = .25; p < .01$) Satellite Model components, but not the *trait* component ($r = .13; p = .13$). These findings informed our tests of the unique relations of scores reflecting Satellite Model components and parent reports of adolescent-parent conflict in the home.

Unique Relations. As a follow-up to our tests of bivariate relations, we constructed a hierarchical regression model to test the unique effects of scores reflecting the *context* and *perspective* Satellite Model components, using the analytic plan described previously. In Step 1 of the regression model testing unique effects in relation to parent reports of adolescent-parent conflict, there was a significant effect of scores reflecting the *context* Satellite Model component ($\beta = -.27; \Delta R^2 = .07; p < .01$). Over-and-above effects observed in Step 1, scores reflecting the

perspective component incrementally contributed a significant and moderate-magnitude effect in Step 2 ($\beta = .23$; $\Delta R^2 = .05$; $p < .01$). In Step 2, scores reflecting the *context* component continued to demonstrate a significant, unique effect ($\beta = -.24$; $p < .01$). Thus, scores reflecting the *context* and *perspective* components demonstrated unique relations with parent reports of adolescent-parent conflict in the home.

Construct Validity of Scores Reflecting the Trait and Perspective Components

Bivariate Relations. Bivariate correlations revealed significant relations between adolescent self-reports of resting arousal and scores reflecting the *trait* ($r = .38$; $p < .001$) and *perspective* ($r = -.27$; $p < .001$) Satellite Model components, but not the *context* component ($r = .02$; $p = .83$). These findings informed our tests of the unique relations of scores reflecting Satellite Model components and adolescent self-reports of resting arousal.

Unique Relations. As a follow-up to our tests of bivariate relations, we constructed a hierarchical regression model to test the unique effects of scores reflecting the *trait* and *perspective* Satellite Model components, using the analytic plan described previously. In Step 1 of the regression model testing unique effects in relation to adolescent self-reports of resting arousal, there was a significant effect of scores reflecting the *trait* Satellite Model component ($\beta = .38$; $\Delta R^2 = .14$; $p < .001$). Over-and-above effects observed in Step 1, scores reflecting the *perspective* component incrementally contributed a significant and moderate-magnitude effect in Step 2 ($\beta = -.26$; $\Delta R^2 = .07$; $p < .001$). In Step 2, scores reflecting the *trait* component continued to demonstrate a significant, unique effect ($\beta = .37$; $p < .001$). Thus, scores reflecting the *trait* and *perspective* components demonstrated unique relations with adolescent self-reports of resting arousal.

Discussion

Main Findings

Recent tests of Kraemer and colleagues' (2003) Satellite Model support the criterion-related validity of the approach (Makol et al., 2020). Yet, we know little about the degree to which scores reflecting the Satellite Model components (i.e., *trait*, *context*, *perspective*) measure what they were intended to measure. Our construct validation test of the Satellite Model yielded two key sets of findings. First, as in prior work (Achenbach et al., 1987; De Los Reyes et al., 2015), we observed low-to-moderate levels of correspondence, with informants observing behavior in different contexts (i.e., parents and UOs) displaying particularly low levels of correspondence (Table 2). In line with recent work (e.g., De Los Reyes et al., 2016; Lerner et al., 2017; Makol et al., 2019, 2021), we also found that the patterns of reports within a given set of informants (e.g., parent > teacher; youth < parent) tended to operate similarly across rated domains. In fact, when we applied the Satellite Model to integrating reports taken from adolescents, parents, and UOs, the eigenvalues and component loadings were remarkably similar across the domains measured with these reports (i.e., social anxiety, avoidance behaviors, fears of negative evaluation, depression, impairments; see Table 3). These findings culminated in creating multi-item scales of the *trait*, *context*, and *perspective* components derived from the Satellite Model; each scale displayed high internal consistency (i.e., $\alpha > .80$; see Nunnally & Bernstein, 1994), particularly for short, six-item scales (see also Youngstrom et al., 2019).

Second, we curated a set of criterion variables that, like the Satellite Model components themselves, systematically varied from each other in the contexts, perspectives, and sources of measurement used to create them. The multi-item *trait*, *context*, and *perspective* scales yielded

distinct patterns of relations with this diverse set of criterion variables (Table 4). Further, these patterns of relations aligned with prior work on how adolescent, parent, and UUO reports relate to each other *and* these criterion variables. For instance, prior work indicates that all three of these informants' reports individually relate to trained independent observers' ratings of adolescent behavior (Botkin et al., 2021; Glenn et al., 2019; Rezeppa et al., 2021). A logical extension of this observation is that scores reflecting the *trait* component would relate to independent observers' ratings, which they did. However, recall that these independent observers based their ratings on adolescents' behavior within interactions that occur in non-home contexts (i.e., within the Unfamiliar Peer Paradigm). Thus, we observed a “match” between the context used to estimate observed adolescent social anxiety and social skills and the *context* component of the Satellite Model. Further, trained independent observers made their ratings from a perspective that, by construction, is distinct from the adolescents, parents, and UUOs whose reports were integrated using the Satellite Model. This explains why we observed non-significant relations between independent observers' ratings and the *perspective* component.

We observed similarly coherent patterns with our two other criterion variables. Specifically, parent reports of home-specific psychosocial experiences (adolescent-parent conflict) displayed unique relations with scores reflecting the *context* and *perspective* components, but not the *trait* component. In light of the low correlations between parent and UUO reports (Table 2), it makes sense that a criterion variable based on the parent report would fail to display relations with scores reflecting the *trait* component, given that, by design, the trait component loads strongly onto all of the informants' reports (Table 3). Thus, we observed a “match” between the context and perspective used to estimate adolescent-parent conflict and the *context* and *perspective* components of the Satellite Model.

In contrast to the effects observed with parent reports of adolescent-parent conflict, adolescent self-reports of a context-neutral, internal experience (resting arousal) displayed unique relations with scores reflecting the *trait* and *perspective* components, but not the *context* component. In line with the Satellite Model, the relation between the *perspective* component and adolescent resting arousal stems from the fact that we estimated resting arousal in a way that “matched” a self-perspective (i.e., using adolescent self-reports). Yet, how do we explain the *trait* component relation? Here too, we can look to prior work. Specifically, resting arousal relates to informants' reports of youth anxiety (see Monk et al., 2001; Thomas et al., 2012a). Further, recent work finds that adolescents' self-reported arousal relates to not only adolescent self-reports on mental health surveys but also the survey reports of other informants like peer confederates and UUOs (Karp et al., 2018; Qasmieh et al., 2018; Rausch et al., 2017; Rezeppa et al., 2021). Thus, multiple informants in our Satellite Model—namely adolescents and UUOs—made survey reports about adolescent mental health that each relate to adolescent self-reports of resting arousal. Further, informants like adolescents and UUOs vary in the contexts in which they observe behavior. This logically results in detecting links between self-reported arousal and scores reflecting the *trait* component. Because no coherent pattern exists between the context-specificity of these informants, we would not expect relations between scores reflecting the *context* component and a context-neutral criterion variable such as resting arousal. Taken together, our findings support the ability to create multi-item, internally consistent scales reflecting components derived from the Satellite Model, and in a way that results in domain-relevant scores, essentially data pertaining to what these components were designed to reflect.

Research and Theoretical Implications

Our study has important implications for research and theory germane to youth mental health assessments. In particular, researchers in youth mental health have both long-observed informant discrepancies in assessments of youth mental health (e.g., De Los Reyes, 2013; Lapouse & Monk, 1958), and *theorized* that these discrepancies might be reflective of the situational specificity of youth mental health concerns (Achenbach et al., 1987). We highlighted a series of studies over the last decade that support the notion of situational specificity (for reviews, see De Los Reyes et al., 2019a, 2019b), as well as a conceptual model to guide work on these issues (OTM; De Los Reyes et al., 2013). Yet, there is uncertainty about how to *integrate* multi-informant data in a way that preserves both the *unique variance* that each report contributes, as well as the *common variance* these reports collectively contribute.

Essentially, the developers of the Satellite Model sought to instantiate notions of situational specificity within integrated scores taken from multi-informant assessments. When leveraged to integrate multi-informant assessments of adolescent mental health, our study supports these notions about what scores taken from the model reflect. In particular, the model appears to capture domain-relevant common variance (i.e., *trait*) as well as domain-relevant unique variance (i.e., *context, perspective*). This balance between emphasizing both common variance and domain-relevant unique variance is a rarity in integrative approaches. In fact, available approaches largely emphasize common variance and treat unique variance as measurement confounds (see Eid et al., 2008; Offord et al., 1996; Rubio-Stipec et al., 2003; Valo & Tannock, 2010; Watts et al., 2021; Youngstrom et al., 2003). In these respects, we see the Satellite Model as facilitating the alignment of conceptual and measurement models of youth mental health assessment. At the same time, we suspect it is not the only model that can facilitate this alignment. Thus, a key direction for future research will involve refining existing approaches to similarly balance the measurement of common variance and domain-relevant unique variance, as well as design new approaches that instill such balance when integrating multi-informant data.

Clinical Implications

Our findings also have important clinical implications. Indeed, as an approach to data integration, the PCA procedures used to estimate the components of the Satellite Model result in sample-level estimates of component loadings, but also individual-level scores for each participant in the sample. This comprises the first step in testing whether scores designed to estimate these components reflect the domains they were intended to reflect (i.e., *trait, context, perspective*). This is also the first step in developing approaches that facilitate interpreting multi-informant data in service settings with individual clients. In these respects, recent work charts a path toward developing these individual-level approaches (for a review, see Talbott & De Los Reyes, 2022). As others have noted (e.g., Makol et al., 2020), PCA is like any other data aggregation technique. In fact, the way in which PCA is incorporated into the Satellite Model shares important similarities with how others have applied this and related factor analytic procedures to enhance the interpretability of individual-level summary scores from client assessments (e.g., total scores from a parent report on a behavioral checklist; see Achenbach & Rescorla, 2001). Consequently, we see two key directions for future research.

First, scaling up the Satellite Model will involve developing normative scores to reflect its constituent components. This would require not only large, representative samples to produce model estimates, but also new assessment strategies for collecting reports from informants that do not tend to appear in traditional clinical assessments of youth mental health, namely UOs.

As we have articulated previously (Rezeppa et al., 2021), the practical value of our approach to collecting reports from UOs lies in our reliance on (a) untrained raters and (b) videotaped segments of therapeutic activities already widely implemented in evidence-based interventions administered to youth clients (i.e., therapeutic exposures; Cannon et al., 2020). As such, we designed our application of the Satellite Model in a way that optimizes its clinical feasibility. In this respect, we encourage future research that probes the clinical feasibility of this approach.

The second step will involve developing validation approaches to interpret scores taken from the Satellite Model with individual cases. That is, to simply assume that these scores accurately reflect the domains they were designed to reflect when working with individual clients would be akin to engaging in the naming fallacy described previously with respect to how researchers interpret the results of factor analytic procedures (see Kline, 2016). Yet, this too is an issue that assessors might overcome with measurement at the individual case level. For instance, consider an assessor who has access to both the multi-informant data used to develop normed Satellite Model scores *and* clients' scores on criterion measures that vary on the context, perspective, and source of measurement. With these data, an assessor can create a client-level version of our study. The goal would be to verify that scores reflecting the Satellite Model components are operating as intended and that patterns of scores reflecting the *trait*, *context*, and *perspective* components “match” or demonstrate consistencies with the battery of criterion variables (see Table 4). These issues merit further study.

Limitations

The limitations of our study highlight directions for future research. A key aim of this study involved testing links between multi-informant assessments of adolescent mental health that incorporated reports from adolescents, parents, and UOs. Importantly, the UOs based their reports on observations of adolescents interacting with unfamiliar peer confederates within the Unfamiliar Peer Paradigm. These confederates were undergraduate and post-baccalaureate personnel who we trained to simulate unfamiliar, same-age peers. Consistent with prior work (Cannon et al., 2020; Deros et al., 2018), we only leveraged the assistance of personnel who appeared youthful and could reasonably appear to adolescents as same-age, unfamiliar peers (e.g., wearing age-appropriate casual clothing, no facial hair for male confederates). Yet, peer confederates were a different age relative to our study participants. Further, we did not examine the degree to which adolescents believed that these confederates were their own age. Importantly, in prior work, we learned that adolescents' reactions to unfamiliar peer confederates within this paradigm predict their reactions to a well-established task where they are (a) told explicitly that they would be interacting with same-age, unfamiliar peers; and (b) provided with photographic stimuli to support this element of the task (i.e., Cyberball; see Karp et al., 2018). Nevertheless, we cannot be certain that adolescents' reactions to the Unfamiliar Peer Paradigm would have been identical to their reactions to interactions with same-age, unfamiliar peers in general. Further, we recruited participants within a fairly limited age range of 14–15 year olds. Thus, our findings might not generalize to adolescents within earlier and later developmental periods as well as from different geographical locations. Future research should examine the generalizability of the findings when using age-matched adolescents as peer confederates, and within samples of older and younger adolescents.

Concluding Comments

Despite the ubiquity of informant discrepancies in multi-informant assessments of youth mental health, no consensus guidelines exist for integrating informants' reports. Without clarity in these areas, researchers and clinicians lack guidance on how to leverage multiple informants' reports to make clinical decisions. In this paper, we described approaches to integrating multi-informant data, with a particular emphasis on Kraemer and colleagues' (2003) Satellite Model, a promising approach designed to yield accurate estimates of both common variance and domain-relevant unique variance observed in multi-informant data. Along these lines, we performed a thorough construct validation test of the Satellite Model, which included (a) estimating the levels of correspondence among multiple informant reports, (b) applying the Satellite Model to a battery of multi-informant surveys on parallel instruments, (c) testing the internal consistencies of aggregated "total scores" of Satellite Model common and unique variance components from this battery, and (d) testing the construct validity of these total scores using a multi-modal battery of domain-relevant criterion variables. Our findings suggest that the Satellite Model instantiates in measurement Achenbach and colleagues' (1987) notion of situational specificity. That is, the Satellite Model captures in measurement the notion that informant discrepancies manifest because: (a) youth vary as to where they display mental health concerns and (b) the informants from whom assessors solicit reports (e.g., parents, teachers, youth, and their peers) vary in the contexts and perspectives from which they observe the youth. Taken together, the Satellite Model facilitates unifying conceptual and measurement models of youth mental health. As such, a key direction for future research involves testing versions of the Satellite Model that can be applied to individual or case-level data.

Table 1

Means (M), Standard Deviations (SD), and Internal Consistency Estimates (α) of Survey Measures Used to Estimate Satellite Model Components

Measure	M	SD	α
Social Interaction Anxiety Scale			
Adolescent Self-Report	28.04	16.14	.93
Parent Report about Adolescent	27.04	16.54	.95
Unfamiliar Untrained Observer	42.22	18.84	.96
Social Phobia Scale			
Adolescent Self-Report	21.42	15.41	.93
Parent Report about Adolescent	16.91	14.38	.94
Unfamiliar Untrained Observer	30.88	18.03	.96
Brief Fear of Negative Evaluation			
Adolescent Self-Report	34.80	9.18	.87
Parent Report about Adolescent	34.40	9.69	.90
Unfamiliar Untrained Observer	39.26	9.63	.92
Subtle Avoidance Frequency Examination			
Adolescent Self-Report	66.19	20.24	.93
Parent Report about Adolescent	64.74	17.43	.92
Unfamiliar Untrained Observer	77.02	18.29	.91
Work and Social Adjustment Scale for Youth			
Adolescent Self-Report	10.07	8.07	.84
Parent Report about Adolescent	8.83	7.71	.84
Unfamiliar Untrained Observer	10.90	7.38	.85
Beck Depression Inventory-II			
Adolescent Self-Report (Raw Score)	13.04	10.72	.92
Adolescent Self-Report (Square Root Transformed Score)	3.31	1.45	—
Parent Report about Adolescent (Raw Score)	6.81	7.95	.90
Parent Report about Adolescent (Square Root Transformed Score)	2.08	1.58	—
Unfamiliar Untrained Observer (Raw Score)	11.54	10.10	.94
Unfamiliar Untrained Observer (Square Root Transformed Score)	3.00	1.60	—

Note. All study aims addressed using Beck Depression Inventory-II scores were based on the square root transformed scores.

Table 2

Bivariate Correlations among Informants' Reports on Parallel Measures

Measure	<i>Adolescent- Parent</i>	<i>Adolescent- Unfamiliar Untrained Observer</i>	<i>Parent- Unfamiliar Untrained Observer</i>
Social Interaction Anxiety Scale	.39***	.32***	.24**
Social Phobia Scale	.31***	.31***	.19*
Brief Fear of Negative Evaluation	.32***	.14	-.07
Subtle Avoidance Frequency Examination	.33***	.22*	-.04
Work and Social Adjustment Scale for Youth	.27**	.12	.06
Beck Depression Inventory-II	.34***	.10	.003

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 3

Eigenvalues and Principal Components Analysis Loadings for the 18-Item Satellite Model

Item	Eigenvalue	Adolescent	Parent	UO
Social Interaction Anxiety Scale Trait Score	1.64	.79	.74	.68
Social Phobia Scale Trait Score	1.54	.78	.68	.68
Brief Fear of Negative Evaluation Trait Score	1.32	.83	.77	.17
Subtle Avoidance Frequency Examination Trait Score	1.38	.85	.70	.41
Work and Social Adjustment Scale for Youth Trait Score	1.32	.77	.73	.44
Beck Depression Inventory-II Trait Score	1.35	.82	.79	.23
Social Interaction Anxiety Scale Context Score	0.77	-.15	-.49	.71
Social Phobia Scale Context Score	0.81	-.003	-.63	.64
Brief Fear of Negative Evaluation Context Score	1.05	.16	-.38	.94
Subtle Avoidance Frequency Examination Context Score	1.04	.04	-.56	.85
Work and Social Adjustment Scale for Youth Context Score	0.95	-.13	-.39	.89
Beck Depression Inventory-II Context Score	1.00	-.01	-.27	.96
Social Interaction Anxiety Scale Perspective Score	0.59	-.59	.46	.19
Social Phobia Scale Perspective Score	0.65	-.63	.36	.35
Brief Fear of Negative Evaluation Perspective Score	0.62	-.53	.50	.29
Subtle Avoidance Frequency Examination Perspective Score	0.58	-.52	.45	.32
Work and Social Adjustment Scale for Youth Perspective Score	0.72	-.61	.56	.15
Beck Depression Inventory-II Perspective Score	0.65	-.57	.55	.15

Note. **UO** = Unfamiliar Untrained Observer.

Table 4

Summary of Findings for Construct Validation Test of the Satellite Model

Satellite Model Component	Criterion Variables Used for Construct Validation Test (Source, Context, and Perspective of Measurement)		
	<i>Social Anxiety and Social Skills</i> (Trained Observer, Non-Home Context, Neutral Perspective)	<i>Adolescent-Parent Conflict</i> (Parent, Home Context, Observer Perspective)	<i>Adolescent Resting Arousal</i> (Adolescent, Neutral Context, Self-Perspective)
<i>Trait</i>	<ul style="list-style-type: none"> Greater <i>trait</i> scores, greater anxiety and lower social skills <p style="text-align: center;">*</p>	X	<ul style="list-style-type: none"> Greater <i>trait</i> scores, greater resting arousal <p style="text-align: center;">*</p>
<i>Context</i>	<ul style="list-style-type: none"> Greater <i>context</i> scores (i.e., direction of more adolescent concerns rated by informant in the <i>non-home context</i>), greater anxiety and lower social skills <p style="text-align: center;">*</p>	<ul style="list-style-type: none"> Lower <i>context</i> scores (i.e., direction of more adolescent concerns rated by informant in the <i>home context</i>), greater conflict <p style="text-align: center;">*</p>	X
<i>Perspective</i>	X	<ul style="list-style-type: none"> Greater <i>perspective</i> scores (i.e., direction of more adolescent concerns rated by <i>observer-report informant</i>), greater conflict <p style="text-align: center;">*</p>	<ul style="list-style-type: none"> Lower <i>perspective</i> scores (i.e., direction of more adolescent concerns rated by <i>self-report informant</i>), greater arousal <p style="text-align: center;">*</p>

Note. * = Statistically significant, unique effect; x = Null effect.

References

- Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*(2), 213-232.
<https://doi.org/10.1037/0033-2909.101.2.213>
- Achenbach, T.M., & Rescorla, L.A. (2001). *Manual for the ASEBA school-age forms & profiles*. University of Vermont Research Center for Children, Youth, & Families.
- Adams, R., & Laursen, B. (2001). The organization and dynamics of adolescent conflict with parents and friends. *Journal of Marriage and Family*, *63*(1), 97-110
<https://doi.org/10.1111/j.1741-3737.2001.00097.x>
- Al Ghriwati, N., Winter, M.A., Greenlee, J.L., & Thompson, E.L. (2018). Discrepancies between parent and self-reports of adolescent psychosocial symptoms: Associations with family conflict and asthma outcomes. *Journal of Family Psychology*, *32*(7), 992-997.
<https://doi.org/10.1037/fam0000459>
- Augenstein, T.M., Visser, K.H., Gallagher, K., De Los Reyes, A., D'Angelo, E.A., & Nock, M.K. (2022). Multi-informant reports of depressive symptoms and suicidal ideation among adolescent inpatients. *Suicide and Life-Threatening Behavior*, *52*(1), 99-109.
<https://doi.org/10.1111/sltb.12803>
- Beck, A.T., Steer, R.A., & Brown, G.K. (1996). *Beck Depression Inventory—Second Edition manual*. San Antonio, TX: The Psychological Corporation.
- Becker-Haimes, E.M., Jensen-Doss, A., Birmaher, B., Kendall, P.C., & Ginsburg, G.S. (2018). Parent–youth informant disagreement: Implications for youth anxiety treatment. *Clinical Child Psychology and Psychiatry*, *23*(1), 42-56. <https://doi.org/10.1177/1359104516689586>
- Beidas, R.S., Stewart, R.E., Walsh, L., Lucas, S., Downey, M.M., Jackson, K., Fernandez, T., & Mandell, D.S. (2015). Free, brief, and validated: Standardized instruments for low-resource mental health settings. *Cognitive and Behavioral Practice*, *22*(1), 5-19.
<https://doi.org/10.1016/j.cbpra.2014.02.002>
- Borsboom, D. (2005) *Measuring the mind*. Cambridge.
- Botkin, T., Racz, S.J., Makol, B.A., & De Los Reyes, A. (2021). Multi-informant assessments of adolescents' fears of negative and positive evaluation: Criterion-related and incremental validity in relation to observed behavior. *Journal of Psychopathology and Behavioral Assessment*, *43*(1), 58-69. <https://doi.org/10.1007/s10862-020-09855-y>
- Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.
<https://doi.org/10.1037/h0046016>
- Cannon, C.J., Makol, B.A., Keeley, L.M., Qasmieh, N., Okuno, H., Racz, S.J., & De Los Reyes, A. (2020). A paradigm for understanding adolescent social anxiety with unfamiliar peers: Conceptual foundations and directions for future research. *Clinical Child and Family Psychology Review*, *23*(3), 338-364. <https://doi.org/10.1007/s10567-020-00314-4>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284-290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Cuming, S., Rapee, R.M., Kemp, N., Abbott, M.J., Peters, L., & Gaston, J.E. (2009). A self-report measure of subtle avoidance and safety behaviors relevant to social anxiety: development and psychometric properties. *Journal of Anxiety Disorders*, *23*(7), 879–883.
<https://doi.org/10.1016/j.janxdis.2009.05.002>

- De Los Reyes, A. (2020, February 15). Materials described in the paper, “A paradigm for understanding adolescent social anxiety with unfamiliar peers: Conceptual foundations and directions for future research.” Retrieved from <https://bit.ly/CCFPRPeerParadigmOSF>
- De Los Reyes, A. (2013). Strategic objectives for improving understanding of informant discrepancies in developmental psychopathology research. *Development and Psychopathology*, 25(3), 669-682. <https://doi.org/10.1017/S0954579413000096>
- De Los Reyes, A. (2011). Introduction to the special section. More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, 40(1), 1-9. <https://doi.org/10.1080/15374416.2011.533405>
- De Los Reyes, A., Augenstein, T.M., Wang, M., Thomas, S.A., Drabick, D.A.G., Burgers, D., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, 141(4), 858-900. <https://doi.org/10.1037/a0038498>
- De Los Reyes, A., Aldao, A., Thomas, S.A., Daruwala, S.E., Swan, A.J., Van Wie, M., Goepel, K.A., & Lechner, W. (2012a). Adolescent self-reports of social anxiety: Can they disagree with objective psychophysiological measures and still be valid? *Journal of Psychopathology and Behavioral Assessment*, 34(3), 308-322. <https://doi.org/10.1007/s10862-012-9289-2>
- De Los Reyes, A., Alfano, C.A., Lau, S., Augenstein, T.M., & Borelli, J.L. (2016). Can we use convergence between caregiver reports of adolescent mental health to index severity of adolescent mental health concerns? *Journal of Child and Family Studies*, 25(1), 109-123. <https://doi.org/10.1007/s10826-015-0216-5>
- De Los Reyes, A., Cook, C.R., Gresham, F.M., Makol, B.A., & Wang, M. (2019b). Informant discrepancies in assessments of psychosocial functioning in school-based services and research: Review and directions for future research. *Journal of School Psychology*, 74, 74-89. <https://doi.org/10.1016/j.jsp.2019.05.005>
- De Los Reyes, A., Cook, C.R., Sullivan, M., Morrell, N., Hamlin, C., Wang, M., Gresham, F.M., Makol, B.M., Keeley, L.M., & Qasmieh, N. (2022b). The Work and Social Adjustment Scale for Youth: Psychometric properties of the teacher version and evidence of contextual variability in psychosocial impairments. *Psychological Assessment*. Advance online publication. <https://doi.org/10.1037/pas0001139>
- De Los Reyes, A., Henry, D.B., Tolan, P.H., & Wakschlag, L.S. (2009). Linking informant discrepancies to observed variations in young children’s disruptive behavior. *Journal of Abnormal Child Psychology*, 37(5), 637-652. <https://doi.org/10.1007/s10802-009-9307-3>
- De Los Reyes, A., Lerner, M.D., Keeley, L.M., Weber, R., Drabick, D.A.G., Rabinowitz, J., & Goodman, K.L. (2019a). Improving interpretability of subjective assessments about psychological phenomena: A review and cross-cultural meta-analysis. *Review of General Psychology*, 23(3), 293-319. <https://doi.org/10.1177/108926801983764>
- De Los Reyes, A., & Makol, B.A. (2022). Informant reports in clinical assessment. In G. Asmundson (Ed.), *Comprehensive clinical psychology*. (2nd ed., 4th vol.) (pp. 105-122). Elsevier. <https://doi.org/10.1016/B978-0-12-818697-8.00113-8>
- De Los Reyes, A., & Makol, B.A. (2019). Evidence-based assessment. In T.H. Ollendick, L. Farrell, and P. Muris (Eds.), *Innovations in CBT for childhood anxiety, OCD, and PTSD: Improving access and outcomes* (pp. 28-51). Cambridge.

- De Los Reyes, A., Makol, B.A., Racz, S.J., Youngstrom, E.A., Lerner, M.D., & Keeley, L.M. (2019c). The Work and Social Adjustment Scale for Youth: A measure for assessing youth psychosocial impairment regardless of mental health status. *Journal of Child and Family Studies*, 28(1), 1-16. <https://doi.org/10.1007/s10826-018-1238-6>
- De Los Reyes, A., Talbott, E., Power, T., Michel, J., Cook, C.R., Racz, S.J., & Fitzpatrick, O. (2022a). The Needs-to-Goals Gap: How informant discrepancies in youth mental health assessments impact service delivery. *Clinical Psychology Review*, 92, 102114. <https://doi.org/10.1016/j.cpr.2021.102114>
- De Los Reyes, A., Thomas, S.A., Goodman, K.L., & Kundey, S.M.A. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology*, 9, 123-149. <https://doi.org/10.1146/annurev-clinpsy-050212-185617>
- De Los Reyes, A., Thomas, S.A., Swan, A.J., Ehrlich, K.B., Reynolds, E.K., Suarez, L., Dougherty, L.R., MacPherson, L., & Pabón, S.C. (2012b). "It depends on what you mean by 'disagree'": Differences between parent and child perceptions of parent-child conflict. *Journal of Psychopathology and Behavioral Assessment*, 34(3), 293-307. <https://doi.org/10.1007/s10862-012-9288-3>
- De Los Reyes, A., Tyrell, F., Watts, A.L., & Asmundson, G.J.G. (2022c). Conceptual, methodological, and measurement factors that disqualify use of measurement invariance techniques to detect informant discrepancies in youth mental health assessments. *Frontiers in Psychology*. Advance online publication. <https://doi.org/10.3389/fpsyg.2022.931296>
- Deros, D.E., Racz, S.J., Lipton, M.F., Augenstein, T.M., Karp, J.N., Keeley, L.M., Qasmieh, N., Grewe, B., Aldao, A., & De Los Reyes (2018). Multi-informant assessments of adolescent social anxiety: Adding clarity by leveraging reports from unfamiliar peer confederates. *Behavior Therapy*, 49(1), 84-98. <https://doi.org/10.1016/j.beth.2017.05.001>
- Dirks, M.A., De Los Reyes, A., Briggs-Gowan, M., Cella, D., & Wakschlag, L.S. (2012). Annual Research Review: Embracing not erasing contextual variability in children's behavior-theory and utility in the selection and use of methods and informants in developmental psychopathology. *Journal of Child Psychology and Psychiatry*, 53(5), 558-574. <https://doi.org/10.1111/j.1469-7610.2012.02537.x>
- DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment*, 23(3), 225-241. <https://doi.org/10.1177/073428290502300303>
- Ehrlich, K.B., Cassidy, J., & Dykas, M.J. (2011). Reporter discrepancies among parents, adolescents, and peers: Adolescent attachment and informant depressive symptoms as explanatory factors. *Child Development*, 82(3), 999-1012. <https://doi.org/10.1111/j.1467-8624.2010.01530.x>
- Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*, 13(3), 230-253. <https://doi.org/10.1037/a0013219>
- Edgeworth, F.Y. (1888). The statistics of examinations. *Journal of the Royal Statistical Society*, 51(3), 598-635. Retrieved from <https://www.jstor.org/stable/2339898>
- Epkins, C.C., & Heckler, D.R. (2011). Integrating etiological models of social anxiety and depression in youth: Evidence for a cumulative interpersonal risk model. *Clinical Child and Family Psychology Review*, 14(4), 329-376. <https://doi.org/10.1007/s10567-011-0101-8>

- Floorean, I.S., Dobrean, A., Balazsi, R., Roşan, A., Păsărelu, C.R., Predescu, E., & Rad, F. (2022). Measurement invariance of Alabama Parenting Questionnaire Across age, gender, clinical status, and informant. *Assessment*. Advance online publication. <https://doi.org/10.1177/10731911211068178>
- Garner, W.R., Hake, H.W., & Eriksen, C.W. (1956). Operationism and the concept of perception. *Psychological Review*, 63(3), 149-159. <https://doi.org/10.1037/h0042992>
- Glenn, L.E., Keeley, L.M., Szollos, S., Okuno, H., Wang, X., Rausch, E., Deros, D.E., Karp, J.N., Qasmieh, N., Makol, B.A., Augenstein, T.M., Lipton, M.F., Racz, S.J., Scharfstein, L., Beidel, D.C., & De Los Reyes, A. (2019). Trained observers' ratings of adolescents' social anxiety and social skills within controlled, cross-contextual social interactions with unfamiliar peer confederates. *Journal of Psychopathology and Behavioral Assessment*, 41(1), 1-15. <https://doi.org/10.1007/s10862-018-9676-4>
- Greenberg, A., & De Los Reyes, A. (2022). When adolescents experience co-occurring social anxiety and ADHD symptoms: Links with social skills when interacting with unfamiliar peer confederates. *Behavior Therapy*. Advance online publication. <https://doi.org/10.1016/j.beth.2022.04.011>
- Hofmann, S.G., Albano, A.M., Heimberg, R.G., Tracey, S., Chorpita, B.F., & Barlow, D.H. (1999). Subtypes of social phobia in adolescents. *Depression and Anxiety*, 9(1), 15-18. [https://doi.org/10.1002/\(SICI\)1520-6394\(1999\)9:1<15::AID-DA2>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1520-6394(1999)9:1<15::AID-DA2>3.0.CO;2-6)
- Humphreys, K.L., Weems, C.F., & Scheeringa, M.S. (2017). The role of anxiety control and treatment implications of informant agreement on child PTSD symptoms. *Journal of Clinical Child and Adolescent Psychology*, 46(6), 903-914. <https://doi.org/10.1080/15374416.2015.1094739>
- Hunsley, J., & Mash, E.J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology*, 3, 29-51. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091419>
- Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D.S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, 167(7), 748-751. <http://dx.doi.org/10.1176/appi.ajp.2010.09091379>
- Karp, J.K., Makol, B.A., Keeley, L.M., Qasmieh, N., Deros, D.E. Weeks, J.W., Racz, S.J., Lipton, M.F., Augenstein, T.M., & De Los Reyes, D. (2018). Convergent, incremental, and criterion-related validity of multi-informant assessments of adolescents' fears of negative and positive evaluations. *Clinical Psychology and Psychotherapy*, 25(2), 217-230. <https://doi.org/10.1002/cpp.2156>
- Kline, R.B. (2016). *Principles and practice of structural equation modeling* (4th ed). Guilford.
- Kraemer, H.C., Measelle, J.R., Ablow, J.C., Essex, M.J., Boyce, W.T., & Kupfer, D.J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry*, 160(9), 1566-1577. <https://doi.org/10.1176/appi.ajp.160.9.1566>
- Lang, P.J. (1980). Behavioral treatment and bio-behavioral assessment: computer applications. In J.B. Sidowski, J.H. Johnson & T.A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 119-137). Norwood, NJ: Ablex.
- Lapouse, R., & Monk, M.A. (1958). An epidemiologic study of behavior characteristics in children. *American Journal of Public Health*, 48(9), 1134-1144. <https://doi.org/10.2105/AJPH.48.9.1134>
- Leary, M.R. (1983). A brief version of the Fear of Negative Evaluation Scale. *Personality and*

- Social Psychology Bulletin*, 9(3), 371-375. <https://doi.org/10.1177/0146167283093007>
- Lerner, M.D., De Los Reyes, A., Drabick, D.G., Gerber, A.H., & Gadow, K.D. (2017). Informant discrepancy defines discrete, clinically useful autism spectrum disorder subgroups. *Journal of Child Psychology and Psychiatry*, 58(7), 829-839. <https://doi.org/10.1111/jcpp.12730>
- Lippold, M.A., Greenberg, M.T., & Collins, L.M. (2014). Youths' substance use and changes in parental knowledge-related behaviors during middle school: A person-oriented approach. *Journal of Youth and Adolescence*, 43(5), 729-744. <https://doi.org/10.1007/s10964-013-0010-x>
- Makol, B.A., De Los Reyes, A., Garrido, E., Harlaar, N., & Taussig, H. (2021). Assessing the mental health of maltreated youth with child welfare involvement using multi-informant reports. *Child Psychiatry and Human Development*, 52(1), 49-62. <https://doi.org/10.1007/s10578-020-00985-8>
- Makol, B.A., De Los Reyes, A., Ostrander, R., & Reynolds, E.K. (2019). Parent-youth divergence (and convergence) in reports of youth internalizing problems in psychiatric inpatient care. *Journal of Abnormal Child Psychology*, 47(10), 1677-1689. <https://doi.org/10.1007/s10802-019-00540-7>
- Makol, B.A., Youngstrom, E.A., Racz, S.J., Qasmieh, N., Glenn, L.E., & De Los Reyes, A. (2020). Integrating multiple informants' reports: How conceptual and measurement models may address long-standing problems in clinical decision-making. *Clinical Psychological Science*, 8(6), 953-970. <https://doi.org/10.1177/2167702620924439>
- Markon, K.E., Chmielewski, M., & Miller, C.J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin*, 137(5), 856-879. <http://dx.doi.org/10.1037/a0023678>
- Marsh, J.K., De Los Reyes, A., & Lilienfeld, S.O. (2018). Leveraging the multiple lenses of psychological science to inform clinical decision-making: Introduction to the special section. *Clinical Psychological Science*, 6(2), 167-176. <https://doi.org/10.1177/2167702617736853>
- Martel, M.M., Eng, A.G., Bansal, P.S., Smith, T.E., Elkins, A.R., & Goh, P.K. (2021). Multiple informant average integration of ADHD symptom ratings predictive of concurrent and longitudinal impairment. *Psychological Assessment*, 33(5), 443-451. <https://doi.org/10.1037/pas0000994>
- Mattick, R.P., & Clarke, J.C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, 36(4), 455-470. [https://doi.org/10.1016/s0005-7967\(97\)10031-6](https://doi.org/10.1016/s0005-7967(97)10031-6)
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Monk, C., Kovelenco, P., Ellman, L. M., Sloan, R.P., Bagiella, E., Gorman, J.M., & Pine, D.S. (2001). Enhanced stress reactivity in pediatric anxiety disorders: Implications for future cardiovascular health. *International Journal of Neuropsychopharmacology*, 4(2), 199-206. <https://doi.org/10.1017/S146114570100236X>
- Murray, A.L., Speyer, L.G., Hall, H.A., Valdebenito, S., & Hughes, C. (2021). Teacher versus parent informant measurement invariance of the Strengths and Difficulties Questionnaire. *Journal of Pediatric Psychology*, 46(10), 1249-1257.

- <https://doi.org/10.1093/jpepsy/jsab062>
- Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Offord, D.R., Boyle, M.H., Racine, Y., Szatmari, P., Fleming, J.E., Sanford, M., & Lipman, E.L. (1996). Integrating assessment data from multiple informants. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35(8), 1078-1085.
<http://dx.doi.org/10.1097/00004583-199608000-00019>
- Okuno, H., Rezeppa, T., Raskin, T., & De Los Reyes, A. (2021). Adolescent safety behaviors and social anxiety: Links to psychosocial impairments and functioning with unfamiliar peer confederates. *Behavior Modification*. Advance online publication.
<https://doi.org/10.1177/01454455211054019>
- Olino, T.M., Finsaas, M., Dougherty, L.R., & Klein, D.N. (2018). Is parent-child disagreement on child anxiety explained by differences in measurement properties? An examination of measurement invariance across informants and time. *Frontiers in Psychology*, 9, 1295.
<https://doi.org/10.3389/fpsyg.2018.01295>
- Osborne, J.W., & Costello, A.B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research, and Evaluation*, 9(1), Article 11.
<https://doi.org/10.7275/ktzq-jq66>
- Pett, M.A., Lackey, N.R., & Sullivan, J.J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Sage.
- Prinz, R.J., Foster, S.L., Kent, R.N., & O'Leary, K.D. (1979). Multivariate assessment of conflict in distressed and nondistressed mother-adolescent dyads. *Journal of Applied Behavior Analysis*, 12(4), 691-700. <https://doi.org/10.1901/jaba.1979.12-691>
- Qasmieh, N., Makol, B.A., Augenstein, T.M., Lipton, M.F., Deros, D.E., Karp, J.N., Keeley, L.M., Truong, M.L., Racz, S.J., & De Los Reyes, A. (2018). A multi-informant approach to assessing safety behaviors among adolescents: Psychometric properties of the Subtle Avoidance Frequency Examination. *Journal of Child and Family Studies*, 27(6), 1830-1843. <https://doi.org/10.1007/s10826-018-1040-5>
- Rausch, E., Racz, S.J., Augenstein, T.M., Keeley, L., Lipton, M.F., Szollos, S., Riffle, J., Moriarity, D., Kromash, R., & De Los Reyes, A. (2017). A multi-informant approach to measuring depressive symptoms in clinical assessments of adolescent social anxiety using the Beck Depression Inventory-II: Convergent, incremental, and criterion-related validity. *Child and Youth Care Forum*, 46(5), 661-683.
<https://doi.org/10.1007/s10566-017-9403-4>
- Rezeppa, T., Okuno, H., Qasmieh, N., Racz, S.J., Borelli, J.L., & De Los Reyes, A. (2021). Unfamiliar untrained observers' ratings of adolescent safety behaviors within social interactions with unfamiliar peer confederates. *Behavior Therapy*. 52(3), 564-576.
<https://doi.org/10.1016/j.beth.2020.07.006>
- Rubio-Stipek, M., Fitzmaurice, G., Murphy, J., & Walker, A. (2003). The use of multiple informants in identifying the risk factors of depressive and disruptive disorders: Are they interchangeable? *Social Psychiatry and Psychiatric Epidemiology*, 38(2), 51-58.
<https://doi.org/10.1007/s00127-003-0600-0>
- Russell, J.D., Graham, R. A., Neill, E.L. & Weems, C.F. (2016). Agreement in youth-parent perceptions of parenting behaviors: A case for testing measurement invariance in reporter discrepancy research. *Journal of Youth and Adolescence*, 45(10), 2094-2107.
<https://doi.org/10.1007/s10964-016-0495-1>
- Smetana, J.G. (2008). "It's 10 o'clock: Do you know where your children are?" Recent

- advances in understanding parental monitoring and adolescents' information management. *Child Development Perspectives*, 2(1), 19-25.
<https://doi.org/10.1111/j.1750-8606.2008.00036.x>
- Smetana, J., & Gaines, C. (1999). Adolescent-parent conflict in middle-class African American families. *Child Development*, 70(6), 1447-1463. <https://doi.org/10.1111/1467-8624.00105>
- Strauss, M.E., & Smith, G.T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, 5, 1-25.
<https://doi.org/10.1146/annurev.clinpsy.032408.153639>
- Szollos, S., Keeley, L.M., Makol, B.A., Weeks, J.W., Racz, S.J., Lipton, M.F., Augenstein, T.M., Beale, A.M., & De Los Reyes, A. (2019). Multi-Informant assessments of individual differences in adolescents' socio-evaluative fears: Clinical correlates and links to arousal within social interactions. *Journal of Child and Family Studies*, 28(12), 3360-3373.
<https://doi.org/10.1007/s10826-019-01517-2>
- Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics*. Allyn and Bacon.
- Talbott, E., & De Los Reyes, A. (2022). Making sense of multiple data sources: Using single case design research for behavioral decision making. In T. Farmer, E. Talbott, K. McMaster, D. Lee, & T. Aceves (Ed.), *Handbook of special education research: Theory, methods, and developmental processes* (pp. 231-244). Routledge.
- Thomas, S.A., Aldao, A., & De Los Reyes, A. (2012a). Implementing clinically feasible psychophysiological measures in evidence-based assessments of adolescent social anxiety. *Professional Psychology: Research and Practice*, 43(5), 510-519
<https://doi.org/10.1037/a0029183>
- Thomas, S.A., Daruwala, S.E., Goepel, K.A., & De Los Reyes, A. (2012b). Using the Subtle Avoidance Frequency Examination in adolescent social anxiety assessments. *Child and Youth Care Forum*, 41(6), 547-559. <https://doi.org/10.1007/s10566-012-9181-y>
- Treutler, C.M., & Epkins, C.C. (2003). Are discrepancies among child, mother, and father reports on children's behavior related to parents' psychological symptoms and aspects of parent-child relationships? *Journal of Abnormal Child Psychology*, 31(1), 13-27
<https://doi.org/10.1023/A:1021765114434>
- Valo, S., & Tannock, R. (2010). Diagnostic instability of DSM-IV ADHD subtypes: effects of informant source, instrumentation, and methods for combining symptom reports. *Journal of Clinical Child and Adolescent Psychology*, 39(6), 749-760.
<https://doi.org/10.1080/15374416.2010.517172>
- Watts, A.L., Makol, B.A., Palumbo, I.M., De Los Reyes, A., Olino, T.M., Lutzman, R.D., DeYoung, C.G., Wood, P.K., & Sher, K.J. (2021). How robust is the p-factor? Using multitrait-multimethod modeling to inform the meaning of general factors of psychopathology in youth. *Clinical Psychological Science*. Advance online publication.
<https://doi.org/10.1177/21677026211055170>
- Youngstrom, E.A., Salcedo, S., Frazier, T.W., & Perez Algorta, G. (2019). Is the finding too good to be true? Moving from "more is better" to thinking in terms of simple predictions and credibility. *Journal of Clinical Child and Adolescent Psychology*, 48(6), 811-824.
<https://doi.org/10.1080/15374416.2019.1669158>
- Youngstrom, E.A., Findling, R.L., & Calabrese, J.R. (2003). Who are the comorbid adolescents? Agreement between psychiatric diagnosis, youth, parent, and teacher report. *Journal of Abnormal Child Psychology*, 31(3), 231-245. <http://dx.doi.org/10.1023/A:1023244512119>
- Zilcha-Mano, S., Shimshoni, Y., Silverman, W.K., & Lebowitz, E.R. (2021). Parent-child

agreement on family accommodation differentially predicts outcomes of child-based and parent-based child anxiety treatment. *Journal of Clinical Child and Adolescent Psychology*, 50(3), 427-439. <https://doi.org/10.1080/15374416.2020.1756300>

Integrating Multi-Informant Reports of Youth Mental Health: A Construct Validation Test of Kraemer and Colleagues' (2003) Satellite Model

Natalie Charamut¹, Sarah J. Racz¹, Mo Wang², Andres De Los Reyes^{1*}

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

²Department of Management, University of Florida, Gainesville, FL, USA

* Correspondence:

Andres De Los Reyes (adlr@umd.edu)

ONLINE SUPPLEMENTARY MATERIAL

Integrating Multi-Informant Reports of Youth Mental Health: A Construct Validation Test of Kraemer and Colleagues' (2003) Satellite Model

ONLINE SUPPLEMENTARY MATERIAL

Characteristics of Unfamiliar Untrained Observers (UOs)

Demographics. Two sets of UOs provided reports on adolescents in this sample. One set of UOs made reports of adolescents specifically on the SAFE. These 28 “SAFE-only” UOs had a mean age of 21.07 years ($SD = 2.92$ years) and included 21 females and 7 males. The SAFE-only UOs self-identified their racial/ethnic background as African American or Black (10.7%); Asian American or Asian (25%); Hispanic or Latino/a (Spanish) (17.9%); or White, Caucasian American, or European (53.6%) (rates total above 100% because UOs could select multiple response options). The SAFE-only UOs self-reported their current educational status at the time they completed their reports as an undergraduate sophomore (25%), junior (32.1%), senior (25%); a post-baccalaureate trainee (i.e., not yet matriculated in a graduate program; 7.1%); or a master’s level graduate student (10.7%).

A second set of UOs made reports of adolescents on an additional five survey measures to complete the six-survey battery (i.e., SIAS, SPS, BFNE, WSASY, & BDI-II). These 45 UOs who provided reports about adolescents on these five measures had a mean age of 19.96 years ($SD = 1.61$ years) and included 36 females and 9 males. The UOs self-identified their racial/ethnic background as African American or Black (11.1%); Asian American or Asian (33.3%); Hispanic or Latino/a (Spanish) (15.6%); White, Caucasian American, or European (48.9%); American Indian or Native American (2.2%); or “other” (i.e., Middle Eastern, Iranian; 4.4%) (rates total above 100% because UOs could select multiple response options). The 45 UOs self-reported their current educational status at the time they completed their reports as an undergraduate freshman (13.3%), sophomore (40%), junior (17.8%), senior (15.6%); a post-baccalaureate trainee (i.e., not yet matriculated in a graduate program; 4.4%); or a master’s level graduate student (8.9%).

Procedures for gathering UO reports. Based on archival videos of the 134 adolescents’ participation in the Unfamiliar Peer Paradigm, UOs were randomly assigned to each observe the video recordings of up to five adolescents. For each adolescent they observed, UOs viewed recordings of their social interactions during the Unfamiliar Peer Paradigm, and completed their survey reports about the adolescent immediately following viewing these recordings. We masked the UOs to all clinical characteristics of the adolescents about whom they provided reports, including referral status and all clinic data (e.g., scores on other instruments). Further, UOs did not participate as peer confederates with the adolescents about whom they provided reports. In this respect and similar to our procedures for peer confederates, UOs had no prior contact with the adolescents about whom they provided reports. In essence, we selected UOs and assigned them to videos of adolescents, such that we could ensure their unfamiliarity with these adolescents.

Characteristics and Training of Independent Observers

For each adolescent, two trained independent observers viewed archived videotapes of their participation in the Unfamiliar Peer Paradigm. All trained independent observers received

training on how to use the behavioral ratings of adolescent social skills. Trained independent observers consisted of post-baccalaureate and undergraduate research assistants. We masked trained independent observers to adolescents' referral status and they did not have access to adolescents' clinical information. Further and as with the UUOs, none of the trained independent observers participated as a peer confederate in the Unfamiliar Peer Paradigm.

To train independent observers on the coding scheme described below, a team of eight to ten researchers (i.e., undergraduate students, post-baccalaureate research assistants, graduate students, and faculty) participated in consensus coding meetings in which team members simultaneously viewed videos of all the social interaction tasks (i.e., SSIT, UCT, IST) performed by five adolescent participants in the sample. Following each task viewing, team members independently rated the adolescent in the video on the levels of social skills they displayed during the task, using the coding scheme below (i.e., a rating for each of the five SSIT role-plays, a rating for UCT, a rating for IST). After each team member made their ratings for a task, the entire team discussed the ratings. During this discussion, the team resolved discrepancies among ratings, and came to a final consensus rating for social skills displayed by the adolescent participant performing the task. We repeated this process for each of the five participants across all seven tasks (i.e., seven social skills consensus ratings per participant).

After creating the consensus ratings for five training cases, we trained the independent observers described previously. Each trained independent observer independently viewed videos for the five training cases and made seven social skills ratings per case. After making their training ratings, we calculated intraclass correlation (ICC) statistics to assess inter-rater reliability between each trained independent observer and the consensus ratings. We set a threshold of a mean ICC of .80 to determine whether a trained independent observer successfully passed the training stage. All trained independent observers passed our criterion ICC, and following training, these observers coded the cases in the sample to which they were assigned.