Advancing a Multidimensional, Developmental Spectrum Approach to Preschool Disruptive Behavior

Dr. Lauren S. Wakschlag, Ph.D.,
Feinberg School of Medicine and Institute for Policy Research, Northwestern University

Dr. Margaret J. Briggs-Gowan, Ph.D.,
University of Connecticut

Dr. Seung W. Choi, Ph.D.,
CTB/McGraw Hill and Northwestern University

Dr. Sara R. Nichols, Ph.D.,
Feinberg School of Medicine, Northwestern University

Ms. Jacqueline Kestler, M.P.H.,
Feinberg School of Medicine, Northwestern University

Mr. James L. Burns, M.S.,
Feinberg School of Medicine, Northwestern University

Dr. Alice S. Carter, Ph.D., and
University of Massachusetts–Boston

Dr. David Henry, Ph.D.
Institute for Health Research and Policy, University of Illinois at Chicago

Abstract

Objective—Dimensional approaches are gaining scientific traction. However, their potential for elucidating developmental aspects of psychopathology has not been fully realized. The goal of this paper is to apply a multidimensional, developmental framework to model the normal-abnormal spectrum of preschool disruptive behavior. The Multidimensional Assessment of Preschool Disruptive Behavior (MAP-DB), a novel measure, was used to model dimensional severity across developmental parameters theorized to distinguish the normative misbehavior of early childhood from clinically salient disruptive behavior. The 4 MAP-DB dimensions are: Temper Loss, Noncompliance, Aggression, and Low Concern for Others.

Correspondence to Lauren S. Wakschlag, Ph.D., Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, 633 N. St. Clair Street, 19th Floor, Chicago, IL 60611; lauriew@northwestern.edu.

Method—Parents of a diverse sample of 1,488 preschoolers completed the MAP-DB. Multidimensional item response theory (IRT) was employed for dimensional modeling.

Results—The 4-dimension, developmentally informed model demonstrated excellent fit. Its factor loadings did not differ across demographic subgroups. All dimensions provided good coverage of the abnormal end of the severity continuum, but only Temper Loss and Noncompliance provided good coverage of milder, normatively occurring behaviors. The developmental expectability and quality of behaviors distinguished normative from atypical behaviors. The point at which frequency of behaviors was atypical varied based on dimensional location for Temper Loss, Noncompliance, and Aggression.

Conclusion—The MAP-DB provides an innovative method for operationalizing developmentally specified, dimensional phenotypes in early childhood. Establishing the validity of these dimensional phenotypes in relation to clinical outcomes, neurocognitive substrates and etiologic pathways will be a crucial test of their clinical utility.

Disruptive behavior (DB) plays a central role in developmental sequences of psychopathology. It is antecedent to up to 60% of common mental disorders across the lifespan,[1,2] often emerges in early childhood,[3] and is the most prevalent disorder of the preschool period.[4] One reason for DB's centrality to both internalizing and externalizing disorders is the heterogeneity of its defining features. For example, irritability is a key feature of oppositional defiant disorder (ODD) and is also central to a number of other disorders with disrupted emotion regulation (e.g., depression).[5] In contrast, aggressive behaviors, particularly callous aggression, are associated with a distinct antisocial pathway.[6] The utility of a multidimensional approach to ODD for clinical prediction has been robustly demonstrated: disaggregation into dimensions such as "irritable," "headstrong," and "hurtful" differentiates internalizing and externalizing patterns and their co-occurrence.[7–11] Conduct disorder (CD) subtypes, including aggressive and nonaggressive rule breaking, and callousness, also distinguish varied clinical risk profiles.[12,13] Parsing the heterogeneity of *emergent* disruptive behavior may illuminate early markers of divergent developmental and clinical pathways. In particular, modeling the dimensional structure of early childhood DB is important for mapping the prodromal phases of clinical patterns, linking them to underlying mechanisms, and targeting prevention prior to the onset of psychopathological conditions.[14] Application of a multidimensional, developmental framework for early childhood DB is the goal of the present paper.

An important first step towards testing DB dimensional models in young children was testing the application of existing dimensional frameworks to early childhood. This has been demonstrated in clinical and community samples of preschoolers with patterns similar to those in older youth.[15,16] There is also some evidence that callous behaviors are distinguishable at preschool age, with short-term predictive utility.[17–19] However, most dimensional approaches to-date have utilized *DSM* symptom sums; their developmental

impossibility/improbability (CD) and/or imprecision (ODD) for capturing early childhood clinical patterns has been previously noted.[3] A key limitation is that dimensions comprised of symptoms that capture only severe behaviors cannot provide an ordered metric for characterizing the dimensions along a normal-abnormal continuum.[20] Thus, an important follow-on for advancing a developmental framework is operationalizing dimensions in a manner that addresses measurement challenges inherent in assessing psychopathological processes in early childhood. These challenges include the high level of behavioral variability and the overlap between normative misbehavior and disruptive behavior. Such a developmental approach captures the developmental variation of early childhood by characterizing a continuum of normal to abnormal behaviors, with atypicality derived from deviation from expectable patterns within the age period.[3,21,22] This requires conceptualization of dimensions that are linked to normative developmental processes and operationalization of age typical behavioral expression.

We previously proposed and provided a preliminary test of a developmentally informed model of DB in early childhood with four distinct dimensions linked to core developmental processes of the preschool period:[3,15] Temper Loss and regulation of frustration, *Noncompliance* and internalization of rules, Aggression and capacity to modulate aggressive tendencies, and Low Concern for Others and the emergence of empathy and conscience (the theoretical rationale for these dimensions has been extensively discussed).[3,15] Using secondary data analysis, this prior study demonstrated the superior fit of a four-dimension model compared to a *DSM*-oriented 2-dimension ODD/CD model, an irritable/headstrong/hurtful model, and a DB/callous model.[15] However, prior testing of this multidimensional model was constrained by the use of existing measures not developed for this purpose. For example, the use of *DSM* symptoms to comprise dimensions has a constricted range (focuses on extreme behaviors), has the same symptoms for all age periods, and does not provide full coverage across dimensional spectra.

Here we use Item Response Theory (IRT)[23] to test the four-dimension developmentally informed model with a novel measure, the Multidimensional Assessment of Preschool Disruptive Behavior (MAP-DB). IRT is useful for dimensional modeling because it can map the locations of both items and respondents on an underlying latent trait continuum, scaled from mild, commonly occurring behaviors to severe, rarely occurring behaviors. Within the framework of IRT, behaviors are psychometrically defined as "abnormal" or severe when they are rarely occurring (e.g., in <5% of the population).

To operationalize behaviors along the normal-abnormal spectrum for early childhood, the MAP-DB incorporates assessment of behavioral frequency, quality, and context. These parameters may provide more nuanced distinction between normative and clinically concerning behaviors in this age period.[24–26] Milder, normative misbehaviors were theorized to occur in developmentally expectable contexts (e.g., "when frustrated"), whereas atypical behaviors were theorized to occur in developmentally unexpectable contexts (e.g., "out of the blue"). Qualitatively atypical behaviors were conceptualized in terms of intensity (e.g., "hurt someone on purpose"), dysregulation (e.g., "difficulty calming down after tantrum"), intransigence (e.g.," refuse to do as asked 'no matter what'"), and provocativeness (e.g., "persist in scaring or upsetting someone"). To test the theory that even normative

misbehaviors would be atypical if they occurred at higher than average frequencies, we utilized an objective frequency format (i.e., ratings of how often the behavior actually occurred). These are in contrast to subjective ratings (e.g., "never," "sometimes," "often"), which may give the same rating (e.g., "often") to varying frequencies depending on factors influencing the judgment of the rater. Given the rapidity of change in developmental capacity across the preschool period and the centrality of sex differences in psychopathology expression, testing for age and sex differences is also an important aspect of this developmental framework.[27,28]

The MAP-DB is employed to test the theorized dimensional spectrum of preschool disruptive behavior in a large, sociodemographically diverse sample of preschool children.

Aims of the present study are:

1.   Test whether the data support the hypothesized 4-dimension, developmentally informed model and its robustness to sociodemographic variation;

2.   Test the hypothesis that location along the dimensional severity continuum will vary based on context, quality, and frequency.

METHOD

Sample and Procedures

The Multidimensional Assessment of Preschoolers (MAPS) Study is comprised of a large, sociodemographically stratified sample of preschoolers recruited from 5 pediatric clinics in the Chicagoland area.[29] All study procedures were approved by Institutional Review Boards and parental informed consent was obtained. Parents were eligible for MAPS if they were the legal guardian of a 3- to 5-year-old child present at the clinic. These children received well-child care at the pediatric clinic. All children were eligible provided their parent had not already participated in the study for a sibling. Psychiatric referral history was not collected at screening; however, any psychiatrically referred children who met other eligibility criteria were eligible to participate. Of 1,814 eligible for the survey, 1,606 (88.5%) consented and 1,516 completed surveys (94.4% completion rate of all consented; 83.6% of all eligible). Compared with the 298 eligible parents who did not participate, the 1,516 participants were significantly more likely to be female (94.3% vs. 88.6%), of minority ethnicity (68.6% vs. 59.9%), or from one of the Chicago-based clinics (85.6% vs. 76.8%), all ps<.01. However, results of a logistic regression model revealed that only recruitment source and female caregiver uniquely predicted participation. Sample size was designed to provide power of at least .8 to detect small (.10) differences in loadings across the four key sociodemographic strata: child age, sex, poverty status, and ethnicity.

Twenty-six children in the autism spectrum were excluded (due to insufficient n to provide informative data on this important sub-group), and 2 participants with more than half the MAP-DB data missing were also excluded. Thus, the analytic sample was 1,488. Ninety-seven percent were biological parents and 91% were mothers. Consistent with the MAPS stratification goals, the sample was fairly evenly distributed by child sex (49% girls, 51% boys), age (35% 3 year olds, 36% 4 year olds, 29% 5 year olds), race/ethnicity (36% African

American, 36% Hispanic, 27% Non-Hispanic White, 1% Other), and poverty status (42% below federal poverty level based on annual household income and household size).[30] A demographically comparable test-retest sample was also assessed (n=76). This sample size is sufficient for power of .8 to detect a test-retest correlation of .28, a magnitude that indicates adequate test-retest reliability of construct stability. The survey, which consisted of the MAP-DB, demographic information, and brief information on disruptive behavior correlates, was administered in English or Spanish. A $20 incentive was provided for survey completion, with a $10 in-clinic completion bonus.

MAP-DB Measure

The MAP-DB was developed by a team of experts in early childhood, clinical assessment and treatment, and developmental epidemiology. The iterative measure development process included theoretical delineation of the core dimensions,[3] preliminary validation utilizing secondary data,[15] pilot testing, focus groups, and review of extant measures. Measurement development was led by the first and second authors including: (1) construct delineation; (2) item generation; (3) review of item pool to assess how well the item captured the intended dimension, ease of comprehension, and wording; and (4) deciding which items to retain in the final pool. Items for which there was disagreement were retained for empirical determination of fit. A range of 4 interactional contexts (with parents, with other adults, with siblings, with other children) and 5 contextual antecedents ("when tired, hungry or sick," "when frustrated, angry or upset," "during daily routines," "to get something s/he wanted," "for no reason or out of the blue") were included, with variation by dimension depending on contextual salience. The final item pool for the present analyses was comprised of 111 MAP-DB items. A Spanish version was generated via certified translation and back-translation. Ratings were done on a 6-point scale (0=Never; 1=Rarely [less than once per week]; 2=Some [1–3] days of the week; 3=Most [4–6] days of the week; 4=Every day of the week; 5=Many times each day), within the range of optimal number of response options for the use of IRT in health assessment.[31]

Temper Loss items ranged from normative expressions, such as tantrums in the face of frustration, to intense, dysregulated tantrums (22 items).[29] Noncompliance items ranged from normative refusal to follow directions to provocative and recalcitrant disobedience (30 items). Aggression items included normal reactive aggression and abnormal manifestations of intentional, hostile aggression (44 items). Low Concern for Others items included disregard for others' feelings and pleasure in others' distress (15 items).

Data Reduction and Modeling

Modeling of the MAP-DB was guided by both theory and data to extract factors as parsimonious as possible, minimize interfactor correlations, best represent the dimensional spectrum, and discriminate well between children with high vs. low levels of each dimension. Seventy-eight of the original 111 items were retained via this process. Confirmatory factor analysis (CFA), a method for testing theories about the structure of items, conducted in Mplus 6.11,[32] tested the fit of the 4-dimension theoretical model. We examined the comparative fit index (CFI) and the root mean square error of approximation (RMSEA) statistics to assess the extent to which the model was a good fit to the data.

RMSEA values <.08 and CFI values >.90 are considered indicators of acceptable fit.[33] We also calculated Cronbach's alpha (α) coefficient as an indicator of the internal reliability of the dimensions. The Bayesian information criterion (BIC) statistic was used to compare whether the same factor structure fit best across demographic subgroups (differences of 10 on the BIC are considered strong evidence of improved fit with lower BIC scores reflecting improvement[34,35]). Finally, the IRTPRO[36] program was used to employ IRT methods to estimate the severity of items along each dimension. Higher severity scores indicate "more difficult" items that are less commonly endorsed.

RESULTS

Aim 1: Examine the Fit of the 4-Dimension, Developmentally Informed Model

Statistical "fit" refers to the extent to which a model is an adequate representation of the actual data. When fit is not adequate, alternate models should be explored. Fit is measured in this study using the root mean squared error of approximation (RMSEA) on which values less than .05 are considered to indicate excellent fit, and the comparative fit index (CFI) on which values above .90 suggest good fit and values above .95 excellent fit. The 4-dimension model fit the data well according to both indicators (RMSEA=.048; CFI=.936). All 4 dimensions demonstrated excellent internal consistency (Temper Loss α=.97, Noncompliance α=.96, Aggression α=.96, Low Concern α=.92) (item-level loadings are provided in Table S1, available online). Multigroup comparisons across child age, sex, race/ethnicity, and poverty status demonstrated equivalent fit across subgroups(indicated by the lower adjusted BIC shown in Table S2, available online), suggesting the generalizability of this 4-dimension model across variable sociodemographic groups. There were, however, mean differences in scores across the subgroups by child sex, age, and ethnicity (Table S3, available online). In particular, there were consistent differences by sex (boys' scores higher on all dimensions), age (3- to 4-year-olds had higher scores on Temper Loss, Noncompliance, and Aggression than 5-year-olds; 3-year-olds were higher than 5-year-olds on Low Concern), and ethnicity (Caucasian children had higher means on all dimensions, relative to Hispanic or African American children [who did not differ from each other]). There were no differences by poverty status. Mothers and other informants did not differ in response patterns (data available from first author).

The 4-dimension model also demonstrated superior fit relative to established alternative models constructed by regrouping MAP-DB dimensions (e.g., a 3-dimension irritable [Temper Loss]/headstrong [Noncompliance]/hurtful [Aggression + Low Concern] model) (Table S4, available online), consistent with prior work in several independent samples.[15,37]

Test-retest reliability of the dimensions was based on completion of the MAP-DB twice over an average period of six months (*M*=3.44 months, SD=1.27 months). Intraclass correlations revealed good to excellent stability for the dimensions (Temper Loss=.80, Noncompliance=.81, Aggression=.85, Low Concern=.83).

Aim 2: Model the Dimensional Severity Spectrum

The severity continuum is like a ruler measuring the overall severity of a child's symptoms on each dimension. The dimensions themselves are latent variables that cannot be directly measured. The severity of a child's behaviors can only be indicated by the reports collected on each item. Just as each child can be given a score on the severity continuum, each item can be assigned a place on it. The category thresholds of each item represent the severity of the behavior measured at different frequencies of occurrence. Table 1 provides the category thresholds and item severity ("location") parameter for items along each dimension. The category thresholds (b1–b5) represent the estimated level of the latent trait at which the probability exceeds 50% that a respondent will choose the next higher category (e.g., choosing "every day" rather than "on most days"). Together, they provide a measure of the item location (indicating its severity) on the latent trait scale. The item location (first column in Table 1) represents the average of these category thresholds for the item. When this item location value exceeds the threshold marking the 95th percentile of the sample of children for the dimension, this is an indicator of "item severity" in psychometric terms (i.e., it is atypical because it occurs in <5% of the population). Category thresholds and item locations that exceed the 95th percentile are bolded in Table 1. The bolded category thresholds indicate at what specific frequency that particular item becomes atypical. For example, item 1, "Lose temper or have a tantrum when frustrated, angry or upset," tends to be severe only when it occurs "every day" or more; item 22, "Stay angry for a long time," tends to be severe when it occurs "some days of the week" or higher.

Across the dimensions, items ranged in severity from just above the mean (0.52) to over 3 standard deviations above the mean (3.66). Across all dimensions, items provided good coverage of the severe end (e.g., Aggression location values ranged from 1.5–3.66). However, the dimensions varied in the extent to which they encompassed the "normal" end of the dimensional continuum. Temper Loss and Noncompliance provided coverage across this full spectrum, including 45–64% of items <95th percentile severity threshold. In contrast, behaviors on the Aggression and Low Concern dimensions were more severe and not commonly occurring, with all items having item severity >95th percentile.

The ordering of item severities in the left-hand column of Table 1 demonstrates the presence of the theorized severity continuum within each dimension. For example, developmentally expectable behaviors (e.g., behaviors that occurred when frustrated or during daily routines, including tantrums, disobedience, aggression, or not caring about others' feelings) tended to have lower item severity (indicated by their severity threshold in the right-hand column, many <95th percentile). In contrast, behaviors that occurred in developmentally unexpected contexts (e.g., hitting a nonparental adult, having a tantrum out of the blue) tended to have higher item severity (>95th percentile). Similarly, across all dimensions, behaviors that were qualitatively atypical, i.e., intense (e.g., hurt someone on purpose), dysregulated (e.g., have a tantrum lasting >5 min.), intransigent (e.g., automatically resist whatever you ask), or provocative (e.g., enjoy making others mad), were more likely to have item severity scores >95th percentile.

Examination of the category thresholds for Temper Loss, Noncompliance, and Aggression revealed within-dimension variation in terms of the response category at which items

crossed into atypicality, which became lower as items progressed across the severity continuum for the dimension. Taking Noncompliance as an example, "Say "no" when told to do something" has low item severity (1.14) and is severe when exhibited "every day" or higher. In contrast, "Misbehaves in ways that are dangerous" has high item severity (2.52) and is severe when exhibited "some days a week" or higher. In contrast, Low Concern did not demonstrate this same variation in frequency at different levels of severity, perhaps due to its more restricted range.

Review of category thresholds also revealed variation across dimensions in the frequency at which severity at the 95th percentile tended to be reached. For Noncompliance and Temper Loss, most items were severe when exhibited "most" or "every day" of the week. In contrast, Aggression and Low Concern items tended to be severe when exhibited "some" or "most" days of the week. Two Aggression items had a severity threshold at a frequency of "rarely."

Figures 1a–b through 4a–b highlight how item category thresholds vary on the dimensions according to the quality and frequency of a behavior (normative misbehavior vs. qualitatively atypical behaviors). The "rug" along the X-axis of each plot is the severity distribution of the dimension being measured. The dots represent the position of each item category threshold along the dimension, with the threshold of atypicality (95th percentile) marked with the dashed vertical (red) line. In the Noncompliance dimension (Figures 1a–b) for the normative misbehavior ("Say "no" when told to do something"), the 95th percentile (1.61) falls at daily frequency. In contrast, the 95th percentile occurs at weekly frequency for a qualitatively atypical Noncompliance behavior ("Misbehaves in ways that are dangerous"). This means that engaging in dangerous misbehavior on a weekly basis is as severe as saying no to requests on a daily basis. Figures 2a–b through 4a–b illustrate differences in severity thresholds for items on the Aggression, Temper Loss, and Low Concern dimensions. In the Aggression (Figures 2a–b) dimension, the 95th percentile for the normative misbehavior ("Act aggressively when frustrated, angry or upset") falls at the weekly frequency, while the 95th percentile for the atypical misbehavior ("Act aggressively to get something s/he wanted") occurs rarely. In the Temper Loss dimension (Figures 3a–b), the 95th percentile for the normative item ("Lose temper or have a tantrum when tired, hungry or sick") falls at the almost daily frequency, while 95th percentile for the atypical item ("Have a tantrum that lasted more than 5 minutes") occurs only weekly. In the Low Concern dimension (Figures 4a–b), the 95th percentile of the comparatively normative misbehavior ("Not seem to care about your or other parents' feelings") falls at weekly, while the 95th percentile for the more atypical misbehavior ("Act like s/he didn't care when someone else felt bad or sad") only needs to occur rarely. Thus, these figures illustrate how the frequency differs for the threshold of typicality for normative vs. atypical misbehaviors.

To test the extent to which there was distinctness in severity patterns across dimensions, we examined cross-dimensional severity overlap. One-hundred eighteen preschoolers (7.8%) had scores in the atypical range on at least one dimension. Within this subgroup, 31% had atypical scores on one dimension, 32% on 2–3 dimensions, and 37% on all 4 dimensions.

DISCUSSION

Findings provide psychometric support for a developmentally informed, dimensional model of preschool disruptive behavior. IRT methods were used to demonstrate that theoretically derived developmental indicators of atypicality differentiate behaviors that are commonly occurring misbehaviors at preschool age from those that are abnormal. In particular, we have shown that normative misbehaviors can be distinguished from abnormal behaviors in early childhood in terms of the developmental expectability of the context in which they occur and their qualitative features. Normative misbehaviors, i.e., those that occur in expectable contexts and/or are qualitatively better modulated and flexible, fell at the milder end of the dimensional spectrum. In contrast, behaviors that were qualitatively intense, dysregulated, intransigent, provocative, and/or occurred in developmentally unexpectable contexts demonstrated heightened severity along the dimensional spectrum. Further, there were distinct dimensional patterns in terms of severity thresholds and in the moderate degree of cross-dimensional overlap. Although the majority of preschoolers exhibited some of the behaviors tapped into by the MAP-DB dimension, <10% fell in the atypical range on any dimension. This suggests that the MAP-DB is effectively distinguishing the normative: atypical distinction for this age group.

The generalizability of our findings for broad-based clinical application is highlighted by the fit of the 4-dimension model across variations in sex, ethnicity, and child age. Mean differences for some subgroups (e.g., uniformly higher scores for boys on all dimensions), however, raise the possibility that clinical thresholds may need to be determined by subgroup. This can only be determined based on norming in large representative samples with sufficient numbers in each subgroup and across strata to ensure that meaningful differences have been accurately characterized and validated in relation to clinical endpoints. This will be an important step for developing the clinical potential of the MAP-DB in diverse populations.

Demarcation of the developmentally informed, behavioral distinctions was further evident in the different frequency thresholds at which they became psychometrically atypical for Temper Loss, Noncompliance, and Aggression. In general, milder, normative misbehaviors had to occur very regularly to be severe (i.e., daily or more for Temper Loss and Noncompliance, weekly for Aggression). This emphasizes that even the common misbehaviors of early childhood are atypical if they occur at high frequencies. For example, in previous findings from this sample, we have shown that while tantrums are a normative misbehavior, <10% of preschoolers tantrum daily and those who tantrum daily are more likely to have problems in other areas.[29,38] In contrast, qualitatively atypical behaviors in each dimension exceeded the threshold with relatively low occurrence.

Low Concern did not demonstrate this same variation in frequency at different levels of severity and had a more restricted range, with average severity values of all items >95th percentile threshold (connoting that they are rare and serious behaviors). Further, virtually all Low Concern items had the same relative frequency at which the 95th percentile threshold was reached. This is not surprising given that this dimension was conceptualized as a developmental substrate of later callous/unemotional (C/U) patterns.[39] C/U is an

atypical variant of antisocial behavior[6] and may not have normative variations. The MAP-DB Low Concern dimension represents one of the first attempts to operationalize callous traits in a clinically and developmentally salient manner specific to early childhood.[19]

A primary limitation of the present findings is that only psychometric (internal) validity was established. Psychometrically, atypicality is delineated in terms of rare occurrence as manifested by scores that occur in a low percentage of the population. In contrast, clinical (external) validity is defined by association with impairment, family history, caseness, differentiated etiologic markers, response to treatment, and prognostic utility. Further research is needed to establish the relationship between MAP-DB psychometric severity and clinical severity. Ultimately, it is this joint consideration which determines clinical utility. In particular, the added value of this dimensional, developmentally informed approach will rest on: (a) establishing whether psychometrically demarcated atypicality, as differentiated by the distinct dimensions, meaningfully predicts varied clinical symptoms and impairment; (b) increasing the accuracy of differentiation of normative versus clinically significant behaviors in early childhood relative to traditional diagnostic categories; and (c) linking established risk processes to distinct dimensions. Further, future research which includes sufficient numbers of children with developmental disabilities for modeling individual differences in patterns, will importantly inform understanding of atypical manifestations of behavior (e.g., Low Concern) as they manifest in children with developmental disabilities relative to children with emotional and behavioral syndromes.

Another limitation is the cross-sectional nature of the data. Recent findings from large representative samples are promising in demonstrating the developmental continuity of ODD dimensions from school age-adolescence as well as their predictive utility.[40] Establishing continuity in MAP-DB dimensions over time will be important. Finally, as the majority of the informants were mothers, calibration of the MAP-DB with fathers and teachers is needed.

Visions for DSM-5,[41] as well as neuroscience-based classifications of psychopathology, such as the Research Domain Criteria (RDoC),[42] converge on the importance of developmentally informed, dimensional phenotypes for pinpointing unfolding clinical sequences as they progress from early prodromal manifestations to frank disorder.[14,43] Our 4-dimension framework integrates key elements of leading dimensional models that have been validated for ODD and CD.[6,8,12] Findings suggest that their varying components may best be considered distinctly, consistent with work by others.[9,16,40,44]

There are disadvantages and benefits to both categorical and dimensional models of childhood psychopathology, and one is not intrinsically preferable to the other.[45] They may best be understood as complementary. Evidence is emerging that some types of child psychopathology (such as melancholic depression) may belong to discrete taxa while others (such as attention-deficit/hyperactivity [ADHD]) may best be understood as dimensions. [45] The present findings highlight the potential utility of psychometric methods for operationalizing the vision of a dimensional classification approach. In particular, IRT methods revealed a developmental spectrum of behavior and identified varying thresholds for individual behaviors along a dimensional severity continuum. We theorize that this

combination of developmental differentiation and articulation of the normal–abnormal continuum will enhance specificity of prediction and linkage to mechanisms because it removes much of the "noise" of developmental variability in early childhood that has impeded accurate clinical identification when traditional a-developmental categorical nosologies have been applied. Variation in dimensional patterns found here were also consistent with revisions of DB symptom criteria in *DSM-5*, i.e., differentiating symptom dimensions within ODD and recognizing the importance of a callous trait specifier for CD.[43] If clinical and incremental utility of this approach is established relative to traditional psychiatric classification systems, the developmental severity continuum delineated here suggests that early childhood classification might benefit from greater specification. Such specification may include incorporation of symptoms that differentiate normative misbehaviors from qualitatively atypical behaviors, and incorporation of specific frequency thresholds based on deviation from expectable developmental patterns of occurrence.

Multidimensional measures, such as the MAP-DB, ultimately have the potential to provide a psychometrically robust, developmentally-informed metric for characterizing behavior as normative, atypical, or "at the boundaries" for an age period. This has the potential to provide an empirical basis for determining whether "watchful waiting," psychoeducational guidance, pharmacologic, or more intensive treatments are warranted. Dimensional measurement tools that enable precise developmental characterization of latent dimensions can advance understanding of the unfolding of psychopathology and etiologic distinctions. Several recent studies provide an intriguing glimpse of the potential clinical utility of this dimensional approach for prevention. Low-intensity interventions that target children based on temperamental or personality risk have shown promise for reducing the likelihood of developing clinical disorder, including one targeting preschool children.[46,47] These studies suggest that targeted prevention along dimensional continua may be fruitful for altering risk of developmental psychopathology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.
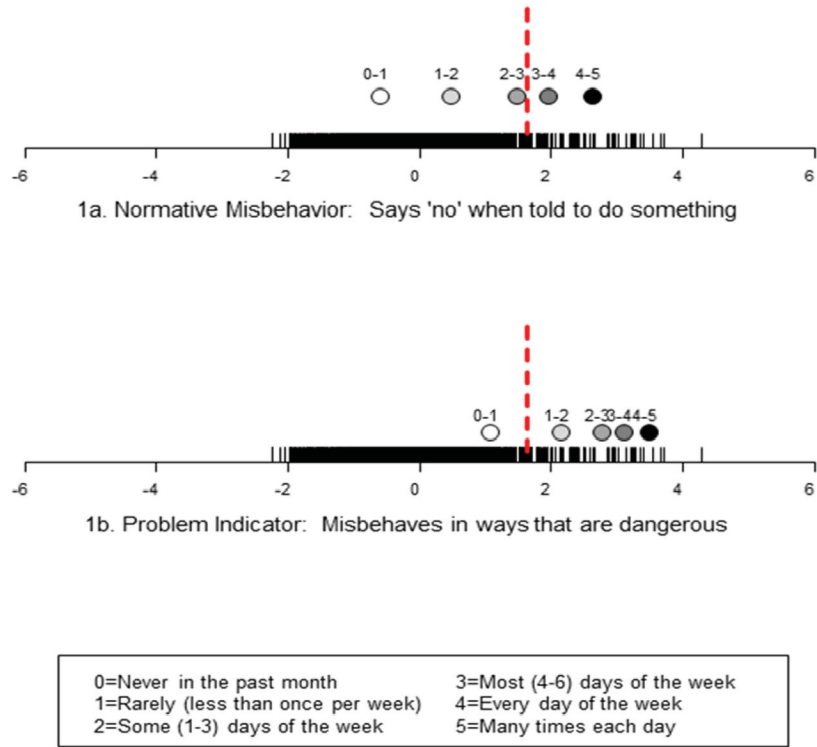
Acknowledgments

References

1. Kim-Cohen J, Caspi A, Moffit T, Harrington H, Milne BJ, Poulton R. Prior juvenile diagnoses in adults with mental disorder: Developmental follow-back of a prospective longitudinal cohort. Arch Gen Psychiat. 2003; 60:709–717. [PubMed: 12860775]
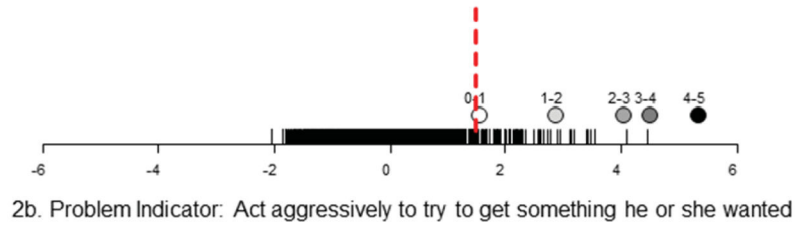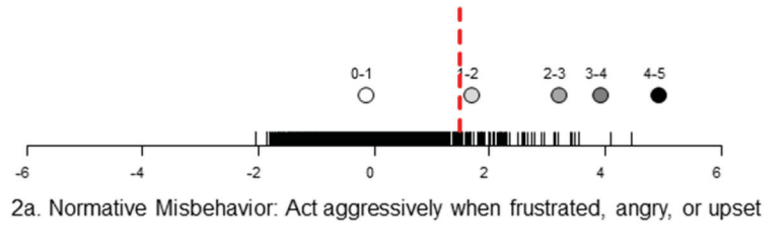
2. Nock M, Kazdin A, Hiripi E, Kessler R. Lifetime prevalence, correlates, and persistence of oppositional defiant disorder: Results from the National Comorbidity Survey replication. Journal of Child Psychology and Psychiatry. 2007:703–713. [PubMed: 17593151]

3. Wakschlag L, Tolan P, Leventhal B. "Ain't misbehavin": Towards a developmentally-specified nosology for preschool disruptive behavior. Journal of Child Psychology and Psychiatry. 2010; 51:3–22. [PubMed: 19874427]

4. Bufferd SJ, Dougherty L, Carlson G, Rose S, Klein D. Psychiatric disorders in preschoolers. Am J Psychiat. 2012; 169:1157–1164. [PubMed: 23128922]

5. Leibenluft E. Severe mood dysregulation, irritability, and the diagnostic boundaries of bipolar disorder in youths. Am J Psychiatry. Feb; 2011 168(2):129–142. [PubMed: 21123313]

6. Frick P. Developmental pathways to conduct disorder: Implications for future directions in research, assessment and treatment. J Clin Child Adolesc. 2012; 41:378–389.

7. Stringaris A, Zavoa H, Leibenluft E, Maughan B, Eley T. Adolescent irritability: Phenotypic associations and genetic links with depressed mood. Am J Psychiat. 2012; 169:47–54. [PubMed: 22193524]

8. Stringaris A, Goodman R. Three dimensions of oppositionality in youth. J Child Psychol Psychiatry. 2009; 50(3):216–223. [PubMed: 19166573]

9. Burke JD, Hipwell AE, Loeber R. Dimensions of oppositional defiant disorder as predictors of depression and conduct disorder in preadolescent girls. J Am Acad Child Adolesc Psychiatry. May; 2010 49(5):484–492. [PubMed: 20431468]

10. Rowe R, Maughan B, Costello J, Angold A. Defining oppositional defiant disorder. J Child Psychol Psyc. 2005; 46:1309–1316.

11. Drabick D, Gadow K. Deconstructing oppositional defiant disorder: Clinic-based evidence for an anger/irritability phenotype. J Am Acad Child Psy. 2012; 51:384–393.

12. Burt S. How do we optimally conceptualize the heterogenety within antisocial behavior? An argument for aggressive versus non-aggressive behavioral dimensions. Clin Psychol Rev. 2012; 32:263–279. [PubMed: 22459789]

13. Frick PJ, White S. Research Review: The importance of callous-unemotional traits for developmental models of aggressive and antisocial behavior. J Child Psychol Psyc. 2008; 49:359–375.

14. Insel T. Disruptive insights in Psychiatry: Transforming a clinical discipline. J Clin Invest. 2009; 119:700–705. [PubMed: 19339761]

15. Wakschlag L, Henry DB, Tolan PH, Carter AS, Burns JL, Briggs-Gowan MJ. Putting theory to the test: modeling a multidimensional, developmentally-based approach to preschool disruptive behavior. J Am Acad Child Psy. Jun; 2012 51(6):593–604.

16. Ezpeleta L, Granero R, de la Osa N, Penelo E, Domenech JM. Dimensions of oppositional defiant disorder in 3-year-old preschoolers. J Child Psychol Psychiatry. Nov; 2012 53(11):1128–1138. [PubMed: 22409287]

17. Ezpeleta L, de la Osa N, Granero R, Penelo E, Domenech JM. Inventory of callous-unemotional traits in a community sample of preschoolers. J Clin Child Adolesc. Oct 24; 2013 42(1):91–105.

18. Kimonis E, Frick P, Boris N, Smyke A, Cornell A, Farrell J, Zeanah C. Callous-unemotional features, behavioral inhibition, and parenting: Independent predictors of aggression in a high-risk preschool sample. J Child Fam Stud. 2006; 15:745–746.

19. Hyde L, Shaw D, Gardner F, Cheong J, Dishion T, Wilson M. Dimensions of callousness in early childhood: Links to problem behavior and family intervention effectiveness. Dev Psychopathol. 2013; 25(2):347–363. [PubMed: 23627949]

20. Achenbach, T. Assessment of psychopathology. In: Sameroff, A.; Lewis, M.; Miller, S., editors. Handbook of Developmental Psychopathology. 2. New York: Kluwer Academic; 2000. p. 41-56.

21. Wakschlag L, Briggs-Gowan M, Carter A, Hill C, Danis B, Keenan K, McCarthy K, Leventhal B. A developmental framework for distinguishing disruptive behavior from normative misbehavior in preschool children. J Child Psychol Psyc. 2007; 48(Special Issue on Preschool Psychopathology): 976–987.

22. Markon K, Krueger R, Watson D. Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. J Pers Soc Psychol. 2005; 88:137–157.

23. Reise S, Waller N. Item response theory and clinical measurement. Annu Rev Clin Psychol. 2009; 5:25–46.

24. Tremblay R, Nagin D, Seguin J, Zoccolillo M, Zelazo P, Boivin M, Perusse D, Japel C. Physical aggression during early childhood: Trajectories and predictors. Pediatrics. 2004; 114:43–50.

25. Hay DF, Castle J, Davies L. Toddlers' use of force against familiar peers: a precursor of serious aggression? Child Dev. 2000; 71(2):457–467. [PubMed: 10834477]

26. Kochanska G, Aksan N. Children's conscience and self-regulation. J Pers. 2006; 74:1587–1617. [PubMed: 17083659]

27. Rutter M, Caspi A, Moffitt T. Using sex differences in psychopathology to study causal mechanisms: Unifying issues and research strategies. J Child Psychol Psychiatry. 2003; 44(8): 1092–1115. [PubMed: 14626453]

28. Baillargeon HR, Morrisset A, Keenan K, Normand LC, Seguin JR, Japel C, Cao G. Development of disruptive behaviors in young children: A prospective population-based cohort study. Inf Mental Hlth J. 2012; 33(6):633–650.

29. Wakschlag L, Choi S, Carter A, Hullsiek H, Burns J, McCarthy K, Leibenluft E, Briggs-Gowan M. Defining the developmental parameters of temper loss in young children: Implications for developmental psychopathology. Journal of Child Psychiatry and Psychology. 2012; 53:1099–1108.

30. Barajas, RG.; Philipsen, N.; Brooks-Gunn, J. Cognitive and emotional outcomes for children in poverty. In: Crane, D.; Heaton, T., editors. Handbook of Families and Poverty. Newbury Park, CA: Sage; 2008. p. 311-333.

31. Bode R, Lai J, Cella D, Heinemann A. Issues in the development of an item bank. Arch Phys Med Rehabil. 2003; 84:S52–S60. [PubMed: 12692772]

32. Muthen, LK.; Muthen, BO. Mplus User's Guide. Vol. 6. Los Angeles, CA: Muthen and Muthen; 2010.

33. Hu L, Bentler P. Cutoff criteria for fit indexes in covariance structure analyses: Conventional criteria versus new alternatives. Structural Equation Modeling. 1999; 6:1–55.

34. Eaton N, Krueger R, Markon K, Keyes K, Skodol A, Wall M, Hasin DS, Grant B. The structure and predictive validity of the internalizing disorders. J Abnorm Psychol. 2013; 122(1):86–92. [PubMed: 22905862]

35. Raftery A. Bayesian model selection in social research. Soc Method. 1995; 25:111–163.

36. IRTPRO. Flexible, multidimensional, multiple categorical IRT modeling [computer program]. Lincolnwood, IL: Scientific Software International; 2011.

37. Wakschlag L, Henry D, Blair R, Dukic V, Burns J, Pickett K. Unpacking the association: Individual differences in the relation of prenatal exposure to cigarettes and disruptive behavior phenotypes. Neurotoxicol Teratol. Jan-Feb;2011 33(1):145–154. [PubMed: 21256429]

38. Hullsiek, H.; Briggs-Gowan, M.; Wakschlag, L. Temper tantrum frequency and associated problems in a diverse sample of preschool children. American Academy of Child and Adolescent Psychiatry and Canadian Academy of Child and Adolescent Psychiatry Joint Annual Meeting; 2011; Toronto, ON.

39. Briggs-Gowan MJ, Nichols SR, Voss J, Zobel E, Carter AS, McCarthy KJ, Pine DS, Blair J, Wakschlag LS. Punishment insensitivity and impaired reinforcement learning in preschoolers. Journal of Child Psychology and Psychiatry. 2013 Epub ahead of print.

40. Whelan Y, Stringaris A, Maughan B, Barker ED. Developmental continuity of oppositional defiant disorder subdimensions at ages 8, 10, and 13 years and their distinct psychiatric outcomes at age 16 years. J Am Acad Child Adolesc Psychiatry. 2013; 52(0):961–9. [PubMed: 23972698]

41. Helzer, J.; Kraemer, HC.; Krueger, RF.; Wittchen, HU.; Sirovatka, PJ.; Regier, DA. Dimensional approaches in diagnostic classification: refining the research agenda for DSM-V. American Psychiatric Publishing, Inc; 2008.

42. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, Sanislow C, Wang P. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiatry. Jul; 2010 167(7):748–751. [PubMed: 20595427]

43. Frick P, Nigg J. Current issues in the diagnosis of Attention Deficit Hyperactivity Disorder, Oppositional Defiant Disorder, and Conduct Disorder. Annual Review of Clincal Psychology. 2012; 8:77–107.

44. Stringaris A, Goodman R. Longitudinal outcome of youth oppositionality: Irritable, headstrong, and hurtful behaviors have distinctive predictions. J Am Acad Child Psy. 2009; 48(4):404–412.

45. Coghill D, Sonuga-Barke EJ. Annual research review: Categories versus dimensions in the classification and conceptualisation of child adolescent mental disorders--implications of recent empirical study. Journal of Child Psychology and Psychiatry. 2012; 53(5):469–489. [PubMed: 22288576]

46. Yeager DS, Miu AS, Powers J, Dweck CS. Implicit theories of personality and attributions of hostile intent: a meta-analysis, an experiment, and a longitudinal intervention. Child Dev. Sep; 2013 84(5):1651–1667. [PubMed: 23402434]

47. Rapee RM. The preventative effects of a brief, early intervention for preschool-aged children at risk for internalising: follow-up into middle adolescence. J Child Psychol Psychiatry. Jul; 2013 54(7):780–788. [PubMed: 23397984]

1a. Normative Misbehavior: Says 'no' when told to do something

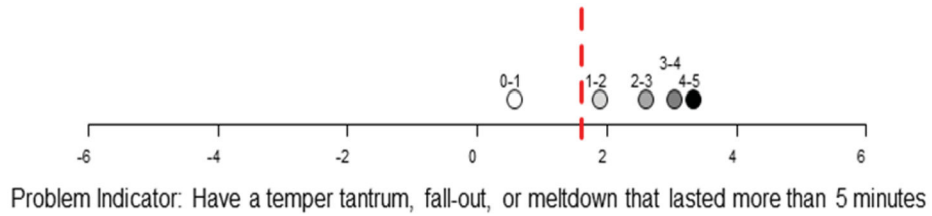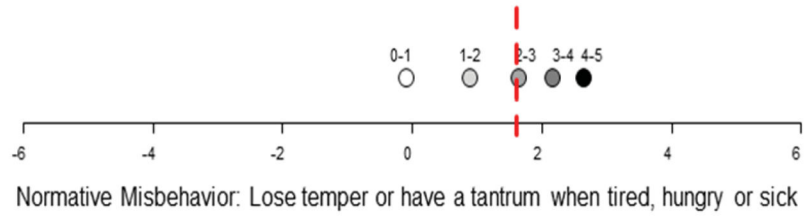1b. Problem Indicator: Misbehaves in ways that are dangerous

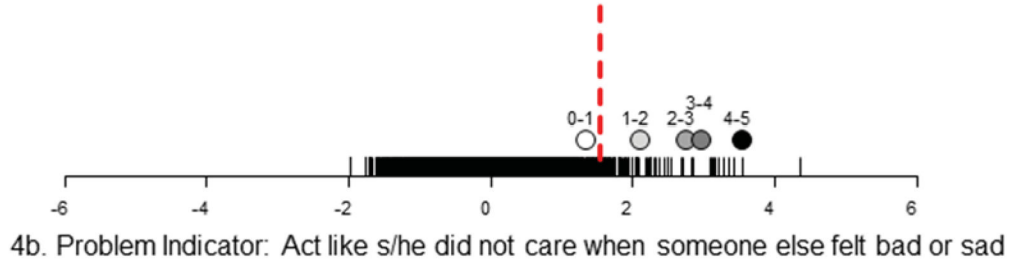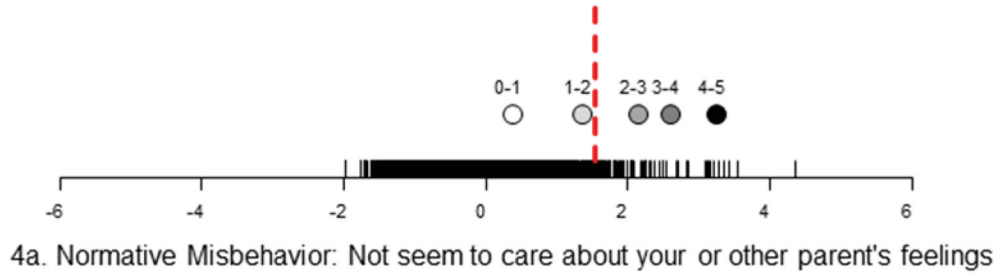|	
0=Never in the past month	3=Most (4-6) days of the week
1=Rarely (less than once per week)	4=Every day of the week
2=Some (1-3) days of the week	5=Many times each day

Figure 1.
Rug[a] and Dot Plot Illustrating Higher Category Thresholds for Normative Misbehavior vs. Problem Indicator from the Noncompliance Dimension Note: 1a. Normative Misbehavior. 1b. Problem Indicator. [a]The "rug" along the X-axis of each plot is the severity distribution of the dimension being measured. The dots represent the position of each item category threshold along the dimension, with the threshold of atypicality (95th percentile) marked with the dashed vertical line.

2a. Normative Misbehavior: Act aggressively when frustrated, angry, or upset

2b. Problem Indicator: Act aggressively to try to get something he or she wanted

0=Never in the past month	3=Most (4-6) days of the week
1=Rarely (less than once per week)	4=Every day of the week
2=Some (1-3) days of the week	5=Many times each day

Figure 2.

Rug[a] and Dot Plot Illustrating Higher Category Thresholds for Normative Misbehavior vs. Problem Indicator from the Aggression Dimension. Note: 2a. Normative Misbehavior. 2b. Problem Indicator. [a]The "rug" along the X-axis of each plot is the severity distribution of the dimension being measured. The dots represent the position of each item category threshold along the dimension, with the threshold of atypicality (95th percentile) marked with the dashed vertical line.

Figure 3.

Figure 3.
Rug[a] and Dot Plot Illustrating Higher Category Thresholds for Normative Misbehavior vs. Problem Indicator from the Temper Loss Dimension. Note: 3a. Normative Misbehavior. 3b. Problem Indicator. [a]The "rug" along the X-axis of each plot is the severity distribution of the dimension being measured. The dots represent the position of each item category threshold along the dimension, with the threshold of atypicality (95th percentile) marked with the dashed vertical line.

4a. Normative Misbehavior: Not seem to care about your or other parent's feelings

4b. Problem Indicator: Act like s/he did not care when someone else felt bad or sad

0=Never in the past month   3=Most (4-6) days of the week
1=Rarely (less than once per week) 4=Every day of the week
2=Some (1-3) days of the week   5=Many times each day

Figure 4.
Rug[a] and Dot Plot Illustrating Higher Category Thresholds for Normative Misbehavior vs. Problem Indicator from the Low Concern Dimension. Note: 4a. Normative Misbehavior. 4b. Problem Indicator. [a]The "rug" along the X-axis of each plot is the severity distribution of the dimension being measured. The dots represent the position of each item category threshold along the dimension, with the threshold of atypicality (95th percentile) marked with the dashed vertical line.

Table 1

Multidimensional Assessment of Preschool Disruptive Behavior Dimensional Severity Thresholds

	Item Location *mean(b)*	Category Thresholds				
		Rarely or higher *b1*	Some days of week or higher *b2*	Most days or higher *b3*	Every day or higher *b4*	Many times a day *b5*
Temper Loss[a] (95th% threshold =1.60)						
1. Lose temper or have a tantrum when frustrated, angry or upset	1.24	−0.30	0.72	1.43	1.97[c]	2.38
2. Lose temper or have a tantrum to get something s/he wanted	1.29	−0.26	0.74	1.49	2.05	2.45
3. Have a temper tantrum, fall-out, or melt-down	1.29	−0.17	0.78	1.50	1.96	2.41
4. Lose temper or have a tantrum when tired, hungry or sick	1.31	−0.31	0.68	1.53	2.13	2.53
5. Become frustrated easily	1.33	−0.52	0.65	1.61	2.21	2.69
6. Lose temper or have a tantrum during daily routines	1.34	−0.26	0.71	1.62	2.05	2.56
7. Lose temper or have a tantrum with parents	1.36	−0.17	0.81	1.56	2.06	2.54
8. Yell angrily at someone	1.43	−0.42	0.80	1.71	2.29	2.78
9. Have a short fuse	1.45	−0.09	0.89	1.64	2.17	2.65
10. Have difficulty calming down when angry	1.58	−0.11	0.97	1.81	2.40	2.82
11. Have a hot or explosive temper	1.66	0.42	1.23	1.80	2.21	2.64
12. Get extremely angry	1.66	0.30	1.18	1.89	2.27	2.66
13. Act irritable	1.66	−0.27	0.93	1.94	2.60	3.11
14. Stamp feet or hold breath during a temper tantrum	1.71	0.09	1.04	1.92	2.55	2.96
15. Keep on having a temper tantrum even when you tried to help calm down	1.76	0.40	1.32	2.00	2.39	2.71
16. Have a temper tantrum lasting >5 min.	1.79	0.20	1.17	2.12	2.57	2.90
17. Lose temper or have a tantrum with other adults	1.91	0.36	1.43	2.23	2.62	2.93
18. Break or destroy things during a temper tantrum	1.92	0.60	1.49	2.13	2.50	2.89
19. Have a temper tantrum until exhausted	1.95	0.67	1.49	2.14	2.54	2.90
20. Lose temper or have a tantrum "out of the blue"	1.96	0.64	1.49	2.17	2.56	2.96
21. Hit, bite, or kick during a temper tantrum	2.04	0.78	1.58	2.12	2.69	3.05
22. Stay angry for a long time	2.30	0.58	1.91	2.62	3.06	3.34
Noncompliance (95th% threshold=1.61)						
23. Break rules even when s/he knew you were watching	0.52	1.93	−0.38	1.08	1.67	2.18

		Category Thresholds				
	Item Location *mean(b)*	Rarely or higher *b1*	Some days of week or higher *b2*	Most days or higher *b3*	Every day or higher *b4*	Many times a day *b5*
24. Argue when asked to do something	0.82	−0.99	0.10	1.09	1.64	2.26
25. Act stubborn	0.96	−0.93	0.25	1.21	1.78	2.49
26. Disobey or break rules with parents	1.05	−0.80	0.44	1.35	1.89	2.38
27. Say "no" when told to do something	1.14	−0.74	0.47	1.35	1.94	2.69
28. Disobey or break rules during daily routines	1.15	−1.01	0.37	1.54	2.07	2.78
29. Act sassy, talk back, or have a "smart mouth"	1.15	−0.60	0.54	1.39	1.93	2.52
30. Disobey or break rules when frustrated, angry or upset	1.19	−0.59	0.48	1.49	1.97	2.63
31. Do exactly what you just said not to do	1.25	−0.71	0.78	1.63	2.08	2.46
32. Refuse to follow directions	1.37	−0.42	0.79	1.64	2.17	2.66
33. Disobey or break rules with other adults	1.37	−0.75	0.61	1.84	2.31	2.85
34. Ignore directions	1.39	−0.58	0.77	1.73	2.24	2.81
35. Disobey or break rules out of the blue	1.53	−0.21	0.91	1.83	2.27	2.86
36. Disobey or break rules when tired, hungry or sick	1.59	−0.20	0.92	1.83	2.37	3.03
37. Disobey or break rules to get something s/he wanted	1.66	0.08	1.11	1.92	2.35	2.85
38. Not do what you asked no matter what	1.68	−0.29	0.92	1.99	2.52	3.25
39. Argue about just about anything	1.69	0.18	1.14	1.85	2.36	2.91
40. Automatically resist whatever you ask	1.73	0.05	1.17	1.99	2.42	3.00
41. Take things s/he was not allowed to have	1.76	−0.21	1.10	2.12	2.60	3.18
42. Show off or laugh while misbehaving	1.99	0.28	1.33	2.22	2.85	3.26
43. Do risky things s/he knew were not allowed	2.05	0.16	1.40	2.37	2.88	3.42
44. Misbehave in ways that are dangerous	2.52	1.08	2.15	2.77	3.12	3.49
Aggression (95th% threshold=1.50)						
45. Act aggressively when frustrated, angry, or upset	1.50	−0.06	1.03	1.80	2.17	2.58
46. Act aggressively to try to get something s/he wanted	1.62	0.11	1.13	1.94	2.30	2.62
47. Act aggressively with you or other parent	1.82	0.38	1.37	2.08	2.42	2.84
48. Throw something at someone	1.93	0.19	1.32	2.17	2.77	3.20
49. Try to hurt someone to get back at them	2.03	0.59	1.42	2.22	2.76	3.14
50. Hit someone with an object	2.14	0.40	1.65	2.44	2.83	3.36

| | | Category Thresholds | | | | |
	Item Location *mean(b)*	Rarely or higher *b1*	Some days of week or higher *b2*	Most days or higher *b3*	Every day or higher *b4*	Many times a day [a] *b5*
78. Act like s/he did not care when someone felt bad or sad	**2.64**	1.02	**2.18**	**2.91**	**3.31**	**3.81**

Note: Data are derived from IRT-graded response models. Numbers in boldface type indicate category thresholds >95th percentile.

[a] In prior work, we documented similar parameters for Temper Loss derived from a unidimensional model.[3] Temper Loss parameters reported here differ slightly as they are derived from a multidimensional model that accounts for correlation with the other dimensions.