

The Structure and Stability of Externalizing and Internalizing Problem Behavior During Early Adolescence

E. Reitz,¹ M. Deković,¹ and A. M. Meijer²

Received January 21, 2003; revised January 6, 2005; accepted January 19, 2005

The first aim of this study was to examine the structure of externalizing and internalizing problem behavior during early adolescence. Our second aim was to determine the stability of these problems for boys and for girls over time. A total of 650, 13–14-year-olds filled out (an expanded version of) the Youth Self-Report [YSR; *Manual for the Youth Self-Report and 1991 Profile*. Department of Psychiatry, University of Vermont, Burlington] 2 times with a 1-year interval. By using confirmatory factor analyses (CFA) to test a series of competing models, a hierarchical model provided the best representation of the structure of problem behavior at both Time 1 and Time 2: externalizing and internalizing problem behavior represent distinct aspects but the model also demonstrates the existence of comorbidity at a higher level. This model appeared to be stable over time for both boys and girls. The relative stability of problem behavior was found to be high for boys and girls. Absolute stability for both externalizing and internalizing problems appeared to be higher for boys scoring in the lower range of problem behavior.

KEY WORDS: externalizing; internalizing; structure; stability; adolescents.

More than 60% of children get involved in some kind of problem behavior in the course of adolescence. The early adolescence is an especially vulnerable period because of the multiple and simultaneously occurring changes both within and outside of the young adolescent (pubertal development, cognitive changes, school transition, etc.). This is indicated by an increase in prevalence of a variety of clinical disorders and behavior problems dur-

ing and following puberty (McCord, 1990; Moffitt, 1993; Siegel and Scovill, 2000).

There has been much debate about whether different problem behaviors during adolescence reflect a single underlying dimension or whether they are better conceived as multiple phenomena (e.g., Jessor and Jessor, 1977; Donovan *et al.*, 1988; Donovan and Jessor, 1985; Gillmore *et al.*, 1991). Use of different instruments and various age groups makes it hard to compare the outcomes of studies that examined the structure of problem behavior. Also, the majority of the results apply to the structure of criminal, delinquent, or otherwise overtly disturbing problem behavior including antisocial behavior, substance use, and precocious sexual behavior. However, these concern only 1 aspect of adolescent problem behavior. Nowadays, adolescent problem behavior is conceptualized as 2 empirically derived syndromes: externalizing problems (including delinquency and aggression) and internalizing problems (including depression, anxiety, and withdrawal) (Achenbach, 1991a, 1991b). Little is known about the structure of internalizing problem behavior. Accordingly, the first aim of this study is to examine the structure of

¹Senior Researcher, Utrecht University, The Netherlands. Received PhD in social sciences, University of Amsterdam. Major research interests include problem behavior, family relations, and peer relations during adolescence. To whom correspondence should be addressed at Department of Child and Adolescent Studies, Utrecht University, Heidelberglaan 1, P.O. Box 80140, 3508 TC Utrecht, The Netherlands; e-mail: e.reitz@fss.uu.nl.

²Full Professor, Utrecht University, The Netherlands. Received PhD in social sciences: child and family studies, from University of Nijmegen. Major research interests include adolescent social development, problem behavior, and family relations during adolescence.

³Assistant Professor, University of Amsterdam, The Netherlands. Received PhD in psychology, University of Amsterdam. Major research interests include family influence on childhood illness and problem behavior.

externalizing *and* internalizing problem behavior during adolescence.

Jessor and Jessor (1977) were one of the first authors who found a 1-factor structure underlying adolescent externalizing problem behavior. Drinking, problem drinking, marijuana use, delinquent behavior, and precocious sexual intercourse correlated positively with each other and correlated negatively with conventional behavior including church attendance and academic performance. The authors proposed that these behaviors might constitute a 'syndrome' of problem behavior in adolescence. That is, they suggested that there is a single factor of unconventionality underlying adolescent problem behavior. Indeed, some studies found support for a 1-factor structure of externalizing problems in adolescence (Allen *et al.*, 1994; Ary *et al.*, 1999a,b; Donovan *et al.*, 1988; Donovan and Jessor, 1985; Farrell *et al.*, 1992; Flannery *et al.*, 1999).

In contrast, other studies showed support for a multi-factor structure (Farrell *et al.*, 2000; Gillmore *et al.*, 1991). These multiple factors are often interrelated, which suggests that there might be a unitary factor of general problem behavior at a higher order of analysis (Gillmore *et al.*, 1991). In other words, these studies have found support for a hierarchical structure of problem behavior. It includes first-order factors that represent distinct types of problem behavior, but at the same time shows that these first-order factors are related to a more general (second-order) problem behavior factor.

Findings of McGee and Newcomb (1992) supported the existence of a syndrome of problem behavior as a meaningful second-order (but not first-order) factor from early adolescence into adult life. More recently, Farrell *et al.* (2000) found a 3-factor structure that differentiated drug use, delinquency, and aggression. However, another model in which these 3 factors loaded on a higher order factor (Problem Behavior) fitted the data equally well, suggesting that each behavior is, in part, a manifestation of a more general tendency of problem behavior and, in part, a unique phenomenon.

The above studies exclusively focused on externalizing problem behavior. Studies that examined both externalizing and internalizing problems consistently found 2 separate factors. For instance, results of Ingersoll and Orr (1989) showed a Behavioral Risk and an Emotional Risk factor. The Behavioral Risk factor consisted of items that clustered in 3 categories: substance abuse (e.g., use of alcohol, use of drugs), sexual behaviors (e.g., had sex, pregnancy), and delinquent behaviors (e.g., arrested, suspended). The Emotional Risk factor was composed of 8 items and reflected feelings of being upset, lonely, tense, sad, etc. Brack *et al.* (1994) found 2 dimensions that

corresponded to internalizing problem behavior (poor health status, lonely, tense, upset, and nervous) and externalizing problem behavior (substance use, school problems, sexual intercourse, running away, and arrest). Hartman *et al.* (1999) used the Child Behavior Checklist (CBCL; Achenbach, 1991a) and compared several models with a confirmatory factor analysis (CFA). They found that the 2-factor model, including externalizing problems and internalizing problems, provided the best fit of the data.

These 3 studies indicate that externalizing and internalizing behavior should be considered as separate forms of problem behavior. Neither of them have tested the possibility of a higher-order factor, despite the fact that there is much evidence for a positive association between internalizing and externalizing problems, suggesting the existence of comorbidity (see for reviews: Angold and Costello, 1993; Loeber and Keenan, 1994; McConaughy and Skiba, 1993; Zeitlin, 1999; and Zoccolillo, 1992). The reviews show no doubt about the existence of a co-occurrence between externalizing and internalizing problem behavior. However, the results are mixed with regard to the strength of the association, depending on whether the measured behaviors are general or specific. For instance Zoccolillo (1992) reported a comorbidity rate of 48–69% for emotional disturbance and conduct disorder (CD)/oppositional disorder (OD), 2 rather general behaviors, but reported smaller percentages between CD/OD and more specific internalizing behaviors like depression (15–31%) and anxiety disorders (7.1–30.5%). Also, Angold and Costello (1993) reported a high rate of comorbidity between depression and CD/OD, ranging from 22 to 83%, but a study by Overbeek *et al.* (2001) showed a very weak co-occurrence between emotional disturbance and delinquency, correlation coefficients ranging from 0.08 to 0.10.

These differences in findings can be explained by the samples studied, the types of behaviors measured, and types of instruments used. For instance, clinically referred children or adolescents may show more co-occurrence between externalizing and internalizing problems than children or adolescents from the general population because subjects with multiple disorders are more likely to be in need of therapy than those with single disorders, the so-called Berkson's bias (Berkson, 1946; Zoccolillo, 1992).

To summarize, the inclusion of a higher-order factor in the analyses is important because externalizing and internalizing problem behavior have been studied (and treated) as 2 completely separate types of problems, although recent research on comorbidity indicate that this distinction might not be so straight forward. It can be hypothesized that adolescents from the general population

will have a relative weak co-occurrence between externalizing and internalizing problem behavior and the structure of problem behavior will therefore result in a 2-factor model. However, considering the fact that each behavior can be, in part, a manifestation of a more general tendency and, in part, a unique phenomenon (Farrell *et al.*, 2000), a hierarchical factor structure can also apply to problem behavior. The present study therefore examines which conceptualization of problem behavior provides a better fit of self-reported problems during adolescence in a community sample. Three models are tested: a 1-factor model (general problem behavior), a 2-factor model (externalizing and internalizing problem behavior), and a 2-factor model with a higher-order factor of general problem behavior.

Though gender differences in prevalence of different types of problem behavior are well established (e.g., Allgood-Merten *et al.*, 1990; Leadbeater *et al.*, 1999; Moffitt *et al.*, 2001), gender differences in the structure of problem behavior are examined less often. Therefore, in the present study the structure is also tested regarding gender differences. Further, because many studies made use of cross-sectional data, stability of the structure of problem behavior has never been tested. The longitudinal data in the present study gives the opportunity to test whether the structure would be consistent over time.

Another aim of this study is to examine how problem behaviors develop as a function of age or time; whether these behaviors change or remain stable. Research shows substantial stability over time of problem behavior of children and adolescents using adults' or teachers' reports (Koot and Verhulst, 1990; Verhulst and Althaus, 1988; Verhulst and van der Ende, 1992). One of the first reports on the stability of self-reported problem behaviors of 11–16-year-old adolescents also showed a high stability over a 2-year period (Verhulst and Van Wattum, 1993): The stability coefficient for total problem scores (an aggregate of 8 problem behavior scales) on the Youth Self-Report (YSR; Achenbach, 1991b) was 0.63. The average stability of the total sample was 0.60/0.61 for internalizing and externalizing problem behavior respectively and stability was higher for girls than for boys.

Also Ferdinand *et al.* (1995) assessed the 4-year course of externalizing and internalizing problem behavior from adolescence into young adulthood in a general population sample, who were initially assessed with the YSR and reassessed with the Young Adult Self-Report (Achenbach, 1990) at follow-up (aged 15–18 years at the beginning of the study). They found an average 4-year stability coefficient for total problem scores of 0.49. The stability coefficients of the separate syndrome scores of

the YSR ranged from 0.31 to 0.47. Boys and girls did not differ significantly.

The above-mentioned results about the stability of problem behavior all apply to *relative stability* (i.e., the consistency of an individual's rank order within a group). Another way to conceptualize or operationalize stability is the so-called *absolute stability* (i.e., the consistency in a construct's absolute level when that construct is measured over time). Conceptually, relative stability is independent of absolute stability (Alder and Scher, 1994; Holsen *et al.*, 2000; Loeber *et al.*, 2000).

Both studies of Verhulst *et al.* (1993) and Ferdinand *et al.* (1995) also examined the absolute stability and show similar results: 42% versus 36% of the adolescents who initially scored high on problem behavior (total problem scores) remained in the higher range, whereas more than 70% of the adolescents in both studies who scored low on problem behavior, remained in the lower range. The percentage of adolescents who moved from the higher to the lower range of problem behavior is thus higher than the percentage of adolescents who moved from the lower to the higher range of problem behavior. Still, more than 2/3 of the adolescents remained scoring high levels of problem behavior over time, which is fairly stable.

In sum, the present study focuses on 2 main questions. First, we will examine the structure of externalizing and internalizing problem behavior during early adolescence by testing 3 competing models in a community sample. The structure of problem behavior will be tested for both genders and consistency of the structure over time will also be investigated. Second, we will examine the relative *and* absolute stability of externalizing and internalizing problem behavior over a 1-year period. Because early adolescence is the most turbulent time during adolescence as many changes are occurring within (pubertal development) and outside of the individual (for instance school transition), it could be expected that the stability of problem behavior is rather low.

METHOD

Sample and Procedure

The sample was drawn from 3 secondary schools in the Netherlands. At Time 1 the sample consisted of 650 adolescents between 12 and 15 years old ($M = 13.36$; $SD = 0.55$ years) from the 8th Grade. All adolescents obtained passive consent from their parents to participate in the study. Adolescents completed a battery of questionnaires during regular school hours. Researchers stayed in the classroom during completion.

After a 1-year interval, the same adolescents were tested once more. The schools were visited again and questionnaires were sent to the homes of the adolescents who had left school or could not be reached at school (due to sickness or truancy), including a letter containing instructions and a postpaid return envelope. Nonresponders were called at home to ask whether they could still fill out the questionnaire and send it back. A total of 563 adolescents participated again at Time 2, i.e. the attrition rate was 13%. The distribution of boys and girls for both waves was about equal (T1/T2: N boys = 328/272; N girls = 322/291) and the self-reported ethnicity was mostly Dutch (T1 = 88.4%; T2 = 89.5%).

The 87 nonresponders at Time 2 were between 13 and 15 years old ($M = 13.67$; $SD = 0.62$ years). There were more boys than girls (N boys = 56; N girls = 31). To examine whether the attrition group differed from the main group on problem behavior, t -tests were performed. Significant differences were found only on Delinquent Behavior ($t = 3.72$; $p < 0.001$), with the attrition group scoring higher.

Instruments

Problem Behavior

The YSR (Achenbach, 1991b; Verhulst *et al.*, 1996) was used to obtain adolescent reports on their own behavioral/emotional problems. The YSR assesses 2 broad-band syndromes, *Externalizing* and *Internalizing*. The Externalizing syndrome consists of the scales Delinquent Behavior (e.g., "I steal from home") and Aggressive Behavior (e.g., "I fight a lot"). Alphas for this study for Time 1/Time 2 were 0.67/0.69 for Delinquent Behavior and 0.82/0.82 for Aggressive Behavior. In addition, 2 new scales were developed conform the YSR format to expand the range of externalizing problems that assess School Problems and Disobedience. A first reason to include these scales is that these problem behaviors occur frequently in non-clinical groups during this developmental period. Since the YSR is originally developed for a clinical group, including these scales broadens the scope of the YSR which might than be more suitable as a measure of problem behavior for a community sample of adolescents. A second reason is that many researchers have often used one or both of these constructs in their studies when examining externalizing problem behavior during adolescence (e.g., Brack *et al.*, 1994; Gillmore *et al.*, 1991; Maggs *et al.*, 1995; Rothbaum and Weisz, 1994). The 2 scales thus seem to be a relevant expansion of externalizing problem behavior.

Items for the new scales were selected from questionnaires used in previous studies that examined problem behavior and were drawn from the literature based on face validity. A total of 7 items were used to measure School Problems, assessing the extent of problem behavior in class, cheating, and not paying attention. Examples are "I have been sent out of class for misbehavior" (Gillmore *et al.*, 1991) and "I copy homework from others" (Fletcher *et al.*, 1999). Alphas of this scale for Time 1/Time 2 were 0.65/0.74.

The second scale, Disobedience of parents, included 8 items. The scale was developed to measure the extend that adolescents disobey or ignore their parents and are pushing through their own wishes or desires. Examples are "I refuse to do domestic tasks" (Peeters, 1994) and "I ignore prohibitions from my parents" (Deković, 1999). The internal consistencies for Time 1/Time 2 were 0.71/0.74. The items of both scales were also rated on a 3-point Likert scale used in the YSR.

To demonstrate that these 2 new scales are indicators of externalizing problem behavior and not internalizing problem behavior, correlations between these scales and the original YSR scales were examined. It appeared that School Problems and Disobedience correlated stronger with the externalizing problem behavior scales (between 0.57 and 0.63 for School Problems and between 0.53 and 0.58 for Disobedience), than with the internalizing problem behavior scales (between 0.19 and 0.25 for School Problems and between 0.21 and 0.27 for Disobedience).

The Internalizing syndrome of the YSR consists of Anxious/Depressed (e.g., "I feel worthless or inferior"), Withdrawn (e.g., "I rather be alone than with others"), and Somatic Complaints (e.g., "I feel overtired"). Alphas for Time 1/Time 2 were 0.84/0.87 for Anxious/Depressed, 0.65/0.68 for Withdrawn, and 0.73/0.76 for Somatic Complaints.

Plan of Analysis

To test the structure of problem behavior we conducted CFA using LISREL 8.53 (Jöreskog and Sörbom, 1993). Three models were tested: a 1-factor model, a 2-factor model without a higher-order factor, and finally, a 2-factor model with a higher-order factor. The first model specifies 7 externalizing and internalizing problem behaviors underlying one single factor of problem behavior. The 2-factor model differentiates between externalizing problem behavior, including delinquent behavior, aggressive behavior, school problems, and disobedience as indicators, and internalizing problem behavior, including anxiety/depression, withdrawal, and somatic complaints

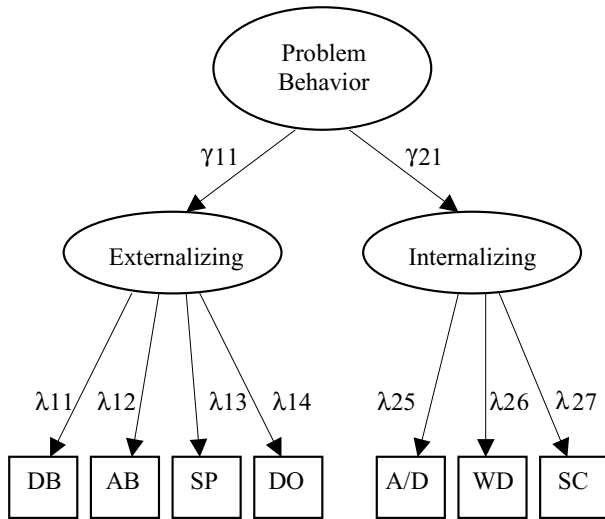


Fig. 1. Two-factor higher-order model. DB: Delinquent Behavior; AB: Aggressive Behavior; SP: School Problems; DO: Disobedience; A/D: Anxious/Depressed; WD: Withdrawn; and SC: Somatic Complaints.

as indicators. The third model is similar to the 2-factor model but includes a single factor of problem behavior at a higher-order level (see Fig. 1).

Because our data are somewhat skewed, violating the normality assumption that is needed for the maximum likelihood estimation method, an unweighted least squares (ULS) estimation method was used (Wolins, 1995). Correlation matrices were used as input. Model fit was evaluated using (1) the chi-square likelihood ratio statistic, (2) the Root Mean Square Error of Approximation (RMSEA), (3) the Non-Normed Fit Index (NNFI), and (4) the Comparative Fit Index (CFI). The chi-square provides a significance test of the null hypothesis that the

model is correct. Because this statistic is strongly dependent on sample size and may cause small differences to be significant, other fit indices were included to evaluate the fit of a model. The fit was to be judged acceptable by an RMSEA of less than 0.07 and by NNFI and CFI values greater than 0.90 (Hartman *et al.*, 1999).

To test which of the 3 competing models fitted the data best at Time 1 and Time 2, the models were compared using a standard “decrement-to-chi-square” test in which the respective goodness of fits (and degrees of freedom) of 2 models were differenced (Willet and Sayer, 1994). The model that best described the data (in terms of parsimony and goodness of fit) was then selected. This selected model was used for multigroup (boys versus girls) comparison tests, to examine gender differences in the magnitude of the structural coefficients.

Next, we examined the stability of externalizing and internalizing problem behavior. The *relative stability* was examined by evaluating the correlation coefficients between individual scores on Time 1 and Time 2 for the same constructs. The *absolute stability* was examined in 2 ways. First, analysis of variance was used to examine changes in the absolute level of (continuous) scores over time. Second, the problem behavior scores were trichotomized (clinical, subclinical, and normal range) and these 3 groups were compared over time.

RESULTS

Structure of Problem Behavior

The Pearson correlations on which the factor analyses were based are reported in Table I, separately for boys and girls at Time 1 and Time 2.

Table I. Intercorrelations Among Indicators of Problem Behaviors at Time 1 and Time 2 for Boys and Girls Separately

Problem behaviors	Girls							Boys						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1. Delinquent behavior	0.45	0.67	0.64	0.56	0.37	0.32	0.42	0.50	0.61	0.61	0.49	0.21	0.20	0.24
2. Aggressive behavior	0.57	0.59	0.61	0.57	0.47	0.32	0.47	0.57	0.61	0.58	0.59	0.33	0.28	0.32
3. School problems	0.58	0.55	0.53	0.45	0.29	0.21	0.31	0.59	0.58	0.48	0.40	0.07 ^a	0.14 ^b	0.17
4. Disobedience	0.53	0.46	0.40	0.56	0.26	0.24	0.35	0.61	0.58	0.50	0.56	0.34	0.29	0.19
5. Anxious/depressed	0.38	0.51	0.24	0.26	0.54	0.74	0.50	0.13 ^b	0.34	0.12 ^b	0.17	0.57	0.64	0.40
6. Withdrawn	0.34	0.33	0.24	0.30	0.71	0.54	0.41	0.22	0.25	0.22	0.24	0.62	0.60	0.36
7. Somatic complaints	0.35	0.39	0.25	0.16	0.55	0.42	0.60	0.28	0.26	0.17	0.29	0.32	0.30	0.49

Note. Intercorrelations at Time 1 for boys and girls are below the diagonal; intercorrelations for Time 2 are above the diagonal. Stability coefficients are on the diagonal.

^aNot significant.

^b $p < 0.05$; all other correlations significant at $p < 0.01$.

Table II. Model Fit Indices and Model Comparison Tests for Time 1 and Time 2

Model label	Model fit indices						Model comparison tests		
	<i>df</i>	χ^2	<i>p</i>	RMSEA	NNFI	CFI	Δdf	$\Delta\chi^2$	<i>p</i> (<i>d</i>)
Time 1									
Model 1: 1-factor	14	631.17	0.000	0.260	0.82	0.90			
Model 2: 2-factor	13	95.54	0.000	0.099	0.99	0.99			
Model 1 versus Model 2							1	535.63	<i>p</i> < 0.001
Model 3: higher order factor	11	11.98	0.370	0.012	0.98	0.99			
Model 1 versus Model 3							3	619.19	<i>p</i> < 0.001
Model 2 versus Model 3							2	83.56	<i>p</i> < 0.001
Time 2									
Model 1: 1-factor	14	520.92	0.000	0.250	0.86	0.91			
Model 2: 2-factor	13	74.21	0.000	0.092	0.99	0.99			
Model 1 versus Model 2							1	446.71	<i>p</i> < 0.001
Model 3: higher order factor	11	26.66	0.005	0.050	0.98	0.99			
Model 1 versus Model 3							3	494.26	<i>p</i> < 0.001
Model 2 versus Model 3							2	47.55	<i>p</i> < 0.001

Chi-square tests and fit indices for the 1-factor model, 2-factor model, and 2-factor model with higher-order factor at Time 1 and Time 2 for the total sample are reported in Table II.

In general, results were consistent across both waves. The 1-factor model did not adequately fit the data. For both Time 1 and Time 2, the chi-square was significant and other indices also showed poor fit. The second model improved significantly in fit over the 1-factor model and although the chi-square was still significant, the fit indices show acceptable fit. The fit of the third model was even better. Comparison of the goodness of fits (and degrees of freedom) of this model with the second model showed a significant improvement of the third model over the 2-factor model. This indicates that a 2-factor model with a higher order factor is the best fitting model for the total sample at Time 1 and Time 2.

After evaluating the overall fit of the model combining data for boys and girls, multigroup comparison

tests were performed to test whether the 2-factor higher-order factor structure was consistent across gender groups. First, all parameters were estimated separately within each group (unconstrained model). Then parameters were constrained to be identical for boys and girls (constrained model). Specifically, factor loadings of the latent variables on their indicators of problem behavior ($=\lambda$) were set equal and loadings of the higher-order factor on the externalizing factor and the internalizing factor ($=\gamma$) were set equal for both genders.

It appeared that the higher-order factor model was not identical for boys and girls for both Time 1 and Time 2 (see Table III). The results of the chi-square differences test indicated that the model in which the loadings were set to be equal differed from the unconstrained model. This means that the factor loadings of boys and girls differ and thus that there are gender differences in structure.

For both Time 1 and Time 2 the results show that 2 of the 3 indicators of internalizing problem behavior, i.e.

Table III. Two Factor Higher-Order Multigroup Analysis for Boys and Girls at Time 1 and Time 2

Model label	Model fit indices						Model comparison tests		
	<i>df</i>	χ^2	<i>p</i>	RMSEA	NNFI	CFI	Δdf	$\Delta\chi^2$	<i>p</i> (<i>d</i>)
Time 1									
Model 1: Free structure	22	23.32	0.380	0.014	0.97	0.99			
Model 2: Equal λ and γ	29	63.18	0.000	0.060	0.98	0.99			
Model 1 versus Model 2							7	39.86	<i>p</i> < 0.001
Time 2									
Model 1: Free structure	22	27.56	0.190	0.030	0.97	0.98			
Model 2: Equal λ and γ	29	65.98	0.000	0.067	0.98	0.99			
Model 1 versus Model 2							7	38.42	<i>p</i> < 0.001

Anxious/Depressed and Somatic Complaints, had somewhat higher factor loadings for girls as compared to boys. Furthermore, the correlations of the 2 first-order factors with the higher-order factor were also higher for girls than for boys on both waves. Because of these gender differences, in the next section, the stability of problem behavior is analyzed for boys and girls separately.

Stability

Stability of Structure

The higher-order factor structure showed the best fit at Time 1 and Time 2 for boys and girls. However, we cannot conclude from these cross-sectional findings that the model is stable over time. In order to test the stability of the model the 2 measurement points were combined in 1 factor analysis. The latent variables were allowed to load only on the variables within their own time frame. The measurement errors of the same construct were allowed to correlate over time (e.g., error of delinquent behavior at Time 1 with the error of delinquent behavior at Time 2). The same first-order factors and the higher-order factor were also allowed to correlate over time.

For boys, this model yielded the following fit: $\chi^2(69) = 127.34, p < 0.000; RMSEA = 0.056; NNFI = 0.99; CFI = 0.99$. The fit indices show an acceptable fit and thus can be concluded that the higher-order factor model is stable over time. The same result holds for girls ($\chi^2(69) = 133.26, p < 0.000; RMSEA = 0.057; NNFI = 1.00; CFI = 1.00$).

Relative stability

To determine to what extent individuals tended to maintain their rank order irrespective of changes in mean

level of the group scores, Pearson correlation coefficients were computed. These stability coefficients are shown in Table I on the diagonal of the 2 matrices. The coefficients range from 0.45 to 0.61 and according to Cohen's criteria most of them can be regarded as large (>0.50) and 3 of them can be regarded as medium (between 0.30 and 0.50). To test whether there are gender differences in stability coefficients, Fisher Z-transformations were used. No significant differences were found between boys and girls, both groups show equal stability in their problem behavior.

Absolute stability

The extent of individuals' consistency in the absolute level of externalizing and internalizing problem behavior over time was measured with a repeated measures model (General Linear Model) with one within-factor (Time) and one between-factor (Gender). Means and standard deviations for time and gender and significant differences are reported in Table IV.

Results showed main effects for Time and Gender. For all problem behaviors, with the exception of Anxious/Depressed and Withdrawn, differences were found in the problem scores between Time 1 and Time 2. Scores on Delinquent Behavior, School Problems, and Disobedience were higher at Time 2 compared to Time 1. Scores on Aggressive Behavior and Somatic Complaints were lower at Time 2 compared to Time 1. Girls scored higher than boys on all 3 internalizing problems but not on externalizing problems. Post hoc analysis showed that the 2 groups differed from each other both at Time 1 and Time 2. Not a single problem behavior showed an interaction effect.

These mean score differences say little about the clinical relevance of these changes (Verhulst and Van Watum, 1993). In the next step, we examined the

Table IV. Means and Standard Deviations of Indicators of Problem Behaviors at Time 1 and Time 2 for Girls and Boys Separately

Problem behavior	Time 1				Time 2				F-value	
	Girls		Boys		Girls		Boys		Time	Gender
	M	SD	M	SD	M	SD	M	SD		
1. Delinquent Behavior	0.29	0.21	0.35	0.23	0.35	0.25	0.36	0.24	10.39**	3.51
2. Aggressive Behavior	0.40	0.23	0.42	0.28	0.38	0.26	0.38	0.24	9.38**	0.06
3. School Problems	0.39	0.26	0.35	0.28	0.42	0.32	0.40	0.32	10.67**	1.96
4. Disobedience	0.35	0.28	0.35	0.31	0.37	0.32	0.40	0.30	6.65*	0.28
5. Anxious/Depressed	0.39	0.31	0.26	0.23	0.38	0.33	0.23	0.23	2.59	39.79**
6. Withdrawn	0.42	0.35	0.33	0.29	0.41	0.33	0.33	0.32	0.18	10.18**
7. Somatic Complaints	0.58	0.36	0.37	0.27	0.56	0.35	0.30	0.28	12.07**	85.06**

* $p < 0.05$; ** $p < 0.01$.

extent of boys' and girls' consistency in the absolute level of externalizing and internalizing problem behavior over time by comparing adolescents who scored in the clinical, subclinical, and normal range of problem behavior.

Following the norms of Achenbach (1990) *T*-values were used as cutoff scores that mark the 3 different problem areas 'clinical,' 'sub-clinical,' and 'normal.' Adolescents with *T*-values of 64 and higher were considered to be high on problem behavior, adolescents with *T*-values of 59 and lower as functioning well, and adolescents with *T*-values in between were on the border of clinical and non-clinical scores (=subclinical). Using these cutoff scores makes it possible to detect clinically relevant changes in adolescents' functioning across time. Because there are no norm scores for the 2 newly developed scales (School Problems and Disobedience), only the 2 scales of the YSR (Delinquent Behavior and Aggressive Behavior) were analyzed.

Figure 2 shows the pathways of the boys and girls who scored in the high or normal range of problem behavior at Time 1 and their changes in position at Time 2. Of the 24 boys who scored high on externalizing problem behavior at Time 1 (panel a), 37% remained scoring high, 42% improved considerably and scored in the normal range, and 21% scored in the subclinical range at Time 2. The girls, on the other hand, showed a less positive picture (panel b): of the 26 girls who scored in the high range, 58% remained in that range, 23% improved to the normal range, and 19% scored in the subclinical range. These differences between boys and girls however, were not significant ($\chi^2(2) = 2.42, p > 0.05$).

Furthermore, 84% of the 194 girls who scored in the normal range of externalizing problems at Time 1 remained in that range at Time 2 but 10% of the girls increased considerably in problem behavior (got more problematic) and scored in the high range at Time 2, as compared to 91% and 3% of the boys, who remained in the normal range and increased, respectively. Here, boys and girls did significantly differ in stability ($\chi^2(2) = 20.76, p < 0.000$), with boys being more stable.

With regard to internalizing problem behavior (panels c and d) it can be seen that more girls than boys scored high on problem behavior, as was expected. Boys showed more stability in the high range than girls (59% versus 43%), although differences in stability were not significant ($\chi^2(2) = 4.87, p > 0.05$). Boys showed also more stability than girls in the normal range (93% versus 82%). There were more girls than boys who showed an increase in problem behavior: 10% versus 5% scored from the normal range at Time 1 in the high range at Time 2. These distributions appeared to be significantly differ-

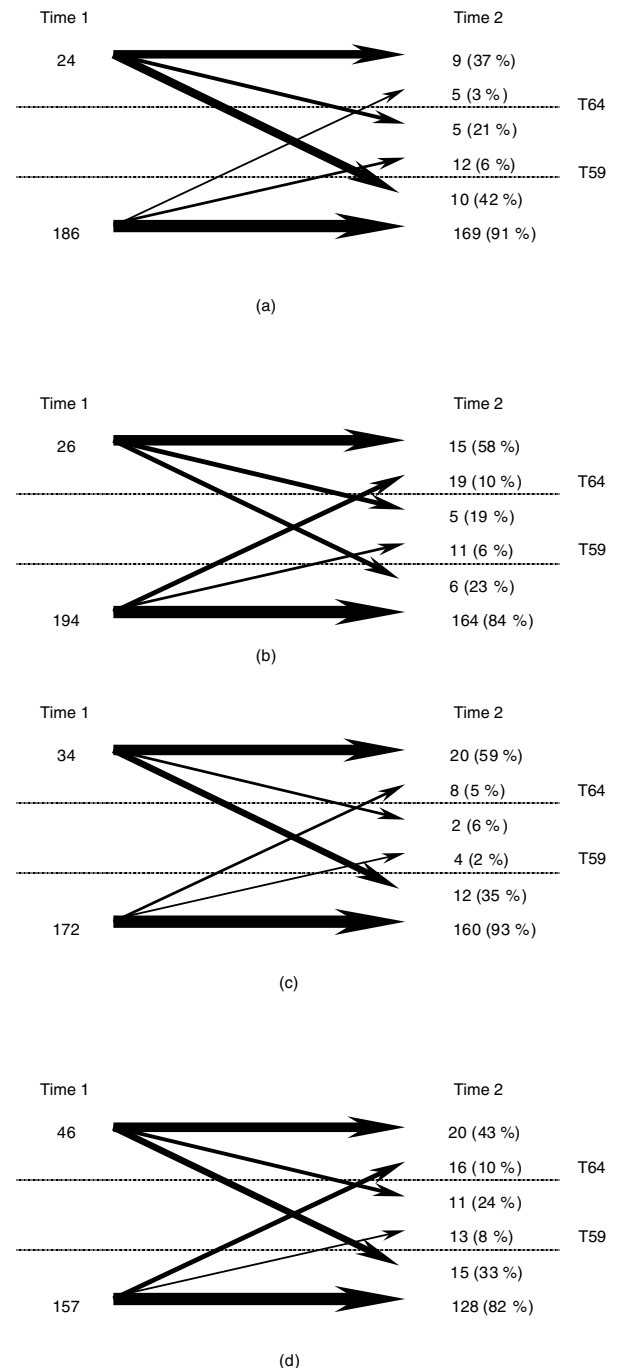


Fig. 2. Developmental pathways of adolescents who scored above and below the cutoff scores for deviant and normal behavior at Time 1, separately for boys and girls and for externalizing and internalizing problem scores. Dotted lines indicate the T64 border (deviant range) and the T59 border (normal range). Percentages are in parentheses. (a) Stability of externalizing problem behavior of boys; (b) Stability of externalizing problem behavior of girls; (c) Stability of internalizing problem behavior of boys; and (d) Stability of internalizing problem behavior of girls.

ent ($\chi^2(2) = 10.33, p < 0.01$) with girls showing more changes than boys.

DISCUSSION

Previous research has focused mainly on the structure of a limited range of problem behavior, namely externalizing problems, and mostly found support for a 1-factor structure. The present study extended this research by studying not only externalizing but also internalizing problems. We examined whether both types of behaviors belonged to 1 single factor of general problem behavior (1-factor model), whether they should be considered as 2 separate constructs (2-factor model), or whether they can be conceptualized as 2 separate constructs belonging to 1 factor of general problem behavior (hierarchical structure model). The third model fitted the data best on both waves, indicating that externalizing and internalizing problem behavior are, in part, 2 unique constructs and, in part, manifestations of a more general tendency of problem behavior. This supports the existence of a syndrome of problem behavior (Jessor and Jessor, 1977) as a meaningful second-order (but not first-order) factor during early adolescence.

This factor of general problem behavior implies the existence of comorbidity between externalizing and internalizing problem behavior. Previous studies have shown that clinically referred children and adolescents show high comorbidity rates between different types of problem behavior (e.g., Meller and Borchardt, 1996). Our data show that there is also considerable comorbidity between externalizing and internalizing problem behavior in a community sample of adolescents. While this seems to be in consistence with findings from other studies that examined comorbidity in community samples (e.g., Verhulst and van der Ende, 1992), our study is the first that actually tested different types of structure *models* of externalizing and internalizing problem behavior, including a hierarchical structure model. Statistically testing such a model offers a more comprehensive representation of the composition of different types of problem behaviors.

Another new concern of the present study was the examination of the stability of the structure of problem behavior over time, as an extension of previous (cross-sectional) research on the structure of problem behavior. Our finding that the hierarchical structure model is stable over time suggests that comorbidity of externalizing and internalizing problem behavior remains an important feature during the period of adolescence. Future research is needed to verify our findings on other samples and for different developmental periods within a child's life.

One might argue that it is not surprising to find a hierarchical structure model because it resembles the structure of the questionnaire that is used to measure problem behavior: the YSR. This questionnaire is composed of different small-band syndromes that belong to one of 2 broad-band syndromes Externalizing or Internalizing. These 2 syndromes can be summed up to provide a Total problem behavior score, which implies the existence of a higher-order factor (general problem behavior). The following reasons can be listed to refute this argument.

The YSR is originally developed for the clinical practice. The structure of the items is examined in a sample of clinically referred adolescents. Principle components analyses were conducted to establish the small-band and broad-band syndromes. Although norms have been provided for a normal population (e.g., Verhulst *et al.*, 1996), the structure of problem behavior has not been statistically tested in this group and might be dissimilar from the clinical group. According to Berkson (1946) and Zoccolillo (1992) this may result in different factor structures for the 2 groups. Furthermore, the possibility of a higher-order factor (general problem behavior) has never been statistically analyzed, both for clinical and normal groups. A 2-factor model could thus have been an equally satisfying solution. A previous study of Hartman *et al.* (1999) found such a model when analyzing the Child Behavior Checklist (CBCL), the version that assesses problem behavior of children as reported by parents. Finally, in the present study not only items from the YSR were used to examine problem behavior but also 2 new scales were added to study problems that occur frequently in a community sample of adolescents: School Problems and Disobedience. These 2 new scales contributed to the hierarchical structure model.

The present study not only focused on the stability of the structure of problem behavior but also examined 2 other types of stability: relative and absolute stability. The relative stability of the different problem behaviors ranged from 0.45 to 0.61. This is in accordance with other research about the stability of problem behavior (Verhulst and Van Wattum, 1993). However, coefficients seem to be somewhat higher than those of Ferdinand *et al.* (1995) (ranging from 0.31 to 0.47), which can be explained by the fact that the stability of their study is over a 4-year period instead of a 1-year period as is conducted in our study. Stability tends to be lower over a longer time period. Because of the turbulent time-period for early adolescents (school transition, pubertal development, etc.) we expected to find low levels of stability for this group of early adolescents. Findings from a study of Osgood *et al.* (1988) with data from high school seniors and thus from an older age group than in the present study (18–22 years old), show stability

coefficients over a 3-year period ranging from 0.53 to 0.93 (!). This seems to indicate that stability becomes higher as adolescents become young adults.

It was further found that delinquency, disobedience and school problems increased over time. Surprisingly, however, aggressive behavior decreased over time. Additional analyses (*t*-tests) showed that mostly minor aggressive behaviors (5 items out of 19) accounted for this decrease, including for example 'I talk too much,' 'I tease others a lot,' and 'I brag about and show off'. More aggressive behaviors like fighting and threatening others do not seem to decrease and even seem to show slight increases, which is in consistence with findings from other studies that show aggression to remain stable or increase over time from early to middle adolescence (Loeber and Stouthamer-Loeber, 1998; Moffitt, 1993; Patterson *et al.*, 1992).

Regarding internalizing problem behavior, somatic complaints seems to decrease over time but only for boys. This is consistent with the norms for a normal population of the YSR where boys show decreasing levels of somatic complaints over time, whereas girls remain stable. As could be expected from previous studies, girls scored higher on anxiety/depression than boys. They remained fairly stable over time, which is also in consistence with the norms for a normal population (not the clinical population) of the YSR.

With regard to the absolute stability, changes within clinical, subclinical, and normal groups were assessed over a 1-year period. It appeared that many adolescents remained in the high range of problem behavior from Time 1 to Time 2, percentages ranging from 37 to 59%. Fortunately, many adolescents also remained in the non-clinical range (82–93%). When using average percentages of the total group of adolescents and comparing these with findings from the 2 other studies that examined the absolute stability (Ferdinand *et al.*, 1995; Verhulst and Van Wattum, 1993), it appeared that our group seemed to be somewhat more stable not only in high levels of problem behavior (49% against 36% and 42%) but also in low levels of problem behavior (87.5% against 73% and 74%).

Regarding absolute stability, the results indicated that girls showed more increases in problem behavior than boys (non-deviant to deviant range), both for externalizing and internalizing problems. Because girls mature earlier than boys, they may experience more stress related to the multiple changes occurring during early adolescence (pubertal development, school transition) (Leadbeater *et al.*, 1999; Petersen *et al.*, 1991). It is found that a stressful developmental period makes girls more likely to become depressed (Hops *et al.*, 1989; Rutter, 1986). Further, there

might be the possibility of a gender paradox: The gender with the lower prevalence of a disorder actually is at higher risk of poor outcomes (Tiet *et al.*, 2001).

Several limitations of the present study are worth noticing. First, findings are based on self-report data of adolescents. It is possible that different patterns of findings would be obtained if other informants were used. On the other hand, because many problems that adolescents experience remain unnoticed by their parents or teachers (Verhulst and van der Ende, 1992; Youngstrom *et al.*, 2000), it is important to have their own report on their behaviors, which might give more reliable findings.

A second limitation considers the attrition group. This group scored higher on Delinquent Behavior than the group of adolescents on which the analyses are based. This might have influenced the results. Regardless of this fact, we still found quite a high percentage of boys and girls who scored in the deviant range of problem behavior.

Finally, our sample is quite homogenous, consisting of mostly Dutch adolescents. It is therefore not possible to generalize the results to other ethnic groups. Replication of these findings is needed in ethnically more diverse samples.

Notwithstanding these limitations, this study is one of the first that systematically tried to answer the question about the structure of both externalizing and internalizing problem behavior, including effects of both gender and time. It is obvious from the literature that both types of problems are interrelated and therefore research should include externalizing as well as internalizing problems in trying to answer the questions about the structure, especially regarding the period of adolescence where both types of problems are quite prevalent. Also, research should focus on the determinants for changes from non-clinical to clinical behavior and the other way around, to improve our understanding and ability to predict change in problem behavior of boys as well as girls.

REFERENCES

- Achenbach, T. M. (1990). *The Young Adult Self-Report*. Department of Psychiatry, University of Vermont, Burlington.
- Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Department of Psychiatry, University of Vermont, Burlington.
- Achenbach, T. M. (1991b). *Manual for the Youth Self-Report and 1991 Profile*. Department of Psychiatry, University of Vermont, Burlington.
- Alder, A. G., and Scher, S. J. (1994). Using growth curve analyses to assess personality change and stability in adulthood. In Heather-ton, T. F. and Weinberger, J. L. (eds.), *Can Personality Change?* American Psychological Association, Washington, DC.
- Allen, J. P., Leadbeater, B. J., and Aber, J. L. (1994). The development of problem behavior syndromes in at-risk adolescents. *Dev. Psychopathol.* 6: 323–342.

- Allgood-Merten, B., Lewinsohn, P. M., and Hops, H. (1990). Sex differences and adolescent depression. *J. Abnormal Psychol.* 99(1): 55.
- Angold, A., and Costello, E. J. (1993). Depressive comorbidity in children and adolescents: empirical, theoretical, and methodological issues. *Am. J. Psychiatry* 150: 1779–1791.
- Ary, D. V., Duncan, T. E., Biglan, A., Metzler, C. W., Noell, J. W., and Smolkowski, K. (1999a). Development of adolescent problem behavior. *J. Abnormal Child Psychol.* 27(2): 141–150.
- Ary, D. V., Duncan, T. E., Duncan, S. C., and Hops, H. (1999b). Adolescent problem behavior: The influence of parents and peers. *Behav. Res. Therapy* 37: 217–230.
- Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics* 2: 47–53.
- Brack, C. J., Brack, G., and Orr, D. P. (1994). Dimensions underlying problem behaviors, emotions, and related psychosocial factors in early and middle adolescents. *J. Early Adolesc.* 14(3): 345–370.
- Deković, M. (1999). Risk and protective factors in the development of problem behavior during adolescence. *J. Youth Adolesc.* 28(6): 667–685.
- Donovan, J. E., and Jessor, R. (1985). Structure of problem behavior in adolescence and young adulthood. *J. Consult. Clin. Psychol.* 53(6): 890–904.
- Donovan, J. E., Jessor, R., and Costa, F. M. (1988). Syndrome of problem behavior in adolescence: A replication. *J. Consult. Clin. Psychol.* 56(5): 762–765.
- Farrell, A. D., Danish, S. J., and Howard, C. W. (1992). Relationship between drug use and other problem behaviors in urban adolescents. *J. Consult. Clin. Psychol.* 60(5): 705–712.
- Farrell, A. D., Kung, E. M., White, K. S., and Valois, R. F. (2000). The structure of self-reported aggression, drug use, and delinquent behaviors during early adolescence. *J. Clin. Child Psychol.* 29(2): 282–292.
- Ferdinand, R. F., Verhulst, F. C., and Wiznitzer, M. (1995). Continuity and change of self-reported problem behaviors from adolescence into young adulthood. *J. Am. Acad. Child Adolesc. Psychiatry* 34(5): 680–690.
- Flannery, D. J., Williams, L. L., and Vazsonyi, A. T. (1999). Who are they with and what are they doing? Delinquent behavior, substance use, and early adolescents' after-school time. *Am. J. Orthopsychiatry* 69(2): 247–253.
- Fletcher, A. C., Steinberg, L., and Sellers, E. B. (1999). Adolescents' well-being as a function of perceived interparental consistency. *J. Marriage Family* 61: 599–610.
- Gillmore, M. R., Hawkins, J. D., Catalano, R. F., Day, L. E., and Moore, M. (1991). Structure of problem behaviors in preadolescence. *J. Consult. Clin. Psychol.* 59(4): 499–506.
- Hartman, C. A., Hox, J., Erol, N., Mellenbergh, G. J., Oosterlaan, J., Shalev, R. S., Auerbach, J., Fonseca, A. C., Nøvik, T. S., Roussos, A. C., and Zilber, N. (1999). Syndrome dimensions of the Child Behavior Checklist and the Teacher Report Form: A critical empirical evaluation. *J. Child Psychol. Psychiatry* 40(7): 1095–1116.
- Holsen, I., Kraft, P., and Vitterso, J. (2000). Stability in depressed mood in adolescence: Results from a 6-year longitudinal panel study. *J. Youth and Adolesc.* 29(1): 61–78.
- Hops, H., Sherman, L., and Biglan, A. (1989). Maternal depression, marital discord, and children's behavior: A developmental perspective. In Patterson, G. R. (ed.), *Depression and Aggression in Family Interactions*. Erlbaum, Hillsdale, NJ, pp. 185–208.
- Ingersoll, G. M., and Orr, D. P. (1989). Behavioral and emotional risk in early adolescents. *J. Early Adolesc.* 9(4): 396–408.
- Jessor, R., and Jessor, S. L. (1977). *Problem Behavior and Psychosocial Development: A Longitudinal Study of Youth*. Academic Press, New York.
- Jöreskog, K. G., and Sörbom, D. (1993). *LISREL 8: User's guide*. Scientific Software, Chicago.
- Leadbeater, B. J., Kuperminc, G. P., Blatt, S. J., and Hertzog, C. (1999). A multivariate model of gender differences in adolescents' internalizing and externalizing problems. *Dev. Psychol.* 35(5): 1268–1282.
- Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., and Van Kammen, W. B. (1998). *Antisocial Behavior and Mental Health Problems: Explanatory Factors in Childhood and Adolescence*. Lawrence Erlbaum, Mahwah, NJ.
- Loeber, R., Green, S. M., Lahey, B. B., Frick, P. J., and McBurnett, K. (2000). Findings on disruptive behavior disorders from the first decade of the Developmental Trends Study. *Clin. Child Family Psychol. Rev.* 3(1): 37–60.
- Loeber, R., and Keenan, K. (1994). Interaction between conduct disorder and its comorbid conditions: Effects of age and gender. *Clin. Psychol. Rev.* 14: 497–523.
- Loeber, R., and Stouthamer-Loeber, M. (1998). Development of juvenile aggression and violence: Some common misconceptions and controversies. *Am. Psychol.* 53(2): 242–259.
- Maggs, J. L., Almeida, D. M., and Galambos, N. L. (1995). Risky business: The paradoxical meaning of problem behavior for young adolescents. *J. Early Adolesc.* 15(3): 344–362.
- McConaughy, S. H., and Skiba, R. J. (1993). Comorbidity of externalizing and internalizing problems. *School Psychol. Rev.* 22(3): 421–438.
- McCord, J. (1990). Problem behaviors. In Feldman, S. S. and Elliott, G. R. (eds.), *At the Threshold. The Developing Adolescent*. Harvard University Press, Cambridge, MA, pp. 414–430.
- McGee, L., and Newcomb, M. D. (1992). General deviance syndrome: Expanded hierarchical evaluations at four ages from early adolescence to adulthood. *J. Consult. Clin. Psychol.* 60(5): 766–776.
- Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychol. Rev.* 100(4): 674–701.
- Moffitt, T. E., Caspi, A., Rutter, M., and Silva, P. A. (2001). *Sex differences in antisocial behaviour. Conduct disorder, delinquency, and violence in the Dunedin longitudinal study*. Cambridge University Press, Cambridge.
- Osgood, D. W., Johnston, L. D., O'Malley, P. M., and Bachman, J. G. (1988). The generality of deviance in late adolescence and early adulthood. *Am. Sociol. Rev.* 53: 81–93.
- Overbeek, G., Vollebergh, W., Meeus, W., Engels, R., and Luijckers, E. (2001). Course, co-occurrence, and longitudinal associations of emotional disturbance and delinquency from adolescence to young adulthood: A six-year three-wave study. *J. Youth Adolesc.* 30(4): 401–426.
- Patterson, G. R., Reid, J., and Dishion, T. J. (1992). *Antisocial boys*. Castalia, Eugene, OR.
- Peeters, J. (1994). Klachtenlijst voor adolescenten: KLAD. [Complaint-list for adolescents] *Gedragstherapie* 27(2): 109–125.
- Petersen, A. C., Sarigiani, P. A., and Kennedy, R. E. (1991). Adolescent depression: Why more girls? *J. Youth Adolesc.* 20: 247–272.
- Rothbaum, F., and Weisz, J. R. (1994). Parental caregiving and child externalizing behavior in nonclinical samples: A meta-analysis. *Psychol. Bull.* 116(1): 55–74.
- Rutter, M. (1986). The developmental psychopathology of depression: Issues and perspectives. In Rutter, M., Izard, C., and Read, P. (eds.), *Depression in Young People: Developmental and Clinical Perspectives*. Guilford, New York, pp. 3–30.
- Siegel, A. W., and Scovill, L. C. (2000). Problem behavior: The double symptom of adolescence. *Dev. Psychopathol.* 12: 763–793.
- Tiet, Q. Q., Wasserman, G. A., Loeber, R., McReynolds, L. S., and Miller, L. S. (2001). Developmental and sex differences in types of conduct problems. *J. Child Family Stud.* 10(2): 181–197.
- Verhulst, F. C., and van der Ende, J. (1992). Agreement between parents' and adolescents' self-reports of problem behavior. *J. Child Psychol. Psychiatry* 33(6): 1011–1023.
- Verhulst, F. C., van der Ende, J., and Koot, H. M. (1996). *Handleiding voor de CBCL/4-18* [Manual for the CBCL: Dutch

- version]. Afdeling Kinder- en jeugdpsychiatrie, Academisch Ziekenhuis/Erasmus Universiteit, Rotterdam.
- Verhulst, F. C., van der Ende, J., and Koot, H. M. (1997). Handleiding voor de Youth Self-Report (YSR) [Manual for the Youth Self-Report (YSR): Dutch version]. Afdeling Kinder- en jeugdpsychiatrie, Academisch Ziekenhuis/Erasmus Universiteit, Rotterdam.
- Verhulst, F. C., and van Wattum, P. J. (1993). Two-year stability of self-reported problems in an epidemiological sample of adolescents. *Acta Psychiatrica Scand.* 87: 322.
- Willet, J. B., and Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychol. Bull.* 116: 363–381.
- Wolins, L. (1995). A Monte Carlo study of constrained factor analysis using maximum likelihood and unweighted least squares. *Educ. Psychol. Measur.* 55(4): 545–557.
- Youngstrom, E., Loeber, R., and Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *J. Consult. Clin. Psychol.* 68(6): 1038–1050.
- Zeitlin, H. (1999). Psychiatric comorbidity with substance misuse in children and teenagers. *Drug Alcohol Dependence* 55: 225–234.
- Zoccolillo, M. (1992). Co-occurrence of conduct disorder and its adult outcomes with depressive and anxiety disorders: A review. *J. Am. Acad. Child Adolesc. Psychiatry* 31: 547–556.