



Published in final edited form as:

J Clin Child Adolesc Psychol. 2011 January ; 40(1): 54–66. doi:10.1080/15374416.2011.533404.

Probing the Depths of Informant Discrepancies: Contextual Influences on Divergence and Convergence

Anselma G. Hartley¹, Audrey L. Zakriski², and Jack C. Wright¹

¹Department of Psychology, Brown University

²Department of Psychology, Connecticut College

Abstract

This research examined how a contextual approach to child assessment can clarify the meaning of informant discrepancies by focusing on children's social experiences and their *if...then* reactions to them. In a sample of 123 children ($M_{age}=13.30$) referred to a summer program for children with behavior problems, parent-teacher agreement for syndromal measures of aggression and withdrawal was modest. Agreement remained low when informants assessed children's reactions to specific peer and adult events. The similarity of these events increased consistency *within* informants, but had no effect on agreement *between* parents and teachers. In contrast, similarity in the pattern of social events children encountered at home and school predicted informant agreement for syndromal aggression and for aggression to aversive events. Our results underscore the robustness of informant discrepancies and illustrate how they can be studied as part of the larger mosaic of person-environment interactions.

Keywords

informant discrepancies; aggression; social environment; context; personality

Informant discrepancies are a robust but poorly understood phenomenon in childhood assessment research and practice. Numerous studies have investigated the child and informant characteristics that predict discrepancies (see Achenbach, 2006; De Los Reyes & Kazdin, 2005), but our understanding of underlying mechanisms remains incomplete. Many have theorized that such discrepancies could result from differences in children's social environments (e.g., home vs. school) and therefore in the behaviors informants observe (Achenbach, McConaughy, & Howell, 1987). Relatively few studies have directly tested these claims (De Los Reyes, Henry, Tolan, & Wakschlag, 2009) or probed the assumptions on which widely used trait or "syndromal" measures are based (Cervone, Shadel, & Jencius, 2001). The child assessment literature continues to be influenced by a nomothetic trait paradigm in which consistency over ratings remains the theoretical expectation, discrepancies are often considered noise, and thus aggregating information from multiple informants' reports is thought to be the solution to improving the signal (Barkley, 1988; Kraemer, Measelle, Ablow, Essex, Boyce, & Kupfer, 2003; Roberts & Caspi, 2001). Recent research on informant discrepancies (Beck, Hartos, & Simons-Morton, 2006; Drabick, Gadow, & Loney 2008; Guion, Mrug, and Windle, 2009), coupled with related developments in the field of personality (Fournier, Moskowitz, Zuroff, 2008; Mischel, 2009), encourages us to consider other possibilities: that informant discrepancies reflect

meaningful variation in children's behavior across situations, and that alternative approaches to assessment are needed to incorporate this variation.

Past studies of informant discrepancies have often relied on standardized syndromal instruments that ask a respondent, typically a parent or teacher, to rate how often a child displays various behaviors over a period of time (e.g., Teacher Report Form [TRF] and Child Behavior Checklist [CBCL], Achenbach & Rescorla, 2001). The popularity of such instruments stems partly from their efficiency—a short inventory can survey a range of behavior problems. Such assessments can be useful when overall rates of problem behaviors are the main concern (e.g., screening in clinical settings), and when functional origins of the behavior are not a central question. Nevertheless, the emphasis on efficiency involves tradeoffs. In essence, these measures adopt an “act frequency” (Buss & Craik, 1983) or aggregationist (Roberts & Caspi, 2001) approach in which behaviors are aggregated to capture the person's “act trend” over a period of observation, usually without specifically examining the situations in which the acts occur. In this view, situational variance in behavior is “filtered out” in order to measure the individual's true disposition (Barkley, 1988). One form of filtering takes place in the mind of the rater when making a frequency judgment about an act statement (e.g., “hits”, “teases”). A second form occurs when various items, including some that provide contextual cues (e.g., “disobedient at school”), are aggregated into broad summary scales or syndrome scores.

Alternative approaches more explicitly incorporate context into the study of personality by examining “*if...then*” relationships between social events and people's behavioral responses to them (Mischel & Shoda, 1995; Wright & Mischel, 1987). Rather than emphasizing what a person does on average, contextual approaches focus on the conditional probabilities of how a person reacts under relevant conditions, or $p(\text{Behavior} | \text{Event})$. For example, Vansteelandt and Van Mechelen's (1998) study of hostility examined several antecedent events (e.g., *if* frustrated, ignored) and adults' reported hostile reactions (e.g., *then* attack, curse). Personality is thus revealed partly in individuals' behavioral “signatures,” or the patterning of behavior around their mean level and across contexts (Fournier et al., 2008; Smith, Shoda, Cumming, & Smoll, 2009; Zakriski, Wright & Underwood, 2005). Instead of implying that consistency across situations should be generally high, these models suggest that it varies with the similarity of the situations in which people are observed (Mischel & Shoda, 1995). Attention to circumscribed consistencies and to contextualized behavior signatures has enhanced our understanding of individual differences in several domains, including anger, anxiety, and dominance (Endler, Parker, Bagby, & Cox, 1991; Fournier et al., 2008; Van Mechelen & Kiers, 1999).

With these issues in mind, we consider some implications of using syndromal assessments to study informant discrepancies. First, several studies suggest that by focusing on overall frequencies, syndromal assessments conflate dispositional and environmental influences on behavior (Smith et al., 2009). Researchers have noted that children who are nomothetically similar (e.g., high overall levels of anxiety) may be psychologically distinct (Haynes, Mumma, & Pinson, 2009; Scotti, Morris, McNeil, & Hawkins, 1996). For example, one boy may act aggressively when peers tease him, but may be teased rarely; another boy may be unlikely to act aggressively when teased, but may be teased often. Syndromal measures appear to be sensitive to overall behavior output, as they are designed to be, but insensitive to the interaction of person and context variables that contribute to that output (Wright, Lindgren, & Zakriski, 2001).

A second concern about the syndromal paradigm is that it can obscure individual differences in the contextual patterning of behavior. Research on anger and sadness in adults has found individual differences in people's reactions to the same situations that are not revealed by

their overall behavior levels (Van Mechelen & Kiers, 1999; Vansteelandt & Van Mechelen, 2006). Related work found that teacher-reported syndrome scores did not distinguish between boys who were equally high in externalizing behavior but had distinct patterns of responses to nonaversive and aversive events in interactions with peers and adults (Wright & Zakriski, 2001). Other work has revealed stability for both mean-level traits aggregated over situations and for disaggregated patterns of behavior across contexts (Fournier et al., 2008). These findings reinforce calls for methods that are sensitive to people's contextualized behavior patterns (Mischel & Shoda, 1995).

Third, to the extent that syndromal assessments conflate environmental and dispositional influences, interpretations can be especially difficult when two syndromal assessments are compared, as in research on informant discrepancies. A discrepancy could occur if a child has different social experiences at home versus school, differs in how he or she responds to those experiences when they occur, or both. Research on gender differences illustrates these issues (Maccoby, 1998). For example, some work has found that gender differences in overall prosocial behavior stem from how often girls versus boys encounter social events rather than how they respond to those events (Zakriski et al., 2005). In contrast, gender differences in overall aggression stemmed from differences in both event rates and reaction rates. Such ambiguity in the meaning of overall differences is a challenge for research that relies on syndrome scores.

By shifting the focus from overall behavioral output to *if...then* relations between contexts and behaviors, a conditional framework helps explain informant discrepancies. If certain aggressive children are primarily aggressive with peers, parent-teacher discrepancies would be expected in cases where parents have few opportunities to observe their child's peer interactions directly. Researchers have speculated that raters may base their ratings on responses to particular situations at home (e.g., when asked to clean up) or at school (e.g., academic or peer challenges) (De Los Reyes et al., 2009; Drabick et al., 2008). They have also speculated that parents may base their ratings on a comparatively limited set of parent-child interactions, whereas teachers may base theirs on a broader set of interactions across situations, interactants, and other children. Laboratory research has found that disruptive behavior with the parent was more closely related to parents' prior ratings of disruptive behavior, whereas disruptive behavior with an examiner was more closely related to teachers' ratings (De Los Reyes et al., 2009).

We suggest that informant discrepancies can be clarified by probing two distinct but inter-related processes highlighted in a conditional approach: the probability that a child encounters a given event, $p(\text{Event})$, and the probability of a behavioral reaction given that an event occurs, or $p(\text{Behavior} | \text{Event})$. Our study examines parents' and teachers' ratings for children accepted into a summer program for youth with behavior problems. We use a syndromal instrument that is often employed in cross-informant research (CBCL/TRF), which we expected to show the modest informant agreement ($r's \leq .30$) often found in this area (De Los Reyes & Kazdin, 2005). We also use a measure that assesses the likelihood of encountering several social events (e.g., adult instruction, peer provocation) and the conditional likelihood of children's aggressive and withdrawn reactions to each event. We then explore how two sets of factors—the properties of the conditioning events themselves, and the frequency with which they are encountered at home and school—predict the coherence of and agreement between informants' ratings.

We examine four questions. First, we study how informants' ratings of behavior are influenced when the task identifies the events to which the child is responding. Past work has shown the expected increases in behavioral consistency as antecedent events become more similar, using both adult self-reports (Fournier et al., 2008; Van Mechelen, 2009) and

adult observations of children (Mischel & Shoda, 1995; Shoda et al., 1993). The latter work defined what we term “event similarity” based on the interactant (peer vs. adult) and the valence of the interactants’ behavior (e.g., peer talk vs. tease; adult praise vs. warn). Although these events seem relevant to parents’ and teachers’ ratings at home and school, whether the event similarity effect generalizes to informant discrepancy research is unknown. Shoda et al.’s (1993) results were based on short observation periods (hourly activities), whereas cross-informant research relies on retrospective ratings over longer periods (typically months). Moreover, the earlier work obtained observations within the same setting, from adults whose general roles with the children were similar. Based on this, and on evidence that teachers are sensitive to social events even when rating children over longer intervals (Wright et al., 2001), we expected the event similarity effect to be present *within* their perspective and setting (school) where role, relationship, and social interactants remain constant. Thus, teachers’ ratings of children’s reactions to two specific events should be weakly related when those events differ in both interactant and valence, and most strongly related when they share both. Likewise, within their own perspective and setting, we expected parents’ ratings to show a parallel event similarity effect.

Second, we test how the similarity of eliciting events is related to agreement *between* parents and teachers. On the one hand, specification of events could reduce differences between parents and teachers in the situations they spontaneously bring to mind when rating behavior (see De Los Reyes & Kazdin, 2005). For example, even when both raters have knowledge of a child’s reactions to peers and adults, parents may be likely to bring to mind certain interactions with adults (e.g., a parent), whereas teachers may bring to mind interactions with peers that are more common in classrooms (Wright & Zakriski, 2001). On the other hand, specifying events may have only a modest effect on agreement when the perspectives and settings differ as much as do parents’ and teachers’. Children have different relationships with parents and teachers, and may behave differently even in response to the “same” situation (e.g., adult instruction) with one adult versus the other (Drabick et al., 2008; Kraemer et al., 2003; Noordhof, Oldehinkel, Verhulst & Ormel, 2008). Parents and teachers may also interpret the same responses differently; for example, one may encode failure to comply as opposition and the other may encode it as anxiety about task competency (Drabick et al., 2008; Ferdinand, van der Ende, & Verhulst, 2004). On balance, this evidence led us to predict that agreement between parents and teachers would increase with event similarity, but that this effect would be weaker than the anticipated effect within rater perspective.

The event similarity effect just described deals with the properties *of* eliciting events, which is distinct from equally important questions about how often children *encounter* those events in their interactions with peers and adults. Some research has examined how cross-informant discrepancies are predicted by single variables such as the amount of conflict or parental acceptance at home (Grills & Ollendick, 2003; Kolko & Kazdin, 1993). Other work has shown how teachers’ frequency ratings of multiple social events (e.g., peer provocation, teacher praise) can be used to clarify certain syndrome groups, but that work did not examine parent-teacher discrepancies (Wright & Zakriski, 2003). The present research integrates and extends both approaches. We describe how children’s social environments at home and school can be studied as a multivariate pattern of event frequencies—using adult and peer events that parallel the ones we use to study children’s reactions. With these patterns of event frequencies, we show how the similarity of children’s home and school environments can be operationalized. We then test our third hypothesis that home-school similarity predicts informant agreement. For both syndromal ratings and event-specific reaction ratings, we expected parent-teacher agreement to be highest when home and school environments were most similar. We expected that our predictive power would increase as more event rates were used to assess the similarity of children’s environments, and that it

would be especially good when we used events involving interpersonal conflict (e.g., peer provocation, adult discipline) to assess home-school similarity.

Finally, we examine the hypothesis that parent-teacher discrepancies emerge in part from the different behaviors they observe, attend to, and perceive as problematic within their perspective and setting (De Los Reyes, Goodman, Kliewer, & Reid-Quinones, 2008; Youngstrom, Loeber, & Southamer-Loeber, 2000). We expected teachers' syndromal ratings to be predicted by a relatively wide range of their ratings of children's reactions to peer and adult events, reflecting the range of teachers' observations at school. We also expected teachers' syndromal ratings to be predicted especially well by their ratings of children's reactions to peer events because of the importance of those reactions to classroom behavior management. Parents' syndromal ratings should be predicted especially well by their ratings of reactions to adult events, particularly adult instruction or discipline, which are important to behavior management at home.

Method

Parents and teachers provided assessments as part of a larger study of children's response to summer residential treatment for behavioral, academic, and social skills problems (see Zakriski et al., 2005, for additional program description). Data were collected before treatment over two consecutive years. To ensure that research participation did not affect access to service, informed consent documents were sealed until admissions decisions were made. With most admissions in May–June, it was difficult to reach teachers before the end of school, but we attempted for all 260 children admitted before the last week of June. We obtained complete parent data on 85% of these cases ($N = 222$), and complete teacher data on 60% ($N = 157$). The intersection of these sets ($N = 123$) became the cross-informant sample. Comparisons of these children with those who had only complete teacher data ($N = 34$) or only complete parent data ($N = 99$) yielded no significant differences for age, sex, or TRF/CBCL behavior problems, respectively. Parents and teachers were paid \$20 for participation; teachers assessing another child in the second year were paid \$30.

The sample was 52% White, 28% African American, 13% Hispanic, 5% mixed, 1% Asian American, and 1% Native American. Children ranged in age from 8 to 18 ($M = 13.30$ years, $SD = 2.54$); 81 (66%) were boys and 42 (34%) were girls. Teacher assessments were completed by teachers (63.6%), special educators/counselors/therapists (29%), or teaching assistants (7.4%). These raters spent 3.41 hours/day with the student ($SD = 1.91$) over 20.28 months ($SD = 13.38$) in classes that contained 13.32 students ($SD = 5.86$) and 2.07 adults ($SD = .84$). Participant age was not significantly related to interaction time, time known, or class composition.

Materials

The 118-item CBCL and TRF (Achenbach & Rescorla, 2001) assesses overall behavior problems using 8 narrow syndromes and two broad ones (externalizing, internalizing). Ratings are made on a 0–2 scale (0=not true, 1=somewhat or sometimes true, 2=very true). We used the aggression and withdrawal scales because our contextual assessment focused on these behaviors.

The Behavior-Environment Transactional Assessment (BETA; Hartley, Zakriski, Wright, & Parad, 2009; see Wright & Zakriski, 2001) is a 134-item instrument based on observations of children in treatment (Zakriski et al., 2005). All items are rated on a scale of 0 (never) to 5 (almost always). We used 48 “reaction” items that assess the likelihood of aggressive or withdrawn behavior if some event occurs. Informants read the overall instruction, “Please rate how this child reacts to the event described”, followed by one of eight event prompts

(e.g., “If a peer teases or bosses this child...”). For each event prompt, they then rate how often the child shows a specific reaction (e.g., “he/she hits, pushes, or physically attacks”). The eight event prompts are: “If a peer talks to this child in a friendly or supportive way,” “If a peer asks or tells the child to do something,” “If a peer argues or quarrels with this child,” “If a peer teases or bosses this child,” “If an adult talks in a friendly or supportive way to this child,” “If an adult gives instructions or directions to this child,” “If an adult warns or reprimands this child,” and “If an adult disciplines or punishes this child.” Aggressive reactions are “argues or quarrels,” “whines or complains,” “teases or bosses,” and “hits, pushes, or physically attacks.” Withdrawn reactions are “withdraws or isolates self” and “looks sad or cries.” For all analyses, the four aggressive reactions were averaged to form an aggressive reaction composite for each event, as were the two withdrawn reaction items. We also used eight additional “event” frequency items from the BETA. Informants are instructed, “In general, how often do peers or adults do the following things to this child.” They then rate each of the eight events listed above (e.g., “Peers tease or boss this child.”). The remaining BETA items were not used, either because they assess prosocial reactions not related to the CBCL/TRF, or because they assess reciprocal responses to child behavior (e.g., “If this child withdraws...” how often do “peers talk in a friendly way to him”?).

Preliminary Analyses

We first examined the internal consistency of aggression and withdrawal scales for the CBCL and TRF. Because CBCL/TRF aggression scales differ in item content, we computed alpha coefficients separately. The parent and teacher alphas (see Table 1) did not differ significantly (Feldt & Kim, 2006), and were comparable to those reported by Achenbach and Rescorla (2001). Using their age and gender norms, raw scores were converted to clinical *T*-scores. Means and standard deviations for parents and teachers for CBCL/TRF aggression scales were 70.54 (*SD* = 9.85) and 70.49 (10.85), respectively, and for withdrawal they were 64.49 (10.52) and 62.04 (7.97). Withdrawal means differed, $t(122) = 2.64, p < .01$, but not aggression, externalizing or internalizing. Percentages of cases meeting the clinical cutoff (*T*-score ≥ 70) were 52.8% versus 41.5% for parent and teacher aggression, respectively, and 26.8% versus 11.4% for withdrawal.

“Event similarity”—We performed preliminary analyses needed to test how the similarity of specific events was associated with parent-teacher agreement for children’s aggressive and withdrawn reactions. Recall that “event” refers to a specific situation on the BETA that might elicit a behavior (e.g., “if a peer teases”). To replicate Mischel and Shoda (1995), we used ratings of reactions to each of the eight events (e.g., withdrawal in response to adult instruction; aggression in response to peer argue; and thus eight reaction scales per behavior). Before proceeding, we tested for differences between parents’ and teachers’ alphas; we set the Type I error rate to .01 because there were 16 comparisons; none was significant. For parents and teachers combined, the alphas for aggressive reactions ranged from .69 (to peer argue) to .85 (to peer talk), with a median of .78. For withdrawn reactions, the alphas ranged from .74 (to adult discipline) to .83 (to peer tease), with a median of .79.

As in Mischel and Shoda (1995), each event has two features: person (adult or peer interactant) and valence (nonaversive or aversive). A similarity index for within-rater agreement of 0, 1, or 2 captures this. For example, peer tease and adult talk share 0 features, peer tease and peer talk share 1 (person), peer talk and adult talk share 1 (valence), and peer tease and peer argue share 2 (valence and person). For parent-teacher agreement, an event can also be “identical” (e.g., “peer tease” for both raters), indexed as a 3. Aggressive (or withdrawn) reactions to events were inter-correlated using Pearson’s *r*. For “within-rater” analyses of parents (or teachers), this yielded the lower half of the 8×8 correlation matrix, or

28 *rs* (Note that each entry in the upper half of the matrix is identical to the corresponding entry in the lower half. Each entry on the diagonal is an autocorrelation and hence $r = 1.0$.) For “between-rater” analyses, this yielded the full 8×8 matrix, or 64 *rs*. (Note that each *r* conveys unique information. Each entry on the diagonal is the correlation between parents’ and teachers’ ratings of children’s reactions to “identical” events). In the Results section, we examine whether within- and between-rater “agreement” (i.e., *r*) increased as a function of event similarity.

Remaining analyses aggregated aggressive and withdrawn reactions over related events to reduce the number of variables and to simplify the presentation. This resulted in four aggressive reaction scales, each comprised of eight BETA reaction items: aggression to peer aversive events (aggression to peer argue combined with aggression to peer tease), to peer nonaversives (peer talk, peer ask), to adult aversives (adult warn, adult discipline), and to adult nonaversives (adult talk, adult instruct). There were four parallel withdrawn reaction scales (each comprised of four original BETA reaction items). Four event rate scales were also used: peer/adult aversive event rates and peer/adult nonaversive event rates (each comprised of two BETA event frequency items). As before, we computed alphas for parents and teachers separately, and tested for differences. Only two were significant: parent versus teacher ratings of aggression to peer aversives (.80 vs. .89, respectively, $p < .005$), and rates of adult aversive events (.83 vs. .93, $p < .001$).¹ Therefore, Table 1 gives alphas for reaction scales for parents and teachers combined. Alphas for the event rate scales were: adult aversives (.88), peer aversives (.70), adult nonaversives (.67), and peer nonaversives (.52). Caution will be needed regarding analyses of nonaversive peer events.

Before standardizing within age and gender (as on the CBCL/TRF), we performed ANOVAs on raw BETA ratings to examine age and gender differences. We also examined whether reaction ratings varied over events, as this would clarify whether raters attended to event cues. The mixed-model ANOVAs used age (< 12 vs. ≥ 12 yrs.) and gender as grouping variables, and rater and event type as repeated measures. To avoid over-fitting, we restricted the model to main effects and two-way interactions. Greenhouse-Geiser adjustments were used. As expected, the mean level of children’s aggressive reactions varied over events, $F(3,360) = 163.49$, $p < .001$, $\eta_p^2 = .58$, ranging from 1.31 ($SD = .79$) (aggression to adult nonaversives) to 2.42 ($SD = 1.00$) (to peer aversives); informants clearly attended to and were influenced by event cues on the BETA. Young children reacted more aggressively than old, $F(1,120) = 9.09$, $p < .01$, $\eta_p^2 = .07$ ($M_y = 1.99$, $M_o = 1.61$), especially to peer aversives, $F(3,360) = 6.60$, $p < .01$, $\eta_p^2 = .05$. Withdrawn reactions also varied in their mean levels over events, $F(3,360) = 89.84$, $p < .001$, $\eta_p^2 = .43$, ranging from 0.96 ($SD = .85$) (withdrawal to peer nonaversives) to 1.85 ($SD = 1.19$) (to adult aversives). The frequency of these eliciting events also varied, $F(3,366) = 192.78$, $p < .001$, $\eta_p^2 = .61$, from 1.97 (peer aversives) to 4.13 (adult nonaversives). Parents reported more adult aversives (3.01) than teachers did (2.69), $F(3,366) = 5.35$, $p < .01$, $\eta_p^2 = .04$.² To parallel the CBCL/TRF, henceforth we use age- and sex-standardized BETA ratings (*z*-scores).

¹We also tested for possible differences between CBCL/TRF scales and the corresponding parent/teacher BETA reaction scales. For aggression, the only difference was that CBCL aggression had a higher alpha than parent BETA aggression to peer aversives (.89 vs. .80, respectively, $p < .002$). For withdrawal, all coefficients for BETA reaction scales were higher than their CBCL/TRF counterparts (the smallest difference .79 vs. .88; p 's $< .01$).

²Parallel analyses for reactions and events using the 8 individual events yielded main effects for events that resembled those reported using 4 event categories. Briefly, η_p^2 s were .56 for event frequencies, .57 for aggressive reactions, and .43 for withdrawn reactions. We also checked whether the means differed for the pairs of events that were most similar; every pair differed by Tukey’s Honestly Significant Difference test, either for event frequency ratings (peer argue vs. peer tease), for ratings of aggressive/withdrawn reactions to events (to adult talk vs. to adult instruct), or both (adult warn vs. adult discipline; peer talk vs. peer ask).

“Environment similarity”—To analyze the similarity of children’s social experiences at home and school we used aggregated ratings of how often children encountered peer nonaversive, peer aversive, adult nonaversive, and adult aversive events. Thus, each child had a vector of four event rates as reported by parents, and a vector of four event rates as reported by teachers (each standardized within rater). A child’s similarity score was the sum of the squared deviations between the vectors, or “Euclidean” distance (Borg & Groenen, 1997). To examine how this environment similarity measure was associated with cross-informant agreement, we then split children at the median into 2 groups (low vs. high similarity). For each group we calculated cross-informant r ’s (e.g., for CBCL/TRF aggression or for BETA aggression to adult aversives).

Note that this procedure computes environment similarity using all four event scales. To examine whether the relative breadth of the similarity measure was associated with rater agreement, we also computed environment similarity using less than the full vectors. For example, one possible subset of two events involves parents’ (and teachers’) ratings of peer aversives and peer nonaversives; another subset involves their ratings of peer aversives and adult aversives. In all, there are 6 possible subsets of 2 events per subset. In this way, we identified all possible subsets of 1 event (there are 4 subsets), of 2 events (6 subsets), and 3 events (4 subsets). For each subset, similarity was re-calculated using the method described above, similarity groups formed using a median split, cross-informant r s computed, and finally averaged within each level of breadth. A second set of analyses calculated similarity for individual event scales (e.g., frequency of peer aversives). Informant agreement was then calculated for children who were more or less similar in how often they experienced that type of interaction. Over the multiple samples generated, the median N s for low/high-similarity groups were 62/61, respectively.

Results

Syndromal Agreement, Reaction Agreement, and Similarity of Conditioning Events

For comparison with other cross-informant research, we assessed informant agreement using the CBCL/TRF. As expected, agreement (r) between parents and teachers was modest (see Table 1). For comparison, Achenbach and Rescorla (2001) report r s of .33 and .24 for withdrawal and aggression, respectively.

We next examined the hypothesis that informant agreement would improve when parents and teachers rated children’s reactions to specific conditioning events. To facilitate comparisons with results for (aggregated) syndromal ratings just reported, we used aggregated reactions (e.g., aggression to peer aversives). As shown in Table 1, agreement remained modest. All r s were positive, but none was significantly different from the r for CBCL/TRF agreement.

The preceding results show that informant agreement was modest for reactions to identical events, but they do not provide a complete test of the event similarity hypotheses. First, as we have noted, the similarity of events could play a large role *within* rater or setting (e.g., parent/home) but less of a role when two different settings are involved (home vs. school). Second, although cross-informant agreement was modest for identical events (see Table 1), it could be even lower (or negative) for dissimilar ones. Each of these results would clarify whether and how the event similarity effect applies to research on informant discrepancies.

Figure 1 provides the relevant results. The left panel shows the mean r s for aggressive reactions for each similarity level (i.e., number and type of features shared). As hypothesized, agreement within rater increased with event similarity. To summarize this effect, we predicted the pairwise r s using 0, 1, or 2 to index event similarity, as previously

noted. For parents, r increased with similarity, $F(1, 26) = 27.19, p < .001, R^2 = .51$ (see Figure 1, left). The same was true for teachers, $F(1, 26) = 13.54, p < .001, R^2 = .34$. Parallel results were found for withdrawal (Figure 1, right): r increased with similarity within parents, $F(1, 26) = 29.02, p < .001, R^2 = .53$, and teachers, $F(1, 26) = 12.36, p < .01, R^2 = .32$.

Figure 1 also examines whether event similarity was related to agreement *between* parents and teachers. We predicted the pairwise r s using the between-rater 0, 1, 2, or 3 similarity index (see Method section); as previously noted, “3” reflects pairs for which parents and teachers rated reactions to an “identical” event (e.g., “peer tease”). Event similarity had no effect on cross-informant agreement; all mean r s were below .20.³

Similarity of Experienced Home and School Environments

Clearly, asking parents and teachers to rate children’s reactions to events does not mean children encounter those events equally often at home and school. Indeed, cross-informant agreement for ratings of how often children encountered events resembled what is found for syndromal measures: peer nonaversives ($r = .16, p < .10$), adult nonaversives ($.14, p < .10$), adult aversives ($.24, p < .01$), and peer aversives ($.38, p < .001$). We have noted how the similarity of children’s experiences at home and school can be defined using rates of events (see Method). We now test the hypothesis that informant agreement will be higher for children whose home and school environments are similar, especially as more events are used to assess similarity.

The results of environment similarity analyses for CBCL/TRF aggression are shown in Figure 2 (top). The cross-informant r presented earlier without considering environment similarity is included (see “All Cases”). The abscissa indicates the number of events used to compute environment similarity, or “breadth” (see Method). For high similarity groups, cross-informant agreement increased with breadth, and was significant at each level. For low similarity groups, agreement decreased with breadth and was never significant. The differences between high and low similarity groups were significant when 3 or 4 events were used, $z_s > 2.16, p_s < .05$.⁴

The same approach was used for CBCL/TRF withdrawal. There were no significant pairwise differences between the low and high similarity groups, regardless of number of events used to compute environment similarity. Although this dictates caution, we note that the only cross-informant correlations that differed from zero were for children with *dissimilar* environments; this occurred at all levels of breadth ($r_s = .28-.39, p_s < .05$).⁵

³We checked the possibility that cross-informant agreement for event-specific reactions might be low because some events did not occur often enough for raters to provide a meaningful response. For example, if a rater states that a child is rarely teased, ratings of the child’s reaction to teasing may not be informative. To examine this, we repeated the analyses just described, each time removing reaction ratings for which a rater gave the corresponding event a frequency of less than “1”, “2”, or “3”. Cross-informant results were similar to those already reported, with one exception: For aggressive reactions, we found small but reliable increases with similarity, with R^2 of .07, .09, and .14, for the 3 frequency thresholds, respectively, $F_s(1,54) > 4.89, p_s < .04$. Cross-informant agreement remained low, with a maximum r of .21, at “identity” for the highest frequency threshold.

⁴One might argue that environment similarity is an indirect measure of children’s behavior (e.g., children with similar environments might be more aggressive than those with dissimilar ones). To examine this, we correlated the similarity measure (based on 4 events) with children’s aggregated aggression (i.e., average of TRF and CBCL). Similarity showed little relationship with overall aggression or withdrawal ($r_s = .14, .17$, respectively, *ns*). Our results indicate that children with similar environments were more *consistent* in their aggression across settings, but they could be either low or high in their overall aggression (or withdrawal).

⁵To examine heterogeneity within groups, exploratory cluster analyses were conducted. Some children were in the “low similarity” group because they often had aversive encounters with peers at school but rarely at home; for others, home and school differed primarily in the rates of nonaversive adult events. Children in the “high similarity” group also showed a variety of event profiles (e.g., high peer aversives at home and school; high adult aversives in both). It will be important in future research with larger samples to examine these functional subgroups.

The previous analyses do not examine whether similarity in the rate of encountering aversive events is especially useful in predicting agreement. As shown in Figure 2 (bottom), agreement for CBCL/TRF aggression was higher for the high- versus low-similarity group when environment similarity was based on either peer or adult aversives, $z > 2.02$, $ps < .05$. Thus, even when narrowly defined using rates of aversive events, environment similarity predicted agreement. Recall that these events were the most reliably assessed by parents and teachers.

The same analyses were conducted for withdrawal. Although some individual cross-informant correlations were significant, parent-teacher agreement for withdrawal did not differ significantly for any of the low- vs. high-similarity comparisons. Thus, we found little evidence that informant agreement for withdrawal was associated with specific environment similarities.

We performed parallel analyses using ratings of children's reactions to events. As Figure 3 shows, the results for aggression to aversive events resembled those for CBCL/TRF aggression. For multiple-event analyses (top row), environment similarity effects were strongest for reactions to peer aversives. For single-event analyses (bottom), environment similarity effects were clear for aggression to both types of aversives, and strongest when similarity was based on adult aversives. Results for withdrawn reactions resembled those for CBCL/TRF withdrawal: few informant correlations were significant (some for children with similar and some for dissimilar environments), and agreement did not differ between low-versus high-similarity groups.

Predicting Syndromal Ratings from Event-Specific Reactions

Finally, we tested hypotheses noted earlier about the contribution of event-specific reactions to the prediction of parents' and teachers' syndromal ratings. For example, for CBCL aggression, the predictors for the multiple regression were parents' ratings of children's aggressive reactions to each of four events (Table 2). Aggression to adult aversives was predictive for both CBCL and TRF aggression, and aggression to peer positives was also predictive for teachers' ratings of TRF aggression. The vector of 4 coefficients for parents (row 1) correlated .53 with the vector for teachers (row 2); none of the pairwise comparisons of slopes (parent vs. teacher) was significant. Withdrawal to peer nonaversives was predictive for both CBCL and TRF withdrawal, and withdrawal to adult aversives was also predictive for parents' ratings of CBCL withdrawal; the vectors of coefficients (rows 3 vs. 4) correlated $-.18$; there was one significant pairwise comparison for withdrawn reactions to adult aversives (warning and discipline), $t(236) = 2.58$, $p < .02$. The vector of coefficients for parent aggression (row 1) correlated .80 with their vector for withdrawal (row 3); for teachers, the parallel r (rows 2 vs. 4) was .01.

Discussion

The current study underscores the robustness of informant discrepancies and illustrates how our understanding of them can be deepened by an analysis of the social events children experience and how they react to them. As expected, parent-teacher agreement for aggression and withdrawal on a widely-used syndromal measure was modest (r 's $\leq .30$). We predicted that parents' and teachers' ratings of reactions to specific events would show greater agreement, but they did not. The similarity of eliciting events had the expected effect on consistency of reaction ratings *within* a given rater perspective, but not on agreement *between* parents and teachers. Agreement was low not only when parents and teachers rated the same behavior in response to different events (e.g., aggression to adult instruction vs. to peer teasing), but equally low when events were nominally identical. In contrast, the similarity of children's social environments—defined in terms of how often children

encountered events at home versus school—was linked to informant agreement for ratings of aggression. Predictions of parents' versus teachers' syndromal ratings by event-specific reactions were more similar than expected, but subtle differences emerged. In what follows, we examine interpretations and implications of these findings for informant discrepancy research and clinical practice.

Our findings demonstrate the reliability and internal organization of raters' contextualized assessments, but they also demonstrate how difficult it is to bridge the gap between informants in different settings. Past research on personality and clinical assessment led us to expect that specifying the eliciting event might clarify the assessment task, capture meaningful variability in children's behavior, and at least modestly improve informant agreement. Nevertheless, informant agreement remained low. This finding would be unremarkable if raters ignored the event cues they were given, but they did not. Mean reaction ratings varied considerably over the four event categories (with η_p^2 s of .43–.58), showing that raters attended to the conditions that elicit behaviors. Even within the pairs of individual events that were most similar (e.g., adult warn vs. adult discipline), raters distinguished between events, either in terms of their frequency and/or in terms of children's reactions to them. Furthermore, analyses revealed the expected event similarity effect *within* rater. The more features (interactant, valence) two events shared, the greater the consistency of the rated reactions to those events. These results extend past research by showing how the event similarity effect found using direct observations of behavior (Mischel & Shoda, 1995) applies to retrospective ratings within a given setting, yet also show that the effect does not apply when parents and teacher rate children's behavior at home versus school.

It is notable that cross-informant agreement remained modest for ratings of reactions to aversive events. Like others (Shoda et al., 1993; Zakriski et al., 2005), we found that problem behaviors were more common and somewhat more variable in response to aversive events (e.g., peer tease, adult discipline), mitigating against floor and restricted range effects. Others have argued that stressful situations engage preferred and stable coping strategies (Hood, Power, & Hill, 2009; Parker & Wood, 2008). Moreover, an aversive "event" (e.g., adult warn) is sometimes a response to a problem behavior previously displayed by the child (Burke, Pardini, & Loeber, 2008). If aggressive children show consistent "bursts" when disciplined (Granic & Patterson, 2006), one would expect better agreement for ratings constrained to this context. These arguments notwithstanding, agreement remained low for reactions to aversive events. The issue is not simply that cross-informant agreement was low in some absolute sense, which researchers have known for some time. Rather, the issue is that parent-teacher agreement was unaffected by factors that one would expect to affect it, and that did affect it within perspective.

Although assessing behavior in response to eliciting events could not bridge the cross-informant gap by itself, attention to how often children *encountered* those events in their interactions was useful. Cross-informant agreement for CBCL/TRF aggression was higher for children whose home and school environments were more similar, and the magnitude of this effect increased as more aspects of children's social experiences were used. Similarity even for single events—namely, conflict with peers or adults—predicted informant agreement. Similarity based on adult instruction and conversation was less useful, and similarity based on peer requests and conversation was even less so. Reliable assessment of nonaversive experiences, especially for peers, deserves attention in future research. Environment similarity effects were also observed for the already-contextualized reaction ratings. When home and school were more similar, we found better parent-teacher agreement for aggression to adult and peer conflict. Similarity in how often children were disciplined was especially useful in predicting agreement for aggression to that event.

These results are consistent with the view that nominally similar events children experience in everyday social interaction (e.g., “peer tease” or “adult discipline”) can be interpreted differently by informants at home and school (Drabick et al., 2008; Ferdinand et al., 2004). The particular peers and adults differ, and the nuances of their teasing, disciplining, or other actions differ as well. Such variation in the meaning of events is, arguably, part and parcel of the larger phenomena of cross-situational variability and informant discrepancies. Nevertheless, it remains possible that further specification of events would improve informant agreement beyond what we found. It is also possible that parents and teachers believe that children’s behavior is cross-situationally variable even when they are given precisely the “same” instruction or discipline by different adults. Future laboratory research examining informants’ perceptions of children’s responses to specific events and interactants would be especially helpful (see De Los Reyes et al., 2009). We hope our apparent focus on “improving” agreement will not be misinterpreted; the larger goal should not be to reduce informant discrepancies, but to measure them well and to understand when and why they do or do not occur.

We found little evidence that cross-informant agreement for withdrawal increased with the similarity of home and school environments. This could be because our sample had fewer clinically withdrawn children, because many of the children with elevated withdrawal were comorbidly elevated for aggression (76% by parent report; 64% by teacher report), or because parents reported higher levels of withdrawal than teachers. Withdrawal was as reliably assessed as aggression for both the CBCL/TRF and the BETA, but for each measure fewer withdrawn behaviors were assessed than for aggression. It is also possible that the social events that predict agreement for withdrawal are more subtle and difficult to assess than are the vivid aversive events we found to be useful in predicting agreement for aggression. This converges with the finding that internal consistency and informant agreement were higher for aversive experiences than for the nonaversive ones (e.g., adult instruction, peer conversation).

Although our home-school similarity measure used a familiar metric (Borg & Groenen, 1997), it should be interpreted with care, especially in light of our “dissimilarity” findings for withdrawal. A variety of patterns could (and did) produce comparable (dis)similarity scores. For example, some children who were seen as withdrawn by both their parents and their teachers experienced a specific type of environment *dissimilarity* (e.g., they encountered more peer aversives at school than at home). A detailed analysis of the distinct event patterns experienced by children with low- or high-similarity environments was beyond what our sample size would allow, but this may be a fruitful issue for future research.

Our regression analyses are relevant to discussions about the nature of parents’ and teachers’ syndrome ratings and the sources of discrepancies between them (De Los Reyes, et al., 2008, Drabick et al., 2008; Youngstrom, et al., 2000). Parents’ and teachers’ syndrome scores could be predicted from children’s event-specific reactions ($R^2=37\%–54\%$), and, as expected, parents’ syndromal ratings of withdrawal were predicted better by withdrawn reactions to adult discipline than were teachers’. This effect was narrower than predicted and was not found for aggression. We did not find clear evidence that teachers’ ratings were better predicted than parents’ by reactions to peers. The overall similarity of parent and teacher regressions dictates caution, but the pattern of findings raised questions that may be worth exploring in future research. One relates to the possible importance of reactions to adult discipline in predicting parents’ ratings of problem behavior. Another relates to the role of children’s withdrawn reactions to positive peer interactions in predicting teachers’ ratings of problem behavior. Interviews and vignette methodologies could be used to further explore how raters prioritize or bring to mind certain interpersonal situations when rating

behavior, and whether they sample from the same or different sets of situations depending on the type of behavior they are asked to assess.

Taken as a whole, our results support the idea that informant discrepancies result from meaningful behavioral variability and are more than measurement error. For readers familiar with debates in the personality literature (Mischel, 2009) this statement may create a sense of déjà vu (Mischel & Peake, 1982) all over again (Roberts & Caspi, 2001). Indeed, a critic might ask whether aspects of this history are repeating themselves in the informant discrepancy literature. Achenbach and colleagues (1987) argued in their influential meta-analysis that informant discrepancies result from children's behavioral variability across settings, yet, over twenty years later, this message has not been fully absorbed. Researchers continue to use measures that are rooted in a nomothetic trait tradition (see Dumenci, Achenbach, & Windle, 2010) and that do not explicitly incorporate psychosocial contexts into the core of the measurement process (see Block, 1995; 2010). To make progress, informant discrepancy researchers need to be alert to the notion, popularized in textbooks, that "consistency across situations lies at the core of the concept of personality" (Weiten, 2004, p. 478). Alternative, more contextualized concepts of personality do not require consistency in this sense, do not insist that complex patterns of behavior be aggregated across situations in order to arrive at a reliable index of a trait (Roberts & Caspi, 2001), and thus may be better able to incorporate the kinds of discrepancies found in the cross-informant literature.

A related suggestion is to note the assumptions on which our measurements are based. To paraphrase Hotelling, Bartky, Deming, Friedman, and Hoel (1948) we sometimes choose our measures as we say our prayers—because they are found in highly respected books written a long time ago. The measures used widely in child assessment have many strengths, but they do not come with a guarantee that they will measure only the "person" merely because they ask the rater about his or her "behavior". We should study not only the broad factors that predict syndromal measures or their differences across settings, but also the narrow micro-contexts that may be implicit in the mind of the rater when rating individual acts, and that give meaning to their ratings and disagreements about them. We should also strive to assess the "environment" with the same rigor that we assess the "child" (e.g., Moos & Moos, 1990). All of this may appear at first to be a distraction from the immediate problem of understanding why parent and teacher assessments are often so different, but our own view is that genuine progress will require a deeper understanding than we now have of just how intertwined behaviors and the surrounding social situation really are.

Acknowledgments

We are deeply grateful to the parents, teachers, children, staff, and administrators whose cooperation made it possible to collect the data reported here. We would especially like to thank Harry Parad, Director of Wediko Children's Services, and Diana Parad, Director of the Wediko Boston School-Based Program, whose support made this work possible. We would like to acknowledge the invaluable contributions of Stephanie Cardoos who coordinated the Wediko Transitions Project, the parent project for this study. Thank you also to Sophia Choukas-Bradley and Lindsay Metcalfe for their devoted research assistance. This research was supported by award number R15MH076787 from the National Institute of Mental Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

References

Achenbach TM. As others see us: Clinical and research implications of cross-informant correlations for psychopathology. *Current Directions in Psychological Science* 2006;15:94–98.

- Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin* 1987;101:213–232. [PubMed: 3562706]
- Achenbach, TM.; Rescorla, LA. *Manual for the ASEBA School-Age Forms & Profiles*. Burlington: Research Center for Children, Youth, & Families, University of Vermont; 2001.
- Barkley, RA. Child behavior rating scales and checklists. In: Rutter, M.; Tuna, AH.; Lann, IS., editors. *Assessment and diagnosis in child psychopathology*. New York: Guilford Press; 1988. p. 113-155.
- Beck KH, Hartos JL, Simons-Morton BG. Relation of parent-teen agreement on restrictions to teen risky driving over 9 months. *American Journal of Health Behavior* 2006;30:533–543. [PubMed: 16893316]
- Block J. A contrarian view of the five-factor approach to personality description. *Psychological Bulletin* 1995;117:187–215. [PubMed: 7724687]
- Block J. The five-factor framing of personality and beyond: Some ruminations. *Psychological Inquiry* 2010;21(1):2–25.
- Borg, I.; Groenen, PJF. *Modern multidimensional scaling: Theory and applications*. New York: Springer; 1997.
- Burke JD, Pardini DA, Loeber R. Reciprocal relationships between parenting behavior and disruptive psychopathology from childhood through adolescence. *Journal of Abnormal Child Psychology* 2008;36:679–692. [PubMed: 18286366]
- Buss D, Craik KH. The act frequency approach to personality. *Psychological Review* 1983;90:105–126.
- Cervone D, Shadel WG, Jencius S. Social-cognitive theory of personality assessment. *Personality and Social Psychology Review* 2001;5:33–50.
- De Los Reyes A, Goodman KL, Kliewer W, Reid-Quinones K. Whose depression relates to discrepancies? Testing relations between informant characteristics and informant discrepancies from both informants' perspectives. *Psychological Assessment* 2008;20:139–149. [PubMed: 18557691]
- De Los Reyes A, Henry DB, Tolan PH, Wakschlag LS. Linking informant discrepancies to observed variations in children's disruptive behavior. *Journal of Abnormal Child Psychology* 2009;37:637–652. [PubMed: 19247829]
- De Los Reyes A, Kazdin AE. Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin* 2005;131:483–509. [PubMed: 16060799]
- Drabick DAG, Gadow KD, Loney J. Co-occurring ODD and GAD symptom groups: Source-specific syndromes and cross-informant comorbidity. *Journal of Clinical Child & Adolescent Psychology* 2008;37:314–326. [PubMed: 18470769]
- Dumenci L, Achenbach TM, Windle M. Measuring context-specific and cross-contextual components of hierarchical constructs. *Journal of Psychopathology and Behavioral Assessment*. 2010
- Endler NS, Parker JDA, Bagby RM, Cox BJ. Multi-dimensionality of state and trait anxiety: The factor structure of the Endler Multidimensional Anxiety Scales. *Journal of Personality and Social Psychology* 1991;60:919–926. [PubMed: 1865327]
- Feldt LS, Kim S. Testing the difference between two alpha coefficients with small samples of subjects and raters. *Educational and Psychological Measurement* 2006;66:589–600.
- Ferdinand RF, van der Ende J, Verhulst FC. Parent-adolescent disagreement regarding psychopathology in adolescents from the general population as a risk factor for adverse outcome. *Journal of Abnormal Psychology* 2004;113:198–206. [PubMed: 15122940]
- Fournier MA, Moskowitz DS, Zuroff DC. Integrating dispositions, signatures, and the interpersonal domain. *Journal of Personality and Social Psychology* 2008;94:531–545. [PubMed: 18284296]
- Granic I, Patterson GR. Toward a comprehensive model of antisocial development: A dynamic systems approach. *Psychological Review* 2006;113(1):101–131. [PubMed: 16478303]
- Grills AE, Ollendick TH. Multiple informant agreement and the Anxiety Disorders Interview Schedule for Parents and Children. *Journal of the American Academy of Child & Adolescent Psychiatry* 2003;42:30–40. [PubMed: 12500074]

- Guion K, Mrug S, Windle M. Predictive value of informant discrepancies in reports of parenting: Relations to early adolescents' adjustment. *Journal of Abnormal Child Psychology* 2009;37:17–30. [PubMed: 18584134]
- Hartley, AG.; Zakriski, AL.; Wright, JC.; Parad, HW. Understanding sources of cross-informant agreement in the assessment of child psychopathology; Poster presented at the 2009 Society for Research in Child Development Biennial; Denver, CO. 2009 Apr.
- Haynes SN, Mumma GH, Pinson C. Idiographic assessment: Conceptual and psychometric foundations of individualized behavioral assessment. *Clinical Psychology Review* 2009;29:179–191. [PubMed: 19217703]
- Hotelling H, Bartky W, Deming WE, Friedman M, Hoel P. The teaching of statistics. *Annals of Mathematical Statistics* 1948;19:95–115.
- Hood B, Power T, Hill L. Children's appraisal of moderately stressful situations. *International Journal of Behavioral Development* 2009;33:167–177.
- Kolko DJ, Kazdin AE. Emotional/behavioral problems in clinic and nonclinic children: Correspondence among child, parent, and teacher reports. *Journal of Child Psychology & Psychiatry* 1993;34:991–1006. [PubMed: 8408380]
- Kraemer HC, Measelle JR, Ablow JC, Essex MJ, Boyce WT, Kupfer DJ. A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry* 2003;160:1566–1577. [PubMed: 12944328]
- Maccoby, EE. *The two sexes: Growing up apart, coming together*. Cambridge, MA: Harvard University Press; 1998.
- Mischel W. From personality and assessment (1968) to personality science, 2009. *Journal of Research in Personality* 2009;43:282–290.
- Mischel W, Peake PK. Beyond déjà vu in the search for cross-situational consistency. *Psychological Review* 1982;89:730–755.
- Mischel W, Shoda Y. A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review* 1995;102:246–268. [PubMed: 7740090]
- Moos, R.; Moos, B. *Life stressors and social resources inventory preliminary manual*. Palo Alto, CA: Stanford University Medical Center; 1990.
- Noordhof A, Oldehinkel AJ, Verhulst FC, Ormel J. Optimal use of multi-informant data on co-occurrence of internalizing and externalizing problems: The TRAILS study. *International Journal of Methods in Psychiatric Research* 2008;17:174–183. [PubMed: 18763693]
- Parker, DA.; Wood, LM. Personality and the coping process. In: Boyle, G.; Matthews, G.; Saklofske, DH., editors. *The SAGE handbook of personality theory and assessment, Vol. 1: Personality theories and models*. Thousand Oaks, CA: Sage Publications; 2008. p. 506-519.
- Roberts BW, Caspi A. Authors' response: Personality development and the person-situation debate: It's déjà vu all over again. *Psychological Inquiry* 2001;12:104.
- Scotti JR, Morris TL, McNeil CB, Hawkins RP. DSM-IV and disorders of childhood and adolescence: Can structural criteria be functional? *Journal of Consulting and Clinical Psychology* 1996;56:1177–1191. [PubMed: 8991304]
- Shoda Y, Mischel W, Wright JC. Links between personality judgments and contextualized behavior patterns: Situation-behavior profiles of personality prototypes. *Social Cognition* 1993;11:399–429.
- Smith RE, Shoda Y, Cumming SP, Smoll FL. Behavioral signatures at the ballpark: Intraindividual consistency of adults' situation-behavior patterns and their interpersonal consequences. *Journal of Research in Personality* 2009;43:187–195.
- Van Mechelen I. A royal road to understanding the mechanisms underlying person-in-context behavior. *Journal of Research in Personality* 2009;43(2):179–186.
- Van Mechelen I, Kiers HAL. Individual differences in anxiety responses to stressful situations: A three-mode component analysis model. *European Journal of Personality* 1999;13:409–428.
- Vansteelandt K, Van Mechelen I. Individual differences in situation-behavior profiles: A triple-typology model. *Journal of Personality and Social Psychology* 1998;75:751–765.

- Vansteelandt K, Van Mechelen I. Individual differences in anger and sadness: In pursuit of active situational features and psychological processes. *Journal of Personality* 2006;74:873–910.
- Weiten, W. *Psychology: Themes and variations*. 6th ed.. Belmont, CA: Wadsworth; 2004.
- Wright JC, Lindgren KP, Zakriski AL. Syndromal versus contextualized personality assessment: Differentiating environmental and dispositional determinants of boy's aggression. *Journal of Personality and Social Psychology* 2001;81:1176–1189. [PubMed: 11761316]
- Wright JC, Mischel W. A conditional approach to dispositional constructs: The local predictability of social behavior. *Journal of Personality and Social Psychology* 1987;53(6):1159–1177. [PubMed: 3694455]
- Wright JC, Zakriski AL. When syndromal similarity obscures functional dissimilarity: Distinctive evoked environments of externalizing and mixed syndrome boys. *Journal of Consulting and Clinical Psychology* 2003;71:516–527. [PubMed: 12795575]
- Wright JC, Zakriski AL. A contextual analysis of externalizing and mixed syndrome boys: When syndromal similarity obscures functional dissimilarity. *Journal of Consulting and Clinical Psychology* 2001;69:457–470. [PubMed: 11495175]
- Youngstrom E, Loeber R, Southamer-Loeber M. Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology* 2000;68:1038–1050. [PubMed: 11142538]
- Zakriski AL, Wright JC, Underwood MK. Gender similarities and differences in children's social behavior: Finding personality in contextualized patterns of adaptation. *Journal of Personality and Social Psychology* 2005;88:844–855. [PubMed: 15898879]

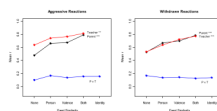


Figure 1. Mean correlations among ratings of event-specific aggressive and withdrawn reactions, as a function of event similarity and source of rating. Teacher = agreement within teachers’ ratings; Parent = agreement within parents’ ratings; P v T = cross-informant agreement between parents and teachers. Person = events sharing person type (peer vs. adult); Valence = events sharing valence (aversive vs. nonaversive); None = events sharing neither; Both = events sharing both; Identity = ratings for identical events by different informants. Asterisks indicate significance of regressions predicting r from event similarity; ** $p < .01$, *** $p < .001$.

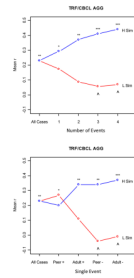


Figure 2. Cross-informant agreement (mean r) between parents' (CBCL) and teachers' (TRF) aggression ratings (AGG), as a function of number and type of events used to assess similarity of children's home and school environments. All Cases = results for all children regardless of environment similarity ($N = 123$). L/H Sim = children with low (L) / high (H) similarity social environments (median split). Top panel shows agreement as a function of the number of events used to assess environment similarity. Bottom panel shows agreement based on specific events used to assess similarity: Peer/Adult + = nonaversive peer/adult events; Peer/Adult - = aversive peer/adult events. Tests of r /Pairwise tests: */ \wedge $p < .05$, ** $p < .01$, *** $p < .001$.

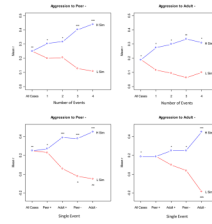


Figure 3.

Cross-informant agreement (mean r) between parents' and teachers' BETA reaction ratings, as a function of number and type of events used to assess similarity of children's home and school environments. All Cases = results for all children ($N = 123$). L/H Sim = children with low (L) / high (H) similarity environments (defined based on median split). Top row shows agreement as a function of the number of events used to assess similarity. Bottom row shows agreement based on specific events used to assess similarity: Peer/Adult + = nonaversive peer/adult events; Peer/Adult - = aversive peer/adult events. Tests of r /Pairwise tests: */^ $p < .05$, **/^ ^ $p < .01$, ***/^ ^ ^ $p < .001$.

Table 1
Internal Consistency and Cross-Informant Agreement for Syndromal Ratings (CBCL/TRF) and for Event-Specific Reactions.

Behavior	CBCL/TRF	Event-Specific Reactions		
		Peer +	Adult +	Peer - Adult -
Internal Consistency ¹				
Aggression	.89/.92	.90	.88	.85 .88
Withdrawal	.78/.79	.88	.88	.90 .90
Cross-Informant Agreement ²				
Aggression	.23**	.18*	.13	.22* .18*
Withdrawal	.30**	.10	.14	.20* .27**

NOTE:

¹ Alpha coefficients based on items within each scale.

² Correlations (Pearson's *r*) between parent and teacher ratings.

Peer/Adult + = nonaversive peer/adult events; Peer/Adult - = aversive peer/adult events. CBCL = Child Behavior Checklist; TRF = Teacher Report Form.

* $p < .05$,

** $p < .01$.

Table 2
 Regressions Predicting Parents' and Teachers' Syndromal Ratings from Event-Specific Reactions.

Dependent Variable	Independent Variable: Event-Specific Reaction				F(4, 118)
	P+	P-	A+	A-	
CBCL AGG	.18	.11	.16	.33**	.44
TRF AGG	.29*	.18	.07	.28*	.54
CBCL WDR	.25*	.11	-.01	.38***	.40
TRF WDR	.39**	-.02	.25	.01	.37

NOTE: AGG = aggression; WDR = withdrawal. CBCL = Child Behavior Checklist; TRF = Teacher Report Form. Predictors for each dependent variable (AGG/WDR) were behaviorally matched reactions. P+ = reactions to peer nonaversives, P- = to peer aversives, A+ = to adult nonaversives, A- = to adult aversives. Entries for predictors are standardized regression coefficients.

* $p < .05$,

** $p < .01$,

*** $p < .001$.