Separating the Domains of Oppositional Behavior: Comparing Latent Models of the Conners' Oppositional Subscale

Ana V. Kuny, Ph.D., Robert R. Althoff, M.D., Ph.D., William Copeland, Ph.D., Meike Bartels, Ph.D., C.E.M. Van Beijsterveldt, Ph.D., Julie Baer, Ph.D., James J. Hudziak, M.D

Objective: Although oppositional defiant disorder (ODD) is usually considered the mildest of the disruptive behavior disorders, it is a key factor in predicting young adult anxiety and depression and is distinguishable from normal childhood behavior. In an effort to understand possible subsets of oppositional defiant behavior (ODB) that may differentially predict outcome, we used latent class analysis of mother report on the Conners' Parent Rating Scales Revised Short Forms (CPRS-R:S). **Method:** Data were obtained from mother report for Dutch twins (7 years old, n = 7,597; 10 years old, n = 6,548; and 12 years old, n = 5,717) from the Netherlands Twin Registry. Samples partially overlapped at ages 7 and 10 years (19% overlapping) and at ages 10 and 12 years (30% overlapping), but not at ages 7 and 12 years. Oppositional defiant behavior was measured using the six-item Oppositional subscale of the CPRS-R:S. Multilevel LCA with robust standard error estimates was performed using the Latent Gold program to control for twin–twin dependence in the data. Class assignment across ages was determined and an estimate of heritability for each class was calculated. Comparisons with maternal report Child Behavior Checklist (CBCL) scores were examined using linear mixed models at each age, corrected for multiple comparisons. **Results:** The LCA identified an optimal solution of four classes across age groups. Class 1 was associated with no or low symptom endorsement (69–75% of the children); class 2 was characterized by defiance (11–12%); class 3 was characterized by irritability (9–11%); and class 4 was associated with elevated scores on all symptoms (5–8%). Odds ratios for twins being in the same class at each successive age point were higher within classes across ages than between classes. Heritability within the two "intermediate" classes was nearly as high as for the class with all symptoms, except for boys at age 12. Children in the Irritable class were more likely to have mood symptoms on the CBCL scales than children in the Defiant class but demonstrated similar scores on aggression and externalizing scales. Children in the All Symptoms class were higher in both internalizing and externalizing scales and subscales. **Conclusions:** The LCA indicates four distinct latent classes of oppositional defiant behavior, in which the distinguishing feature between the two intermediate classes (classes 2 and 3) is the level of irritability and defiance. Implications for the longitudinal course of these symptoms, association with other disorders, and genetics are discussed. J. Am. Acad. Child Adolesc. Psychiatry; 2013;52(2):172-182. **Key Words:** oppositional defiant disorder, twin, latent class analysis.

Along with conduct disorder (CD) and attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD) is one of the leading reasons for referral to youth mental health services.[1] In contrast to CD, which is seen as a severe and inflexible condition,[2] ODD has often been considered a fairly mild condition,[3] possibly because some of the behaviors associated with it approximate normative child development (e.g., losing one's temper, arguing with adults). This thought has persisted, despite evidence that ODD is in fact distinguishable from normal childhood behavior[2,4] and is present in up to 2% of girls and nearly 5% of boys.[5] Despite differences between CD and ODD, research on these disorders has typically combined the two, collapsing them into a single construct.[6–8] In doing so, many studies

Supplemental material cited in this article is available online.

involving ODD and CD fail to consider the two disorders distinctly, and often ODD is excluded altogether. Because ODD and CD are often studied in concert, the specific environmental and genetic contributants to ODD remain elusive. It is often assumed that ODD is caused by poor parenting or by environmental causes; yet, research demonstrates that, separate from CD, there is support for specific genetic factors associated with ODD.[9] Although studies have found that ODD and CD are correlated, the symptoms appear to represent distinct processes.[10] As researchers have begun to separate the disruptive behavior disorders and to examine ODD individually, it has become clear that ODD may not be as benign as previously thought. Instead of serving as prodrome for CD, the ODD diagnosis may in fact play a significant role on its own in the development of a wide range of child psychopathology, including, depression, anxiety, CD, and later the development of antisocial personality disorder.[11]

Subtyping of the ODD diagnosis may be especially important if we hope to understand its association with later development of psychopathology (e.g., more defiant behavior predicting something distinct from ODD with more irritable or reactive features), as well as its association with service use and prescribed treatment adherence. Copeland *et al.*[12] found that ODD emerges as a key disorder in predicting young adult anxiety and depression. Earlier age at onset of ODD symptoms generally results in a poorer prognosis in terms of progression to CD and, ultimately, antisocial personality disorder. In fact, it has been estimated that approximately 30% of children who have an early onset of ODD later progress to develop CD.[1,13] However, it may be important to differentiate between boys and girls, as findings have been mixed. In one study, ODD in girls was found to be associated with increased risk of depression, anxiety, and later ODD but was not associated with increased risk for later development of CD.[14] In examining the course of the disorder, preschool children with ODD are likely to exhibit additional disorders several years later, and with increasing age, comorbidity with ADHD, anxiety, or mood disorders begins to appear.[15] In fact, ODD as a long-term predictor of many other disorders holds in childhood and adolescence even when controlling for other disorders.[12] Furthermore, the distinction among ADHD, ODD, and CD seems to be supported by research, but findings have again been mixed.[4,13,16,17] Similar to CD, the association of ODD and ADHD appears to indicate more severe psychopathology. Compared to children with ADHD only, children with both ODD and ADHD tend to be more aggressive, show a greater range and persistence of problem behaviors, are rejected at higher rates by peers, and underachieve more severely in the academic domain. Children and adolescents with ODD not only appear to have significantly higher rates of comorbid psychiatric disorders, but they also seem to have significantly greater family and social dysfunction relative to other youths with psychopathology.[13,18] Understanding the subtypes of ODD that might predict differential outcomes seems prudent.

A study by Stringaris and Goodman[19] attempted to subtype ODD using three distinct *a priori*–derived dimensions of oppositionality: irritable, hurtful, and headstrong. This study found that all three dimensions were associated with differing manifestations of CD; therefore, the authors concluded that distinct subtypes of oppositionality likely do not exist. Furthermore, they concluded that the three dimensions may suggest differing origins and trajectories to oppositionality, based on the cross-sectional and longitudinal associations that they had seen. This has been followed by studies from Aebi *et al.*,[20] who demonstrated similar dimensions in preschoolers,[21,22] and Rowe *et al.*,[23] who demonstrated that there were few cases of "pure" headstrong behavior. They examined differential prediction of the dimensions and showed that the headstrong dimension was associated with substance disorder and that irritability was associated with later anxiety disorder. Similarly, Kolko and Pardini[24] studied dimensions of treatment resistance and showed that irritability was associated with treatment-resistant ODD, whereas hurtfulness was associated with later treatment-resistant CD. We questioned whether defining subtypes using a bottom-up approach, rather than using *a priori* dimensions might produce a slightly different result. Specifically, we questioned whether latent class analysis (LCA) could be used to refine the ODD phenotype. LCA is a form of person-centered categorical data analysis that allows one to identify latent classes that account for the distribution of cases that have similar categorical response variables.[25] By the nature of the analysis, these classes are mutually exclusive, with each having its own particular pattern of item endorsement. LCA presupposes the existence of discrete latent categories of responding and groups individuals, distinguishing it from factor analysis, which assumes continuous latent variables that group symptoms. LCA results in two metrics: the probability of class membership for each individual, and symptom endorsement probabilities for each class. The class that is most probable for a particular individual or the posterior probability of class assignment can then be used in subsequent analyses. The advantage to this approach is that it is free of preconceived notions about which items should go together and thus allows for a manner of classifying individuals empirically using a bottom-up approach. This approach has been used to study classes of ADHD,[26–28] OCD,[29] juvenile bipolar disorder,[30] tic disorders,[31] and alcohol use disorders,[32] among others.

The objective of the current analysis was to determine whether specific ODB subclasses could be identified using a LCA of mother report on the Conners' Parent Rating Scales Revised Short Forms (CPRS-R:S). Given that the CPRS-R:S does not have hurtful items, we hypothesized that subjects would differ on their levels of headstrong (or defiant) and irritable symptoms, based on the previous literature. With this in mind, we hypothesized that a person-centered, latent class analysis would reveal four latent classes of individual responding: a class with no or few symptoms, a class with mainly irritable symptoms, a class with mainly defiant symptoms, and a class with high levels of all symptoms. Given findings of the stability of the heritability of ODB over time, we expected that the same latent structure would hold at ages 7, 10, and 12 years.

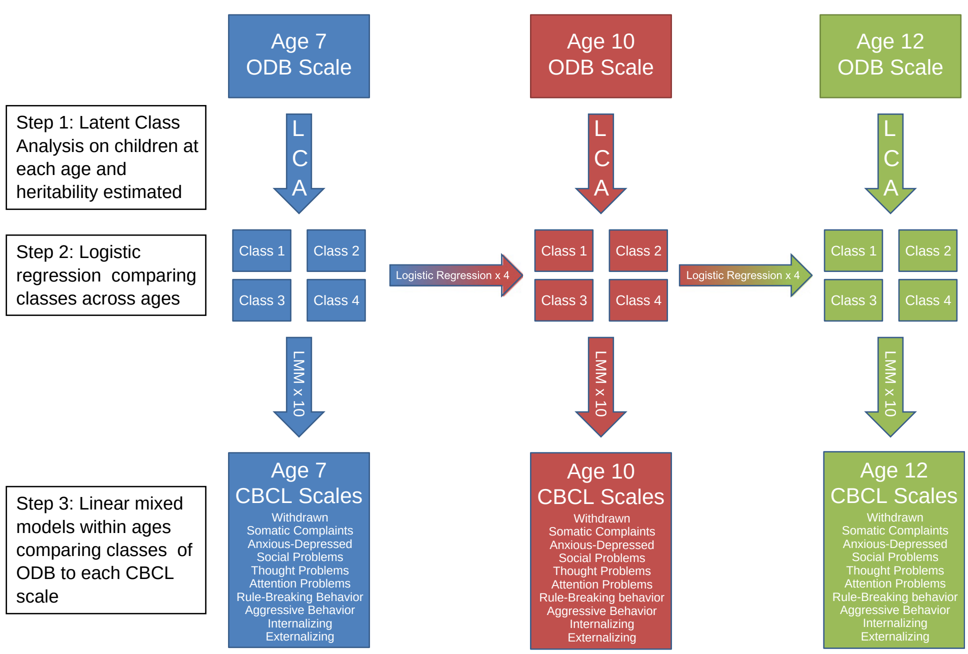
METHOD

This study proceeded in three steps. First, latent class analysis was performed within each age group, and heritability was estimated. Next, the across-age stability of these classes was tested by comparing across ages 7 to 10 years and 10 to 12 years. Third, a comparison of concurrent validity was performed within each age group. This can be seen graphically in Figure 1.

Subjects and Procedure

Data were obtained using mother report for Dutch twin pairs from the Netherlands Twin Registry, kept by the Department of Biological Psychology at the Free University in Amsterdam.[33–35] Starting in 1987, families with twins were recruited a few months after birth. Currently, 40% to 50% of all multiple births are registered by the Netherlands Twin Registry. The data of the present study are derived from a large ongoing longitudinal study that examines the genetic and environmental influences on the development of problem behavior in families with twin's ages 3 to 12 years of age.[33,34] Information from the Conners' forms used

FIGURE 1 Analysis work flow demonstrating the cross-sectional nature of the latent class analysis (LCA) and mixed models with partial longitudinal analysis between ages 7 and 10 years for class membership. Note: CBCL = Child Behavior Checklist; LMM = linear mixed models; ODB = oppositional defiant behavior.



here were introduced later in the data collection using a cohort-based data collection (detailed in Bartels *et al.*[33]). The data from the original 7-year-old cohort are only now turning 12; therefore, there are no individuals with full longitudinal data from ages 7 to 12. The final sample for LCA consisted of Conners' forms for 14,844 children. A total of 5,018 children were sampled more than once (2,214 sampled at both age 7 and 10 years, and 2,804 sampled at both age 10 and 12 years; there were no children sampled at both age 7 and age 12). Thus, a total of 19,862 observations were entered into the latent class analysis. Of the children, 7,597 had data at age 7 (38.2% of total observations), 6,548 had data at age 10 (33.0% of total observations), and 5,717 had data at age 12 (28.8% of total observations). For examining the concurrent validity, Child Behavior Checklist (CBCL) data were included using mother report.

For the present study, data of mother report for 7-, 10-, and 12-year-old twin pairs was examined separately for each age group. Mothers of twins were asked to fill out questionnaires about problem behavior separately. After 2 months, a reminder was sent to the nonresponders, and, when finances permitted, families who had not responded after 4 months were contacted by telephone. Families who did not participate at a certain age were subsequently contacted and allowed to participate in the next scheduled study contact. The overall participation rate for the age groups used in the present study is 66% at age 7, 64% at age 10, and 64% at age 12 (this includes all registered families with a twin pair at a particular age). Previous work on this sample has demonstrated that attrition was random with respect to childhood psychopathology.[33] This study was approved by the institutional review boards of the Free University, Amsterdam, and the University of Vermont.

Measures

Mothers of participants completed the Conners' Parent Rating Scales Revised Short Form (CPRS-R:S). The questionnaire consists of 27 items rated on a four-point Likert scale for symptom severity (i.e., 0 = not true at all, 1 = just a little true, 2 = pretty much true, 3 = very much true). The items are summarized on four scales: Oppositional, Cognitive Problems/Inattention, Hyperactivity, and the ADHD Index. Three of these scales, Oppositional, Cognitive Problems/Inattention, and Hyperactivity, were originally derived from the Conners' Rating Form: Long Form. To provide brief versions of these scales, only items loading the highest (loadings 0.40) from an exploratory factor analysis of the factor scale items on the long form were used.[36] This study specifically used the Oppositional subscale, which consists of six items (Table 1). The internal consistency coefficient for both scales was greater than 0.80 for boys and girls, and the test–retest reliability coefficients for scales were between 0.63 and 0.85 during a period of 6 to 8 weeks.[36]

For the purpose of the LCA, items on the oppositional subscale were recoded such that 0 and 1 were recoded to be 0. Items scored 2 and 3 were recoded to be 1. This approach has been used in the analysis of the ADHD Index on the same scale, and the use of truncation strategies did not change the overall pattern, only the number of children placed into each class.[27] Before using this truncation strategy on these data, we compared and contrasted three possible truncation strategies. Dichotomizing data with 0 and 1 responses grouped together and 2 and 3 responses grouped together resulted in lower residuals and higher explained variance and with model fits that were, quite similar. With all truncation strategies, if the best fitting model was actually a three- or five-class model, these models were essentially equivocal with the four-class model. This information, along with detailed information about the model fitting, is available in Supplement 1, available online.

For examining the concurrent validity, information from the scales of the Child Behavior Checklist/4–18[37] were used. The CBCL is a questionnaire of 118 items developed to measure problem behavior in children and adolescents 4 to 18 years old. Mothers were asked to rate the behavior of the child of the preceding 6 months on

TABLE 1 Comparison of Connors' Subscale Questions and *DSM-IV* Oppositional Defiant Disorder (ODD) Items

***DSM-IV* ODD Items**	**CPRS-R:S Oppositional Items Parent Form**
1. Often loses temper	Q11. Loses temper
2. Is often touchy or easily annoyed by others	Q16. Irritable
3. Often actively defies or refuses to comply with adults' requests or rules	Q20. Actively defies and refuses to comply with adults' requests
4. Often deliberately annoys people	Q24. Deliberately does things that annoy other people
5. Is often angry and resentful	Q2. Angry and resentful
6. Often argues with adults	Q6. Argues with adults
7. Often blames others for his or her mistakes or misbehavior	—
8. Is often spiteful or vindictive	—

Note: Items on the Conners' Parent Rating Scale: Revised Short Form (CPRS-R:S) are labeled by the numbering system from the CPRS-R:S.

a three-point scale. Eight syndrome scales plus two broadband scales (internalizing and externalizing) were composed according to the Dutch scales for the 1991 version, which are the same as the American scales.[38]

Data Analysis

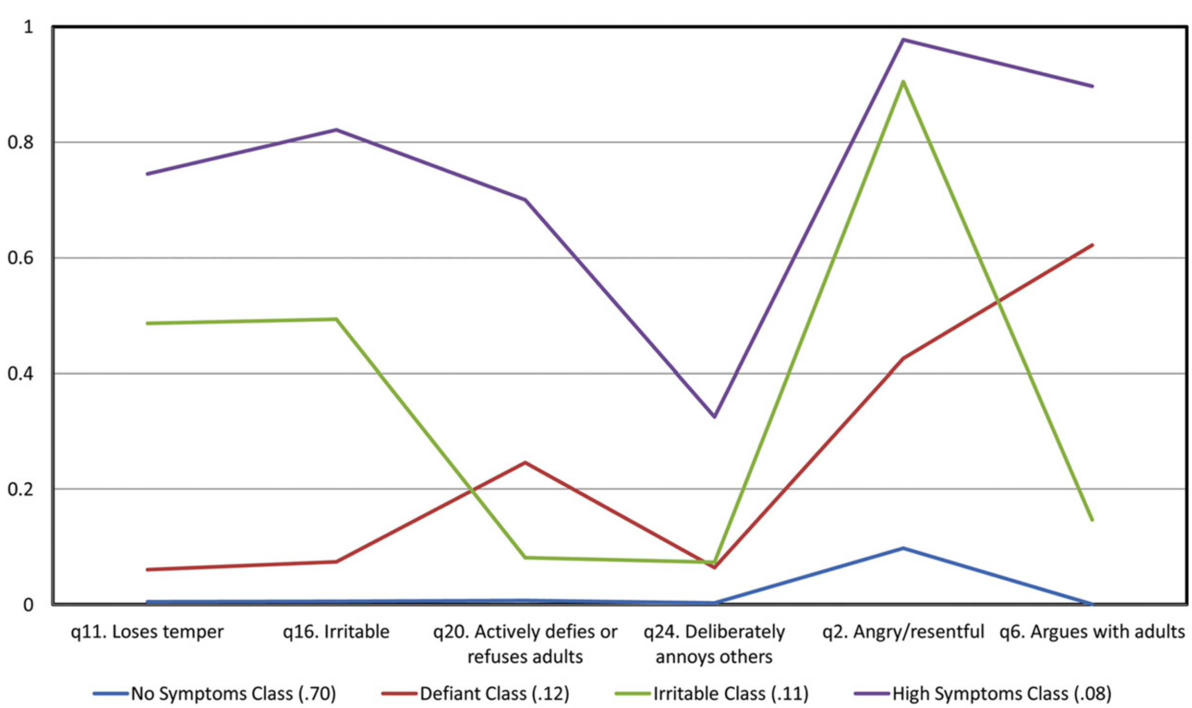
The data analytic workflow is shown graphically in Figure 1. Latent class models were fitted by means of an Expectation Maximization (EM) algorithm[39] with the program Latent Gold.[40] To control for twin-dependence, a multilevel model was used with family number as a grouping variable and standard errors adjusted using the robust (Sandwich) standard error estimator. Models estimating one-class through five-class solutions were compared. Changes in the Bayesian Information Criterion (BIC; a goodness-of-fit index that considers the rule of parsimony) were primarily used, although other metrics were considered as was a factor-mixture model of the data that yielded results consistent with those reported from the LCA (see Supplement 1, available online). LCA proceeded in five steps for each age group. First, models were fit without any restrictions; then bivariate residuals were reduced by allowing for direct effects, and the role of the sex covariate was considered; then significance of the model was examined using nonparametric bootstrapping; and, finally, the fits with models with one additional or one fewer class were examined (see Supplement 1 and Tables S1, S2, and S3, available online).

To examine heritability of the latent classes, the posterior probability of class membership for each latent class for each twin was compared to the posterior probability of class membership for that same latent class in their cotwin. This was performed using intraclass correlations in SPSS software (SPSS Inc., Cary, NC). To calculate a rough estimate of the heritability, the Falconers formula[41] was used by calculating two times the difference of monozygotic (MZ) intraclass correlation and dizygotic (DZ) intraclass correlation ($2*[ICC_{MZ} - ICC_{DZ}]$). In situations in which genetic dominance might be evident (i.e., the MZ correlation was more than twice the DZ correlation), the MZ correlation itself was taken as the estimate of heritability.

Logistic regression was used to predict stability of class membership by examining the likelihood that being in a particular class at one age predicted the categorical outcome of being in all other classes at the next age.

Finally, a set of linear mixed models were performed to examine the relations between the classes and CBCL scales. We controlled for family clustering by choosing one random MZ twin from any MZ twin pair and including a family clustering variable for the DZ pairs. These models used CBCL scale as the dependent variable, family as a random factor, and latent class, sex, and sex $\times$ latent class interaction as categorical fixed effects. Each latent class was compared with the all-symptoms class in the model, and a criterion of $p < .005$ was set for the significance value for each test to control for multiple comparisons. For comparisons

FIGURE 2 Four-class solution for 7-year-old Conners' Parent Rating Scale: Revised, short latent class analysis.



within a fixed effect (e.g., for comparing between latent classes), we examined the confidence interval around the estimate, using the 99.5% confidence interval, again, to be conservative with multiple comparisons.

RESULTS

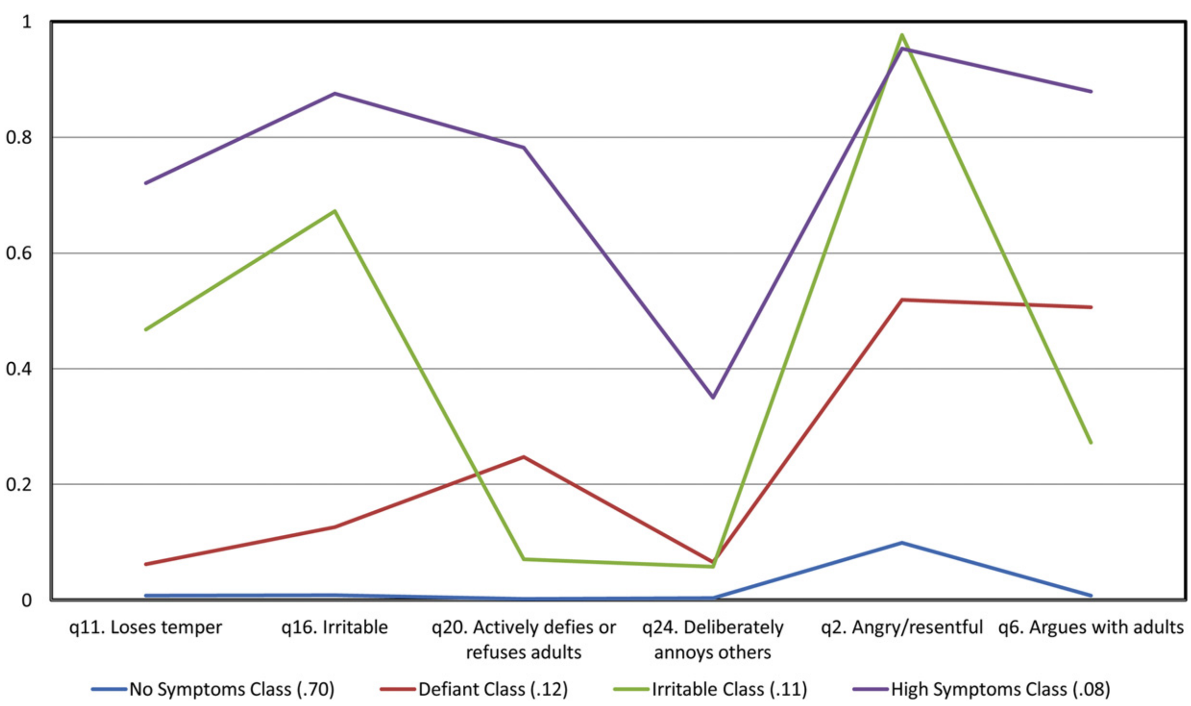
The LCA identified an optimal solution of four classes across age groups co-varying for sex (Figures 2–4). The best model that controlled for twin dependence did not include sex as a covariate, and included direct effects (except age 10) to account for significant bivariate residuals. Distributions of the groups are shown in Table 2.

The across-twin intraclass correlations and estimated heritabilities are provided in Table 3. On the whole, twin correlations within a particular class were higher for MZ twins than for DZ twins, indicating the role of genetics. Estimated heritabilities of the latent classes ranged from 0.13 (12-year-old boys in the No Symptoms class) to 0.59 (7 year-old boys in the Defiant Class). Heritability estimates for males generally decreased in each class from 7 through 12 years, whereas estimates were equivalent or increased for females from 7 through 12. At each age, the estimates for the Defiant and Irritable classes were in the same general range as the estimates for the All Symptoms class, with the exception of age 10, for which correlations were generally lower for the Irritable class. A one-way analysis of variance (ANOVA) did not demonstrate Bonferroni-corrected differences among the classes in terms of heritability.

The results of the logistic regression are shown in Table 4, which presents the ratio of the odds of being in a particular class versus the odds of being in any other class. On the whole, odds ratios were significantly higher between age groups (on the basis of nonoverlapping confidence intervals) for comparisons within a particular latent than across latent classes. In addition, being in class 2 at age 7 did predict being in either Class 2 or class 4 at age 10 and being in class 3 at age 10 did predict being in either class 3 or class 4 at age 12. However, there was no significant crossover in switching between classes 2 and 3.

At all ages, linear mixed models demonstrated a significant effect of latent class on all CBCL scales. Controlling for multiple comparisons, children in the Irritable class had significantly higher mean scores on the anxious–depressed subscale than children in the Defiant class at all ages, and higher mean scores for both withdrawn behavior and the internalizing problems at age 7. Although children in the Defiant class had higher

FIGURE 3 Four-class solution for 10-year-old Conners' Parent Rating Scale: Revised, short latent class analysis.



mean scores on aggressive behavior and externalizing problems than the children in the Low or No Symptoms class, they had equivalent scores on these scales to the children in the Irritable class and lower scores, at all ages, on aggressive behavior and externalizing than children in the High symptoms class. It was only at age 12 that children in the Defiant class began to separate statistically from the No symptoms class in terms of rule-breaking behavior. Full model results are provided in Table S4, available online.

DISCUSSION

The current findings indicate four distinct latent classes of ODB. As expected, the majority of children had low or no symptom endorsement. This should be expected in a general population sample of children. Furthermore, consistent with the literature, which suggest a decrease in ODD diagnoses at 3-year follow-up,[1,13] approximately 75% of children were in the low symptom class by age 12, compared with 69% at both age 7 and 10. The level of either irritability or defiance was the distinguishing difference between class 2 and class 3 in the LCA (Table 2). Specifically, these findings may indicate some ability to separate children who present with oppositional behavior into different patterns of behavior. Children classified into class 2 by mother report were more likely to argue with adults and to be actively defiant, however, this same class of children was not likely to be rated as irritable or hot-tempered. Moreover, this class was also unlikely to demonstrate more internalizing symptoms than the low symptoms class. This finding may suggest that these children's low level of irritability and higher rates of defiance are indicative of children with lower levels of pro-social behavior and more anti-social–like behaviors. This is in contrast to class 3 in the LCA, which includes children whose mothers' endorsed items related to very high levels of irritability (e.g., anger, resentment, and hot-temperedness) accompanied by low levels of defiance. In fact, unlike children in class 4 (the high symptoms class), children in class 3 were not any more defiant than children with low or no symptom endorsement. This finding may indicate a pattern of behavior more associated with the later development of mood disorders, consistent with higher levels of internalizing symptoms on the CBCL in the this class compared with class 1 or 2. This distinction between "irritable non-defiance" and "defiant non-irritability" is consistent with findings in the literature of the distinction between Reactive–affective–defensive–impulsive (RADI) vs. Proactive–instrumental–planned–predatory (PIPP). RADI refers to aggression that is unplanned and accompanied by negative emotions such as anger,

FIGURE 4 Four-class solution for 12-year-old Conners' Parent Rating Scale: Revised, short latent class analysis.

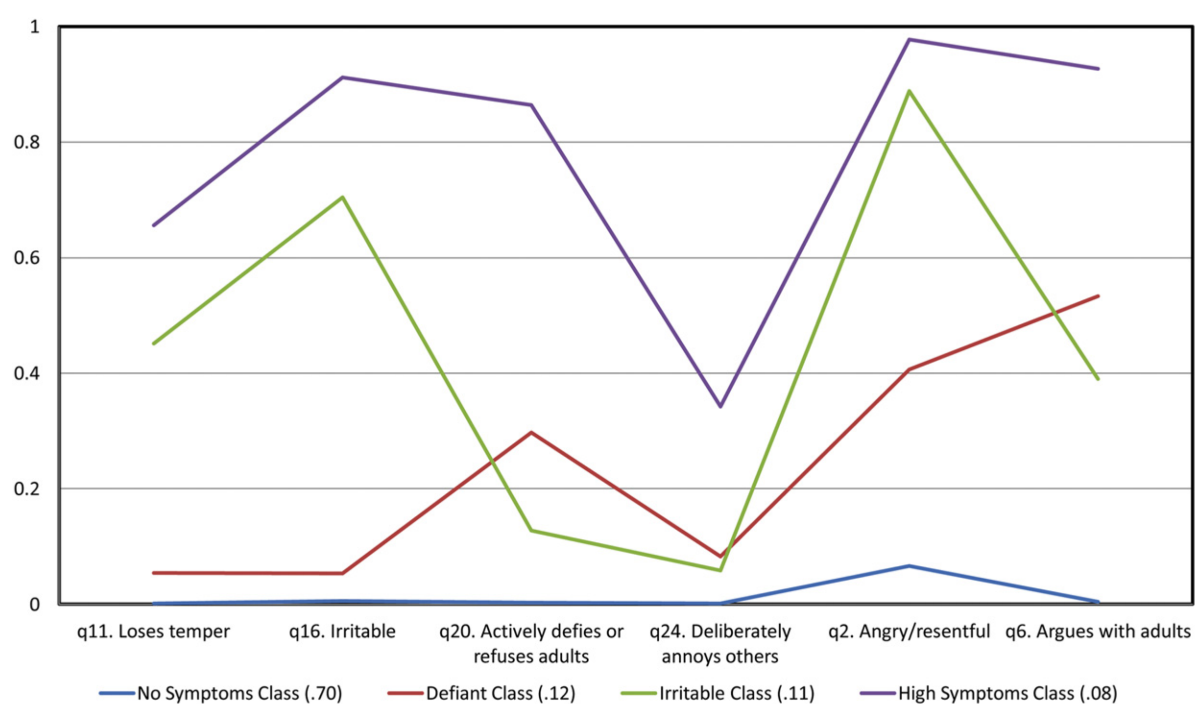


TABLE 2 Item Endorsement Probabilities and Class Membership for Four-Class Latent Class Solution

Age	Class	Class Membership	Prop. Female	Loses Temper	Is Irritable	Actively Defies or Refuses Adults	Deliberately Annoys Others	Is Angry or Resentful	Argues With Adults
7	1	0.70	0.52	0.0046	0.0058	0.0074	0.003	0.0977	0.001
	2	0.12	0.48	0.0609	0.0741	0.2463	0.0643	0.4258	0.6216
	3	0.11	0.44	0.4863	0.4937	0.0812	0.0735	0.9049	0.1472
	4	0.08	0.40	0.7452	0.8213	0.7003	0.3254	0.9779	0.8971
10	1	0.69	0.54	0.0076	0.0088	0.0019	0.0036	0.0992	0.0077
	2	0.12	0.51	0.0619	0.1261	0.2477	0.0657	0.5189	0.506
	3	0.10	0.47	0.4675	0.6727	0.0703	0.058	0.9772	0.2723
	4	0.09	0.42	0.7212	0.8754	0.7825	0.3504	0.9537	0.879
12	1	0.75	0.53	0.0011	0.0054	0.0031	0.0015	0.066	0.004
	2	0.11	0.47	0.0543	0.0531	0.2973	0.0828	0.4064	0.5336
	3	0.09	0.45	0.4517	0.7044	0.1276	0.0584	0.8883	0.3899
	4	0.05	0.46	0.6561	0.9118	0.8645	0.3423	0.9774	0.9268

Note: Class Membership = proportion of the sample at that age placed into the class. Prop. Female = proportion of the class that is female.

TABLE 3 Intraclass Correlations and Estimated Heritabilities Between Twins Within Latent Classes at Ages 7, 10, and 12 Years

Twin Type	Pairs (n)	Class			
		No or Low Symptoms	Defiant	Irritable	High Symptom
Age 7 years					
MZM	605	0.65	0.643	0.333	0.476
DZM	679	0.457	0.349	0.014	0.137
MZF	675	0.617	0.63	0.252	0.405
DZF	622	0.435	0.415	0.016	0.134
DOS	1,223	0.421	0.505	0.079	0.175
Heritability male		0.386	0.588	0.333	0.476
Heritability female		0.364	0.43	0.252	0.405
Age 10 years					
MZM	498	0.69	0.524	0.279	0.574
DZM	510	0.43	0.365	–0.009	0.244
MZF	637	0.711	0.658	0.395	0.522
DZF	513	0.503	0.477	0.081	0.179
DOS	1,122	0.393	0.325	0.018	0.132
Heritability male		0.52	0.318	0.279	0.574
Heritability female		0.416	0.362	0.395	0.522
Age 12 Years					
MZM	488	0.636	0.494	0.31	0.39
DZM	459	0.571	0.377	0.155	0.185
MZF	580	0.748	0.623	0.45	0.411
DZF	423	0.608	0.374	0.123	0.108
DOS	920	0.531	0.445	0.245	0.236
Heritability male		0.13	0.234	0.31	0.39
Heritability female		0.28	0.498	0.45	0.411

Note: Heritability is calculated using Falconer's formula , except in situations where genetic dominance may be present (monozygotic [MZ] correlation ≥ 2 × dizygotic [DZ] correlation) in which case the MZ correlation is used as the estimate. DOS = dizygotic twins of opposite sex; DZF = dizygotic female; DZM = dizygotic male; MZM = monozygotic male; MZF = monozygotic female.

irritability, or fear. PIPP aggression, on the other hand, is associated with positive emotions and is willfully planned and executed.[42] This is the first example, however, that we are aware of in which these distinctions have been reified within an oppositionality scale using a person-centered approach.

The results of the logistic regression (Table 3) performed on the LCA classes suggest that class membership is relatively stable. In fact, at all ages, there was a significant likelihood of homotypic continuity. The only class with significant drift regarding class membership was the 10-year-old group in class 4 (high symptom class), although these children were likely to maintain membership in the high symptom class, they were also likely to shift to class 2 or class 3. This finding is in line with previous studies that suggest that a significant portion of kids with ODD, exit the diagnosis by the age of 12.

There are some limitations that need to be acknowledged. First, despite the fact that the ODD checklist began with all eight items, and factor analyses done by Conners[36] yielded evidence that the six items retained were the most highly loaded, the Conners' Oppositional scale does not include all eight of the *DSM-IV* ODD criteria. Thus the scale used for this analysis assesses oppositional defiant behavior rather than the ODD diagnosis specifically, and the items relating to the hurtful dimension was not included in these analyses. Second, having only one informant means that we cannot be sure whether the results would be different if teachers, other caregivers, or the children themselves provided information. This is work that we are continuing to explore. Moreover, using an all-Dutch sample means that we cannot be sure whether these results are generalizable to other groups of children, although these children have been demonstrated to be similar to the Dutch general population,[33] and the overall levels of psychopathology in children in the Netherlands

TABLE 4 Across-Age Comparison: Odds Ratios and 95% Confidence Intervals

	Age 10 Class 1 No or Low Symptoms	Age 10 Class 2 Defiant	Age 10 Class 3 Irritable	Age 10 Class 4 High Symptom
Age 7 Class 1	**5.22 (4.2–6.4)***	0.51 (0.38–0.69)*	0.27 (0.20–0.36)*	0.12 (0.08–0.17)*
Age 7 Class 2	.57 (0.42–0.77)*	**2.00 (1.3–2.93)***	0.90 (0.54–1.49)	2.04 (1.28–3.28)*
Age 7 Class 3	.27 (.20, .37)*	1.18 (0.73–1.91)	**5.62 (3.97–7.96)***	1.73 (1.04–2.88)
Age 7 Class 4	.15 (.11, .21)*	1.88 (1.22–2.86)*	2.34 (1.56–3.49)*	**10.85 (7.41–15.90)***
	Age 12 Class 1 No or Low Symptoms	**Age 12 Class 2 Defiant**	**Age 12 Class 3 Irritable**	**Age 12 Class 4 High Symptom**
Age 10 Class 1	**7.83 (6.31, 9.71)***	0.29 (0.22–0.38)*	0.12 (0.09–0.18)*	0.07 (0.04–0.13)*
Age 10 Class 2	.57 (.43, .76)*	**2.64 (1.85–3.74)***	1.12 (0.73–1.97)	0.60 (0.26–01.39)
Age 10 Class 3	.38 (.29, .50)*	1.21 (0.77–1.88)	**5.08 (3.54–7.28)***	1.29 (0.68–2.46)
Age 10 Class 4	.08 (.06, .11)*	3.51 (2.50–4.92)*	5.29 (3.70–7.56)*	**7.83 (6.31–9.71)***

Note: Boldface type is used to demonstrate the within-class, across age comparisons.
*$p < .05$.

has been demonstrably similar to that in U.S. populations.[43] In addition, although attrition for general psychopathology was demonstrably missing at random, this might not necessarily hold for latent class assignment. Similarly, it is not completely clear that these model fits would generalize to another sample; although, here, because ages 7 and 12 contain completely different children, with absolutely no overlap, the fact that these two models are so similar represents a large replication in a separate sample. Furthermore, we have conducted a second study on an entirely different sample using a different instrument and have demonstrated similar latent structure and that these classes have predictive validity (Althoff *et al.*, manuscript in preparation). In addition, the limits of odds ratios need to be acknowledged, especially when numbers in classes get small. However, Pearson correlations were also performed for probability of class membership in each class, and the results were essentially the same. Finally, these data were from a mixed cross-sectional/longitudinal sample yielding no individuals with full longitudinal data from ages 7 to 12. Full longitudinal data will be available when all waves reach age 12, at which time the full longitudinal genetic model for these classes can also be fitted.

An understanding of distinct differences between classes is necessary if clinicians and researchers wish to tease apart the specific contributions of environmental and genetic factors to ODD. The assumption that ODB in general and ODD in particular are entirely due to poor parenting to or environmental causes has not been supported by research. Future research must evaluate the complex etiology of ODD apart from CD, which may allow a more accurate and complete picture of presenting oppositional defiant behaviors in both research and clinical settings. The current findings suggest that there are subsets of ODB in the population that may have differential presentation and course. These findings are consistent with recent proposed changes to the ODD diagnosis by the American Psychiatric Association (APA) *DSM-5* committee. Specifically, a proposed change in the reorganization of ODD: "Recommendation 3. Organize symptoms in the criteria for ODD to distinguish emotional and behavioral symptoms." In examining possible changes, the committee found that although behavioral and emotional symptoms both predicted disruptive behavior disorders, mood and anxiety disorders were predicted independently by emotional symptoms.[44] This recommendation is supported by the results presented here that person-centered analyses can distinguish between children with "irritable non-defiance" and "defiant non-irritability." We would predict that children with irritable non-defiance would be more at risk for later mood disorders versus children with defiant nonirritability, who would be more at risk for conduct disorders. New work performed in our laboratory using a similar construct has suggested that this is the case (Kuny, unpublished doctoral thesis; Althoff *et al.*, manuscript in preparation) with children in the defiant nonirritable group demonstrating higher levels of criminal behavior in adulthood, compared with children in the irritable but not defiant group showing a higher rate of mood disorders in adulthood. &

Accepted October 9, 2012.

This article was reviewed under and accepted by ad hoc editor, Daniel S. Pine, M.D.

Drs. Kuny and Althoff contributed equally to this article.

Drs. Kuny, Althoff, Baer and Hudziak are with the Vermont Center for Children, Youth, and Families at the University of Vermont. Dr. Kuny is also with the New Mexico Veterans Administration Hospital in Albuquerque, New Mexico. Dr. Baer is also with the Nationwide Children's Hospital in Columbus, Ohio. Dr. Copeland is with Duke University. Drs. Bartel, Van Beijsterveldt, and Hudziak are with Vrije Universiteit Amsterdam, the Netherlands. Dr. Hudziak is also with Erasmus Medical Center, Rotterdam, the Netherlands.

This work was supported by National Institute of Mental Health (NIMH) grants MH082116 and MH58799, and the Netherlands Organisation for Scientific Research grants 575-25-006, 575-25-012, and 904-57-94.

The authors thank Dorret Boomsma of Vrije Universiteit Amsterdam, the Netherlands, for allowing the use of Netherlands Twin Registry data.

Disclosure: Dr. Althoff has received grant or research support from NIMH and the Klingenstein Third Generation Foundation. Dr. Copeland has received grant or research support from NIMH and the Brain and Behavior Research Foundation. Dr. Bartels has received funding from the National Institute of Diabetes and Digestive and Kidney Disease (NIDDK). Dr. Hudziak has received funding from NIMH and NIDDK. His primary appointment is with the University of Vermont. He has additional appointments with Erasmus University in Rotterdam, Vrije University in Amsterdam, Dartmouth Medical School in Hanover, New Hampshire, and the Avera Institute of Human Behavioral Genetics in Sioux Falls, South Dakota. Drs. Van Beijesterveldt, Kuny, and Baer report no biomedical financial interests or potential conflicts of interest.

Correspondence to Robert Althoff, M.D., Ph.D., Department of Psychiatry, 364 SJ3 FAHC, 1 S. Prospect, Burlington VT 05401; e-mail: ralthoff@uvm.edu

http://dx.doi.org/10.1016/j.jaac.2012.10.005

REFERENCES

1. Loeber R, Burke JD, Lahey BB, Winters A, Zera M. Oppositional defiant and conduct disorder: a review of the past 10 years, part I. J Am Acad Child Adolesc Psychiatry. 2000;39:1468-1484.
2. Keenan K, Shaw DS. Starting at the beginning: Exploring the etiology of antisocial behavior in the first years of life. In: Lahey BB, Moffitt TE, Caspi A, eds. Causes of Conduct Disorder and Juvenile Delinquency. New York, NY: Guilford Press; 2003: p. 153-181.
3. Rey JM, Bashir MR, Schwarz M, Richards IN, Plapp JM, Stewart GW. Oppositional disorder: fact or fiction? J Am Acad Child Adolesc Psychiatry. 1988;27:157-162.
4. Rutter M, Giller H, Hagell A. Antisocial behavior by young people. J Am Acad Child Adolesc Psychiatry. 1999;38:1320-1321.
5. Maughan B, Rowe R, Messer J, Goodman R, Meltzer H. Conduct disorder and oppositional defiant disorder in a national sample: developmental epidemiology. J Child Psychol Psychiatry. 2004; 45:609-621.
6. Burke JD. Relationship between conduct disorder and oppositional defiant disorder and their continuity with antisocial behaviors: evidence from longitudinal clinical studies. In: Shaffer D, Leibenluft E, Rohde LA, eds. Externalizing Disorders of Childhood: Refining the Research Agenda for DSM-V. Arlington, VA: American Psychiatric Association; 2009.
7. Greene RW, Biederman J, Zerwas S, Monuteaux MC, Goring JC, Faraone SV. Psychiatric comorbidity, family dysfunction, and social impairment in referred youth with oppositional defiant disorder. Am J Psychiatry. 2002;159:1214-1224.
8. Kuhne M, Schachar R, Tannock R. Impact of comorbid oppositional or conduct problems on attention-deficit hyperactivity disorder. J Am Acad Child Adolesc Psychiatry. 1997;36:1715-1725.
9. Hudziak JJ, Derks EM, Althoff RR, Copeland W, Boomsma DI. The genetic and environmental contributions to oppositional defiant behavior: a multi-informant twin study. J Am Acad Child Adolesc Psychiatry. 2005;44:907-914.
10. Frick PJ, Lahey BB, Loeber R, Tannenbaum L. Oppositional defiant disorder and conduct disorder: a meta-analytic review of factor analyses and cross-validation in a clinic sample. Clin Psychol Rev. 1993;13:319-340.
11. Loeber R, Burke JD, Pardini DA. Development and etiology of disruptive and delinquent behavior. Annu Rev Clin Psychol. 2009; 5:291-310.
12. Copeland WE, Shanahan L, Costello EJ, Angold A. Childhood and adolescent psychiatric disorders as predictors of young adult disorders. Arch Gen Psychiatry. 2009;66:764-772.
13. Connor DF. Aggression and Antisocial Behavior in Children and Adolescents: Research and Treatment. New York: Guilford; 2002.
14. Rowe R, Maughan B, Pickles A, Costello EJ, Angold A. The relationship between DSM-IV oppositional defiant disorder and conduct disorder: findings from the Great Smoky Mountains Study. J Child Psychol Psychiatry. 2002;43:365-373.
15. Lavigne JV, Cicchetti C, Gibbons RD, Binns HJ, Larsen L, DeVito C. Oppositional defiant disorder with onset in preschool years: longitudinal stability and pathways to other disorders. J Am Acad Child Adolesc Psychiatry. 2001;40:1393-1400.
16. Burke JD, Loeber R, Birmaher B. Oppositional defiant disorder and conduct disorder: a review of the past 10 years, part II. J Am Acad Child Adolesc Psychiatry. 2002;41:1275-1293.
17. Hinshaw SP, Anderson CA. Conduct and oppositional defiant disorders. In: Mash EJ, Barkley RA, eds. Child Psychopathology. New York, NY: Guilford Press; 1996:113–149.
18. 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. 2010;49:484-492.
19. Stringaris A, Goodman R. Three dimensions of oppositionality in youth. J Child Psychol Psychiatry. 2009;50:216-223.
20. Aebi M, Muller UC, Asherson P, *et al*. Predictability of oppositional defiant disorder and symptom dimensions in children and adolescents with ADHD combined type. Psychol Med. 2010; 40:2089-2100.
21. Ezpeleta L, Granero R, de la Osa N, Penelo E, Domènech JM. Dimensions of oppositional defiant disorder in 3-year-old preschoolers. J Child Psychol Psychiatry. 2012;53:1128-1138.
22. Wakschlag LS, 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 Adolesc Psychiatry. 2012;51593-604:e594.
23. Rowe R, Costello EJ, Angold A, Copeland WE, Maughan B. Developmental pathways in oppositional defiant disorder and conduct disorder. J Abnorm Psychol. 2010;119:726-738.
24. Kolko DJ, Pardini DA. ODD dimensions, ADHD, and callous-unemotional traits as predictors of treatment response in children with disruptive behavior disorders. J Abnorm Psychol. 2010; 119:713-725.
25. McCutcheon AL. Latent Class Analysis. Newbury Park: Sage Publications; 1987.
26. Neuman RJ, Heath A, Reich W, *et al*. Latent class analysis of ADHD and comorbid symptoms in a population sample of adolescent female twins. J Child Psychol Psychiatry. 2001;42:933-942.
27. Althoff RR, Copeland WE, Stanger C, *et al.* The latent class structure of ADHD is stable across informants. Twin Res Hum Genet. 2006;9:507-522.
28. Todd RD, Lobos EA, Sun LW, Neuman RJ. Mutational analysis of the nicotinic acetylcholine receptor alpha 4 subunit gene in attention deficit/hyperactivity disorder: evidence for association of an intronic polymorphism with attention problems. Mol Psychiatry. 2003;8:103-108.

29. Althoff RR, Rettew DC, Boomsma DI, Hudziak JJ. Latent class analysis of the Child Behavior Checklist Obsessive-Compulsive Scale. Compr Psychiatry. 2009;50:584-592.
30. Althoff RR, Rettew DC, Faraone SV, Boomsma DI, Hudziak JJ. Latent class analysis shows strong heritability of the child behavior checklist-juvenile bipolar phenotype. Biol Psychiatry. 2006;60:903-911.
31. Nestadt G, Addington A, Samuels J, *et al*. The identification of OCD-related subgroups based on comorbidity. Biol Psychiatry. 15 2003;53:914–920.
32. Rindskopf D. Heavy alcohol use in the ''fighting back'' survey sample: Separating individual and community level influences using multilevel latent class analysis. J Drug Issues. Special Issue: The ''Fighting Back'' program. 2006;36:441-462.
33. Bartels M, van Beijsterveldt CE, Derks EM, *et al*. Young Netherlands Twin Register (Y-NTR): a longitudinal multiple informant study of problem behavior. Twin Res Hum Genet. 2007;10:3-11.
34. Boomsma DI, de Geus EJ, Vink JM, *et al*. Netherlands Twin Register: from twins to twin families. Twin Res Hum Genet. 2006;9:849-857.
35. Boomsma DI, Vink JM, van Beijsterveldt TC, *et al*. Netherlands Twin Register: a focus on longitudinal research. Twin Res. 2002; 5:401-406.
36. Conners CK. Conners Rating Scales–Revised. New York, NY: Multi-Health Systems; 2001.
37. Achenbach TM. Manual for the Child Behavior Checklist/4-18 and 1991 Profile. Burlington, VT: University of Vermont, Department of Psychiatry; 1991.
38. Verhulst FC, van der Ende J, Koot HM. Handleiding voor de CBCL/4-18. Rotterdam: Sophia Kinderziekenhuis/Academisch Ziekenhuis Rotterdam/Erasmus Universiteit Rotterdam, Afdeling Kinder-en jeugdpsychiatrie; 1996.
39. Dempster A, Laird N, Rubin D. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc, Series B.. 1977;39:1-38.
40. Vermunt JK, Magidson J. Latent GOLD User's Manual. Belmont, MA: Statistical Innovations; 2000.
41. Falconer D, MacKay T. Introduction to Quantitative Genetics. Harlow, Essex, UK.: Longmans Green; 1996.
42. Steiner H, Saxena K, Chang K. Psychopharmacologic strategies for the treatment of aggression in juveniles. CNS Spectrums. 2003; 8:298-308.
43. Ivanova MY, Achenbach TM, Dumenci L, *et al*. Testing the 8-syndrome structure of the child behavior checklist in 30 societies. J Clin Child Adolesc Psychol. 2007;36:405-417.
44. Association AP. DSM-5 Development. 2011. Available at: http://www.dsm5.org/ProposedRevisions/Pages/proposedrevision.aspx?rid=106. Accessed July 12, 2011.

SUPPLEMENT 1

METHOD

Model Fitting Description for "Separating the Domains of Oppositional Behavior: Comparing Latent Models of the Conners' Oppositional Subscale." This supplemental material is provided to explain in detail the model fitting for the Latent Class Analysis (LCA), including information about truncation of the response variables and model fitting decisions used to decide an optimal fit for each age group, and to provide information about a confirmatory factor mixture model that was performed on these data.

Latent Class Analysis. Latent class models were fitted by means of an Expectation Maximization (EM) algorithm[39] with the program Latent Gold.[40] To control for twin dependence, a multilevel model was used with family number as a grouping variable, and standard errors were adjusted using the robust (Sandwich) standard error estimator. Models estimating one-class through five-class solutions were compared. Changes in the Bayesian Information Criterion (BIC; a goodness-of-fit index that considers the rule of parsimony) were primarily used, although other metrics were considered, including the Akaike Information Criterion (AIC), the entropy-R2, the number of classification errors, and the average bivariate residual among predictor variables. We expected that the best-fitting models would have low bivariate residuals, high entropy-R2, while retaining relatively low classification errors and lower BIC.

For each age group separately, with every dichotomization of the responses, model-fitting proceeded through the following steps.

Step 1: Fit All Models Without Restrictions. One-class through five-class models were fit using all responses as dichotomous variables (0 = unaffected, 1 = affected), with sex included as a covariate, robust (Sandwich) standard error estimates, all subjects included (missing variables were set by the program to the most common response = 0), and family dependence modeled as a factor with two levels (related or not related) using the multilevel G-Classes approach in Latent Gold. If the BIC was continuing to decrease in the five-class model, a six-class model was also fitted. All metrics were computed on these first, uncorrected models and the reduction in the average bivariate residual and entropy were also considered, but as secondary, supportive measures.

Step 2: Reduce Bivariate Residuals. Bivariate residuals between all variables in the best-fitting model, based on the lowest BIC, were then examined. If the residual was greater than 3.84 (bivariate residuals greater than 3.84 indicate that the remaining association between variables are significant, because, for 1 degree of freedom, 3.84 is significant at the $p < .05$ level), a direct effect among those particular items would then be included in the model and the model re-run. This was repeated until there were no significant bivariate residuals in the model.

Step 3: Examine Sex Covariate. At this point, sex was dropped from the model. If there was a decrease in the BIC with dropping sex, it was dropped from the model.

Step 4: Examine Significance of Model Using Nonparametric Bootstrapping. At this point, the multilevel variable was dropped from the model to allow for a bootstrapping estimate of model fit (bootstrapping is not available for models using a multilevel approach). Bootstrapping of the log-likelihood of the model was then performed using 500 samples, and an estimate of the significance of the model fit was obtained. This bootstrapping procedure generated 500 replication samples from the best (maximum likelihood) solution and re-estimated the model with these replication samples. Hence, a p value greater than .05 indicates a good fit, as the bootstrapped p value indicates the proportion of replication samples that had a higher log-likelihood than the actual sample.

Step 5: Examine N+/−1 Class Solution if Necessary. If the bootstrapped p value for the best model identified by the BIC was not greater than .05, the next most appropriate model was then fitted, by returning to Step 2 and running through the same steps with this class and supplanting the previously determined best model if appropriate.

Once the best model for a particular dichotomization of variables was determined, then all ages and dichotomizations were examined together to decide on the best dichotomization and models to use at each age and overall. Narratives for each age are provided and refer to the findings in Table S1, available online. Narratives are provided for each dichotomization type. "012_3" Refers to 0, 1, and 2 being

TABLE S1 Steps in Analysis of Age 7 Latent Class Analysis

Analysis	LL	BIC	AIC	Npar	Bootstrap *p* Value	Classif. Error	Entropy	Average Bivariate Residual
012_3								
1-Class, Multilevel	−6314.79	12692.12	12643.58	7		0.00	1.00	866.01
2-Class, Multilevel	−4861.15	9874.20	9756.30	17		0.01	0.82	6.89
3-Class, Multilevel	−4791.91	9825.08	9637.82	27		0.03	0.68	4.45
4-Class, Multilevel	−4744.28	9819.18	9562.57	37		0.03	0.70	0.33
5-Class, Multilevel	−4731.99	9883.95	9557.98	47		0.03	0.69	0.13
4-Class, Multilevel, no sex	−4759.54	9796.09	9581.09	31		0.03	0.70	0.32
4-Class, No Multilevel, no sex	−4861.22	9963.69	9776.43	27	.62	0.02	0.75	0.27
01_23								
1-Class, Multilevel	−17412.71	34887.97	34839.42	7		0.00	1.00	1194.17
2-Class, Multilevel	−13605.71	27363.32	27245.41	17		0.03	0.84	24.22
3-Class, Multilevel	−13365.74	26972.73	26785.47	27		0.10	0.68	14.99
4-Class, Multilevel	−13154.70	26640.01	26383.40	37		0.09	0.71	0.67
5-Class, Multilevel	−13129.09	26678.16	26352.19	47		0.10	0.51	0.20
4-Class, Multilevel, direct q20-24	−13147.21	26633.97	26370.42	38		0.09	0.71	0.48
4-Class, Multilevel, direct q20-24, no sex	−13171.40	26628.73	26406.80	32		0.09	0.71	0.47
4-Class, No Multilevel, direct q20-24, no sex	−13405.65	27061.49	26867.30	28	.26	0.09	0.68	0.36
0_123								
1-Class, Multilevel	−28740.38	57543.31	57494.76	7		0.00	1.00	1001.83
2-Class, Multilevel	−24113.80	48379.50	48261.60	17		0.06	0.78	45.19
3-Class, Multilevel	−23583.93	47409.12	47221.86	27		0.14	0.69	20.67
4-Class, Multilevel	−23200.33	46731.27	46474.66	37		0.16	0.70	1.47
5-Class, Multilevel	−23147.51	46714.98	46389.01	47		0.20	0.68	1.36
6-Class, Multilevel	−23102.79	46714.91	46319.58	57		0.24	0.66	0.23
5-Class, Multilevel, q2-q6	−23116.84	46662.58	46329.68	48		0.19	0.68	0.47
5-Class, Multilevel, q2-q6, no sex	−23165.94	46689.30	46411.88	40		0.20	0.67	0.32
5-Class, No Multilevel, q2-q6	−23878.84	48106.17	47835.69	39	<.001	0.23	0.61	0.28
4-Class, Multilevel, q2-q6	−23169.72	46678.98	46415.43	38		0.16	0.70	0.60
4-Class, Multilevel, q2-q6, no sex	−23196.98	46679.90	46457.96	32		0.16	0.70	0.61
4-Class, No Multilevel, q2-q6, no sex	−23950.39	48150.98	47956.79	28	<.001	0.19	0.63	0.30

Note: AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; Classif. = Classification; LL = log-likelihood.

considered "negative" and 3 being considered "positive" for endorsement of a symptom. "01_23", Similarly groups 0 and 1 responses together versus 2 and 3 responses. "0_123" Follows suit with 0 being considered "negative" and any other response considered "positive." We did make an attempt at not dichotomizing variables at all. However, when entered into the analysis in an ordinal LCA in Latent Gold, the models did not converge on any parsimonious solution, presumably in part because of the low proportion of items endorsed in the highest category in this general population twin sample and in part because the ordinal LCA tries to capture the continuous nature of the variables and ends up as a latent-trait-like model, better captured by factor mixture modeling (FMM).

Age 7 Model Fitting

012_3. Step 1 demonstrated the lowest BIC in the four-class analysis (see Table S1, available online). Examination of the average bivariate residuals indicated a 13-fold drop-off between three-class and four-class solutions. Step 2 indicated no significant bivariate residuals. Step 3 demonstrated that dropping sex from the model improved the BIC. Step 4 demonstrated that without the multilevel twin modeling, the model had a *p* value greater than .05, indicating an acceptable fit. Step 5 was not necessary. A model with four classes, without direct effects and without a sex covariate, was the best model.

01_23. Step 1 demonstrated the lowest BIC in the four-class analysis. Examination of the average bivariate residuals indicated a 22-fold drop off between three-class and four-class solutions. Step 2 indicated a significant bivariate residual between items 20 and 24 so a direct effect was included. Step 3 demonstrated that dropping sex from the model improved the BIC by five points. Step 4 demonstrated that without the multilevel twin modeling, the model had a *p* value greater than .05, indicating an acceptable fit. Step 5 was not necessary. A model with four classes, with one direct effect and without a sex covariate, was the best model.

0_123. Step 1 demonstrated the lowest BIC in the six-class analysis, although it was essentially identical to the five-class BIC, but with more classification errors and worse entropy. Based on this, the five-class solution was chosen. Examination of the average bivariate residuals indicated a 14-fold drop off between three-class and four-class solutions, and little change between four-class and five-class. Step 2 indicated a significant bivariate residual between items 2 and 6 in the five-class solution, so a direct effect was included. Step 3 demonstrated that dropping sex from the model increased the BIC. Step 4 demonstrated that without the multilevel twin modeling, the model had a *p* value less than .05, indicating a less than acceptable fit. Step 5 re-ran the models using a four-class solution, which also required a direct effect between items 2 and 6. The best-fitting four-class model also did not have a bootstrapped *p* value greater than 0.05 and had a BIC that was higher than the best-fitting five-class model. A model with five classes, with one direct effect, with sex covariate, was the best model, but was essentially equivocal with the four-class model.

Age 10 Model Fitting

012_3. Step 1 demonstrated the lowest BIC in the 3-class analysis (see Table S2, available online). Examination of the average bivariate residuals indicated a very small drop-off between two-class, three-class, and four-class solutions. Step 2 indicated significant bivariate residuals between items 6 and 20. Step 3 demonstrated that dropping sex from the model improved the BIC. Step 4 demonstrated that without the multilevel twin modeling, the model had a *p* value less than .05, indicating a less than acceptable fit. Step 5 re-ran the models using a four-class solution, which required a direct effect between items 20 and 24. The best fitting four-class model also did not have a bootstrapped *p* value greater than .05 and had a BIC that was higher than the best-fitting three-class model. Similarly, the five-class model also did not fit. A model with three classes, with one a direct effect and without a sex covariate, was the best model, but was essentially equivocal with the four-class or five-class model.

01_23. Step 1 demonstrated the lowest BIC in the 4-class analysis. Examination of the average bivariate residuals indicated an 18-fold drop off between 3-class and 4-class solutions. Step 2 indicated no significant bivariate residuals, so no direct effects were included. Step 3 demonstrated that dropping sex from the model improved the BIC by 3.7 points. Step 4 demonstrated that

TABLE S2 Steps in Analysis of Age 10 Latent Class Analysis

Analysis	LL	BIC (LL)	AIC (LL)	Npar	Bootstrap *p* Value	Classif. Error	Entropy	Average Bivariate Residual
012_3								
1-Class, Multilevel	−5944.40	11950.30	11902.80	7		0.00	1.00	756.36
2-Class, Multilevel	−4532.16	9213.69	9098.32	17		0.01	0.83	3.95
3-Class, Multilevel	−4475.06	9187.37	9004.12	27		0.05	0.64	2.25
4-Class, Multilevel	−4443.34	9211.79	8960.68	37		0.06	0.63	0.89
5-Class, Multilevel	−4432.18	9277.35	8958.37	47		0.05	0.68	0.46
3-Class, Multilevel, q20-q6	−4460.88	9167.79	8977.76	28		0.05	0.62	0.74
3-Class, Multilevel, q20-q6, no sex	−4471.76	9154.41	8991.52	24		0.05	0.62	0.74
3-Class, No Multilevel, q20-q6, no sex	−4535.65	9255.82	9113.30	21	.01	0.04	0.67	0.48
4-Class, Multilevel, q20-q24	−4439.31	9212.53	8954.63	38		0.05	0.65	0.52
4-Class, Multilevel, q20-q24, no sex	−4450.66	9182.50	8965.31	32		0.05	0.65	0.49
4-Class, No Multilevel, q20-q24, no sex	−4533.36	9312.75	9122.72	28	.01	0.04	0.68	0.26
5-Class, Multilevel, q20-q24	−4428.32	9278.41	8952.64	48		0.04	0.71	0.13
5-Class, Multilevel, q20-q24, no sex	−4440.46	9232.39	8960.92	40		0.06	0.63	0.07
5-Class, No Multilevel, q20-q24, no sex	−4527.62	9362.79	9125.25	35	.02	0.04	0.70	0.11
01_23								
1-Class, Multilevel	−15756.46	31574.42	31526.92	7		0.00	1.00	1127.98
2-Class, Multilevel	−12014.01	24177.40	24062.02	17		0.03	0.85	20.60
3-Class, Multilevel	−11792.28	23821.81	23638.56	27		0.09	0.72	7.84
4-Class, Multilevel	−11663.48	23652.08	23400.97	37		0.10	0.70	0.42
5-Class, Multilevel	−11640.50	23693.98	23375.00	47		0.12	0.68	0.14
4-Class, Multilevel, no sex	−11688.19	23648.78	23438.38	31		0.10	0.70	0.45
4-Class, No Multilevel, no sex	−11883.60	24004.45	23821.21	27	.06	0.10	0.68	0.37
0_123								
1-Class, Multilevel	−24650.30	49362.10	49314.59	7		0.00	1.00	941.55
2-Class, Multilevel	−20357.65	40864.68	40749.30	17		0.06	0.80	35.20
3-Class, Multilevel	−19899.37	40035.98	39852.74	27		0.14	0.70	14.96
4-Class, Multilevel	−19606.22	39537.55	39286.44	37		0.16	0.71	1.67
5-Class, Multilevel	−19539.93	39492.85	39173.86	47		0.19	0.69	1.55
6-Class, Multilevel	−19515.60	39532.06	39145.21	57		0.21	0.67	1.48
5-Class, Multilevel, q2-q6	−19518.07	39457.91	39132.14	48		0.19	0.68	0.64
5-Class, Multilevel, q2-q6, no sex	−19578.99	39509.45	39237.97	40		0.17	0.70	0.54
5-Class, No Multilevel, q2-q6	−20124.10	40590.88	40326.19	39	<.001	0.24	0.61	0.26
4-Class, Multilevel, q2-q6	−19582.57	39499.04	39241.14	38		0.15	0.71	0.90
4-Class, Multilevel, q2-q6, q2-q16	−19593.27	39529.23	39264.54	39		0.16	0.69	1.12
4-Class, Multilevel, q2-q6, no sex	−19609.70	39500.59	39283.40	32		0.15	0.71	0.78
4-Class, No Multilevel, q2-q6, no sex	−20223.77	40693.58	40503.54	28	<.001	0.20	0.63	0.36

Note: AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; Classif.= Classification; LL = log-likelihood.

without the multilevel, twin modeling, the model had a *p* value greater than .05, indicating an acceptable fit. Step 5 was not necessary. A model with four classes, with no direct effects and without a sex covariate, was the best model.

0_123. Step 1 demonstrated the lowest BIC in the five-class analysis. Examination of the average bivariate residuals indicated a ninefold dro- off between three-class and four-class solutions, and little change between four-class and five-class solutions. Step 2 indicated a significant bivariate residual between items 2 and 6 in the five-class solution, so a direct effect was included. Step 3 demonstrated that dropping sex from the model increased the BIC. Step 4 demonstrated that without the multilevel twin modeling, the model had a *p* value less than .05, indicating a less than acceptable fit. Step 5 re-ran the models using a four-class solution, which also required a direct effect between items 2 and 6. When a direct effect between items 2 and 16 were included, the entropy dropped and the BIC increased, so only one direct effect was included, but sex could not be dropped. The best fitting four-class model also did not have a bootstrapped *p* value greater than .05 and had a BIC that was higher than the best-fitting five-class model. A model with five classes, with one direct effect, with sex covariate, was the best model, but was essentially equivocal with the four-class model.

Age 12 Model Fitting

012_3. Step 1 demonstrated the lowest BIC in the three-class analysis (see Table S3, available online). Examination of the average bivariate residuals indicated a very small drop-off between two-class and three-class solutions, but a ninefold drop-off between three- and four-class solutions. Step 2 indicated three significant bivariate residuals (q20-q6, q20-q24, and q24-q16) requiring three direct effects. Step 3 demonstrated that dropping sex from the model improved the BIC. Step 4 demonstrated that without the multilevel twin modeling, the model had a *p* value greater than .05, indicating acceptable fit. Step 5 was not necessary. A model with three classes, with three direct effects, and without a sex covariate was the best model.

01_23. Step 1 demonstrated the lowest BIC in the four-class analysis. Examination of the average bivariate residuals indicated a sevenfold drop-off between three-class and four-class solutions. Step 2 indicated one significant bivariate residual so a direct effect between items 11 and 24 were included. Step 3 demonstrated that dropping sex from the model improved the BIC by 16 points. Step 4 demonstrated that without the multilevel twin modeling, the model had a *p* value less than .05, indicating a less than acceptable fit. Step 5 examined the five-class solution. No direct effects were needed. Sex could be dropped, but it also had an unacceptable fit on bootstrapping. A model with four classes, with one direct effect, and without sex covariate was the best model.

0_123. Step 1 demonstrated the lowest BIC in the five-class analysis. Examination of the average bivariate residuals indicated a ninefold drop off between three-class and four-class solutions, with a two-fold change between four-class and five-class. Step 2 indicated a significant bivariate residual between items 2 and 6 in the five-class solution, so a direct effect was included. Step 3 demonstrated that dropping sex from the model increased the BIC. Step 4 demonstrated that without the multilevel twin modeling, the model had a *p* value less than .05, indicating a less than acceptable fit. Step 5 re-ran the models using a four-class solution, which also required a direct effect between items 2 and 6 in addition to items 2 and 16. Sex could not be dropped. The best-fitting four-class model also did not have a bootstrapped *p* value greater than .05 and had a BIC that was higher than the best-fitting five-class model. A model with five classes, with one direct effect, and with a sex covariate was the best model, but was essentially equivocal with the four-class model.

LCA Overall Conclusions and Choice of Best Models to Continue

The analyses were very consistent across the various ages. When items were dichotomized with "just a little true" grouped with "not true at all" (01_23), the best-fitting models had the highest entropy and, generally, lower<<? residuals. In each of these situations, a four-class model was the best, and the profiles (Figure 1) look essentially identical. Consequently, they were chosen as the best-fitting models, and subsequent analyses were performed on these data.

TABLE S3 Steps in Analysis of Age 12 Latent Class Analysis

Analysis	LL	BIC (LL)	AIC (LL)	Npar	Bootstrap *p Value*	Classif. Error	Entropy	Average Bivariate Residual
012_3								
1-Class, Multilevel	−3786.86	7634.28	7587.72	7		0.00	1.00	791.65
2-Class, Multilevel	−2721.89	5590.85	5477.78	17		0.01	0.85	5.16
3-Class, Multilevel	−2668.72	5571.02	5391.44	27		0.02	0.72	2.92
4-Class, Multilevel	−2631.71	5583.52	5337.42	37		0.02	0.75	0.32
5-Class, Multilevel	−2619.25	5645.11	5332.51	47		0.02	0.77	0.26
3-Class, Multilevel, q20-q6	−2652.33	5546.89	5360.66	28		0.03	0.67	1.61
3-Class, Multilevel, q20-q6, q20-q24	−2647.76	5546.41	5353.53	29		0.03	0.67	0.90
3-Class, Multilevel, q20-q6, q20-q24, q24-q16	−2642.56	5544.65	5345.11	30		0.03	0.66	0.49
3-Class, Multilevel, q20-q6, q20-q24, q24-q16, no sex	−2658.97	5542.87	5369.94	26		0.04	0.65	0.53
3-Class, No Multilevel, q20-q6, q20-q24, q24-q16, no sex	−2741.75	5682.47	5529.49	23	.61	0.03	0.61	0.40
01_23								
1-Class, Multilevel	−11396.69	22853.95	22807.39	7		0.00	1.00	961.33
2-Class, Multilevel	−8492.28	17131.64	17018.57	17		0.02	0.86	14.88
3-Class, Multilevel	−8337.46	16908.50	16728.92	27		0.08	0.72	7.40
4-Class, Multilevel	−8254.47	16829.04	16582.94	37		0.07	0.72	1.01
5-Class, Multilevel	−8222.65	16851.91	16539.30	47		0.10	0.69	0.49
4-Class, Multilevel, q11-q24	−8247.33	16823.40	16570.65	38		0.08	0.72	0.57
4-Class, Multilevel, q11-q24, no sex	−8265.08	16807.01	16594.17	32		0.07	0.72	0.53
4-Class, No Multilevel, q11-q24, no sex	−8475.46	17193.14	17006.91	28	<.001	0.06	0.74	0.34
5-Class, Multilevel, no sex	−8250.56	16838.51	16579.11	39		0.08	0.73	0.24
5-Class, No Multilevel, no sex	−8470.08	17234.31	17008.17	34	<.001	0.06	0.77	0.46
0_123								
1-Class, Multilevel	−21412.11	42884.79	42838.23	7		0.00	1.00	833.19
2-Class, Multilevel	−17479.07	35105.21	34992.14	17		0.05	0.82	36.74
3-Class, Multilevel	−17111.36	34456.31	34276.73	27		0.10	0.76	15.83
4-Class, Multilevel	−16775.24	33870.57	33624.48	37		0.14	0.74	1.76
5-Class, Multilevel	−16704.86	33816.32	33503.71	47		0.17	0.71	0.80
6-Class, Multilevel	−16667.57	33828.26	33449.15	57		0.19	0.70	0.96
5-Class, Multilevel, q2-q6	−16683.66	33782.58	33463.32	48		0.17	0.71	0.10
5-Class, Multilevel, q2-q6, no sex	−16745.57	33837.19	33571.14	40		0.15	0.72	0.11
5-Class, No Multilevel, q2-q6	−17353.79	35044.97	34785.57	39	<.001	0.22	0.64	0.07
4-Class, Multilevel, q16-q2	−16752.42	33833.58	33580.84	38		0.15	0.72	1.34
4-Class, Multilevel, q16-q2, q2-q6	−16737.97	33813.34	33553.95	39		0.15	0.71	0.30
4-Class, Multilevel, q16-q2, q2-q6, no sex	−16770.49	33826.46	33606.97	33		0.15	0.70	0.24
4-Class, No Multilevel, q16-q2, q2-q6	−17410.51	35097.85	34885.01	32	<.001	0.18	0.64	0.20

Note: AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; Classif = Classification; LL = log-likelihood.

TABLE S4 Child Behavior Checklist Scale (CBCL) Mean Scores by Latent Class for Each Age

CBCL Scale	Class Mean (SD)				Significance Pattern (Criterion $p < .005$)
	No or Low Symptoms	Defiant	Irritable	High Symptom	
Age 7 years					
Withdrawn	1.33 (1.52)	1.91 (1.86)	2.70 (2.25)	3.11 (2.51)	No<Def<Irr=Hi
Somatic complaints	0.78 (1.29)	1.18 (1.60)	1.26 (1.66)	1.61 (1.96)	No<Def<>Irr<Hi
Anxious–depressed	1.71 (2.21)	2.87 (3.04)	4.09 (3.86)	4.95 (4.31)	No<Def<Irr=Hi
Social problems	0.95 (1.38)	1.71 (1.93)	2.22 (2.20)	2.93 (2.54)	No<Def, Def<>Irr, Def<Hi, Irr=Hi
Thought problems	0.27 (0.71)	0.57 (1.11)	0.82 (1.51)	1.26 (1.89)	No<>Def<>Irr, Def<Hi, No<Irr,Hi, Irr=Hi
Attention problems	2.27 (2.39)	3.95 (3.05)	4.58 (3.39)	6.06 (3.84)	No<>Def<>Irr<Hi
Rule-breaking behavior	0.77 (1.11)	1.64 (1.62)	1.62 (1.59)	2.91 (2.47)	No<>Def<>Irr<Hi
Aggressive behavior	4.52 (4.03)	9.22 (5.12)	10.7 (5.59)	15.88 (6.93)	No<Def<>Irr<Hi
Internalizing	3.77 (3.77)	5.83 (4.89)	7.85 (5.96)	9.34 (6.73)	No<>Def<Irr=Hi
Externalizing	5.29 (4.76)	10.86 (6.20)	12.32 (6.65)	18.79 (8.79)	No<Def<>Irr<Hi
Age 10 years					
Withdrawn	1.25 (1.50)	2.14 (2.10)	2.83 (2.50)	3.29 (2.51)	No<Def<>Irr<Hi
Somatic complaints	0.81 (1.34)	1.15 (1.65)	1.27 (1.79)	1.68 (2.11)	No<Def<>Irr<>Hi, Def<Hi
Anxious–depressed	1.88 (2.47)	3.27 (3.32)	4.70 (4.10)	5.82 (4.89)	No<>Def<Irr<Hi
Social problems	1.06 (1.54)	2.06 (2.16)	2.39 (2.37)	3.33 (2.86)	No<Def, Def<>Irr<Hi
Thought problems	0.26 (0.73)	0.52 (1.06)	0.72 (1.21)	1.37 (2.03)	No<>Def, Def<>Irr<Hi
Attention problems	2.25 (2.45)	3.97 (3.22)	4.62 (3.37)	6.12 (3.93)	No<Def<>Irr<Hi
Rule breaking behavior	0.63 (1.05)	1.60 (1.62)	1.54 (1.53)	3.29 (2.82)	No<>Def<>Irr<Hi
Aggressive behavior	3.82 (3.72)	8.40 (5.08)	10.13 (5.20)	16.15 (7.1)	No<Def<>Irr<Hi
Internalizing	3.85 (3.99)	6.37 (5.54)	8.44 (6.56)	10.45 (7.5)	No<Def<>Irr<Hi
Externalizing	4.45 (4.40)	10.01 (6.16)	11.67 (6.19)	19.44 (9.3)	No<Def<>Irr<Hi
Age 12 years					
Withdrawn	1.12 (1.54)	1.94 (1.95)	3.02 (2.51)	3.92 (3.04)	No<>Def<>Irr<Hi
Somatic complaints	0.70 (1.26)	1.01 (1.47)	1.28 (1.77)	1.78 (2.17)	No<>Def<>Irr<>Hi, Def<Hi
Anxious–depressed	1.73 (2.38)	3.14 (3.28)	5.64 (4.89)	6.88 (5.18)	No<>Def<Irr<>Hi
Social problems	1.01 (1.52)	1.97 (2.25)	2.59 (2.59)	3.62 (2.98)	No<>Def, Def<>Irr<Hi
Thought problems	0.21 (0.69)	0.42 (1.00)	0.85 (1.56)	1.38 (1.98)	No<>Def<>Irr<Hi
Attention problems	2.12 (2.37)	3.76 (2.94)	5.01 (3.63)	6.46 (4.28)	No<Def<All, Def<>Irr<>Hi
Rule-breaking behavior	0.61 (1.15)	1.72 (1.70)	1.92 (1.82)	3.52 (2.71)	No<Def<>Irr<Hi
Aggressive behavior	3.55 (3.51)	8.62 (4.90)	11.04 (5.68)	17.00 (6.95)	No<Def<>Irr<Hi
Internalizing	3.50 (4.02)	5.93 (5.14)	9.49 (7.19)	11.92 (8.33)	No<>Def<>Irr<Hi
Externalizing	4.16 (4.25)	10.34 (6.10)	12.96 (6.97)	20.52 (8.95)	No<Def<>Irr<Hi

Note: For Significance Pattern column: Def = Defiant class; Hi = High Symptoms class; Irr = Irritable class; No = No Symptoms class; < denotes significantly different or 99.5% confidence intervals not overlapping; <> denotes 99.5% confidence intervals overlapping; = denotes not significantly different.

Factor Mixture Models

One of the limitations of latent class analysis is that it assumes that the items in the analysis are independent. Because we know from previous research that some of these items fall into groups, or factors, we wanted to model also this aspect of these data. To combine the strengths of the person-centered LCA with the correlations among the items, we used FMM.[25] < <? FMM consists of two types of latent variables: categorical latent class variables, which indicate class membership; and continuous latent factor(s), which indicate the common content of observed variables. The results from an FMM analysis combines the latent class model and the common factor model, and has a single categorical and one or more continuous latent variables.[25] < <? This approach is different from the latent class model, because observed variables within classes are not assumed to be independent and the covariation among observed variables is modeled using underlying continuous factors. In the present study, FMM allowed us to determine whether our latent classes were fitting the data even when we included the covariation among the items. It also allowed us to test the assumptions of the underlying factor structure of the items in a confirmatory fashion.

Confirmatory factor mixture modeling was completed using FMM software (Mplus). Models estimating a two-factor and a one- or two-class solution, with and without correlation among the factors, were performed for the largest sample (the 7-year-old age group). Factors in the one-factor model were set up so that all variables loaded on a continuous factor. Factors in the two-factor models were determined based on the literature and represented ''irritability'' with the following items: loses temper, is irritable, and is angry and resentful, versus ''defiance'' with the items actively defies and refuses to comply with adults' requests, deliberately does things that annoy other people, and argues with adults. Classes were set up to confirm or deny the presence of underlying latent classes that, in a model that already has the items correlated with each other, were set to indicate severity. We compared the change in the Bayesian Information Criterion (BIC) when moving from a one-class to a two-class solution along with the Lo–Mendell–Rubin and bootstrapped log-likelihood ratio tests.

FMM Results

The results of the FMM identified that a two class, two factor model solution was better than a one-class, two-factor model based on the following: a decrease in the BIC from 37,495.62 in the one-class solution to 37,476.39 in the two-class solution; a significant Lo–Mendell–Rubin adjusted LRT test with the values 107.39 and $p < .001$, and a significant parametric bootstrap likelihood ratio test result ($p < 0.04$). This finding is consistent with the results of the multilevel LCA, and indicates that defiance and irritability can be modeled as continuous variables alongside the person-centered categorical variables. Furthermore, the distinction between classes in the FMM is similar to the distinction between classes 2 and 3 in the LCA. Specifically, the difference between the two classes in the LCA and the two factors in the FMM was irritability and defiance, although, notably, this was a confirmatory FMM, so it was designed to demonstrate this difference.