



Published in final edited form as:

*Psychol Assess.* 2010 December ; 22(4): 893–911. doi:10.1037/a0020703.

## Cross-Informant Symptoms from CBCL, TRF, and YSR: Trait and Method Variance in a Normative Sample of Russian Youths

**Elena L. Grigorenko,**

Yale University, New Haven, CT

**Christian Geiser,**

Arizona State University, Phoenix, AZ

**Helena R. Slobodskaya,** and

Research Institute of Physiology, Russian Academy of Medical Sciences, Siberian Division, Novosibirsk, Russian Federation

**David J. Francis**

University of Houston, Houston, TX

### Abstract

A large community-based sample of Russian youths ( $n = 847$ , mean age = 13.17,  $sd = 2.51$ ) was assessed with the *Child Behavior Checklist* (mothers and fathers separately), *Teacher's Report Form*, and *Youth Self-Report*. The multiple indicator-version of the Correlated Trait-Correlated (Method Minus One) [CT-C(M-1)] model was applied to analyze (1) the convergent and divergent validity of these instruments in Russia, (2) the degree of trait-specificity of rater biases, and (3) potential predictors of rater-specific effects. As expected, based on the published results from different countries and in different languages, the convergent validity of the instruments was rather high between mother and father reports, but rather low for parent, teacher, and self reports. For self- and teacher reports, rater-specific effects were related to age and gender of the children for some traits. These results, once again, attest to the importance of incorporating information from multiple observers when psychopathological traits are evaluated in children and adolescents.

### Keywords

Cross-Informant Symptoms; multi-trait-multi-method; CBCL; TRF; YSR; Russian children and adolescents

---

The use of multiple informants for the purpose of assessing child and adolescent psychopathology is viewed as a best practice and is recommended as a source of both convergent and divergent but helpful information (Jensen et al., 1999; Lahey et al., 1996). Such assessments typically involve data obtained with parallel self-report and collateral-report forms, starting from middle childhood. It is widely accepted in this field that aggregate and comparative data from multiple informants are essential for accurate and effective assessment and treatment (Achenbach, 2006).

---

The evidence underlying this belief is abundant. For example, it was found in a large-scale meta-analysis that the mean correlations between different types of informants reporting on child and adolescent psychopathology (e.g., parents, teachers, mental health workers, observers, peers, and the youth themselves), although statistically significant, varied dramatically pairwise. Specifically, the mean correlations were .60 between similar informants (e.g., pairs of parents), .28 between different types of informants (e.g., parent/teacher), and .22 between youth and other informants (Achenbach, McConaughy, & Howell, 1987). Subsequent meta-analyses of cross-informant correlations for ratings of psychopathology in children and adolescents have yielded similar findings (Duhig, Renk, Epstein, & Phares, 2000; Renk & Phares, 2004). This pattern of low to moderate cross-informant correlations with regard to reports on child and adolescent psychopathology is cited as one of the most robust findings in psychopathology (De Los Reyes & Kazdin, 2005). Similarly, when adult data are considered, the mean cross-informant correlations are estimated at .43 for internalizing and .44 for externalizing problems, if parallel instruments are used, and at .30 if different, non-parallel instruments are used (Achenbach, Krukowski, Dumenci, & Ivanova, 2005).

An accumulation of studies that involve data from multiple informants and careful investigations into the patterns of results depending on the specifics of the samples of participants recruited have revealed a number of interesting specifications to the patterns of agreement and disagreement between different informants on different aspects of psychopathology. For example, there is variability in the patterns of correlations between informants depending on whether a clinical or a normative sample of children is evaluated (Stanger, MacDonald, McConaughy, & Achenbach, 1996). Specifically, it has been observed rather consistently and in a number of countries that in community samples children and adolescents tend to report more symptoms about themselves than their parents report about them [e.g., (Begovac, Rudan, Skocic, Filipovic, & Szivovicza, 2004; Goodman, Meltzer, & Bailey, 1998; Stanger & Lewis, 1993; van den Ende & Verhulst, 2005)], whereas in clinical samples parents tend to report more symptoms about their offspring than the children and adolescents report about themselves (Becker, Hagenberg, Roessner, Woerner, & Rothenberger, 2004; Berg-Nielsen, Vika, & Dahl, 2003; Goodman, et al., 1998; Martin, Ford, Dyer-Friedman, Tang, & Huffman, 2004; Sawyer, Baghurst, & Mathias, 1992).

In general, complaints about internalizing problems are more often endorsed by youngsters than their parents (Carlson, Kashani, Thomas, Vaidya, & Daniel, 1987; Herjanic, 1982; Reich, Herjanic, Welner, & Gandhi, 1982; Zahn-Waxler, Klimes-Dougan, & Slattery, 2000). Similarly, youths tend to report more internalizing problems about themselves than their teachers do, however, teachers often report more externalizing problems than youths do, especially for particular minority groups (Fabrega, Ulrich, & Loeber, 1996; Youngstrom, Loeber, & Stouthamer-Loeber, 2000). And, whereas parents and teachers tend to be in higher agreement with each other than with the youths (Loeber, Green, Lahey, & Stouthamer-Loeber, 1989), teachers tend to report fewer problems than parents (Zimmerman, Khoury, Vega, Gil, & Warheit, 1995). These differential profiles of reporting and their clinical implications have been extensively discussed in the literature [e.g., (Smith, 2007)]. Agreements between mothers and fathers are not studied as often as those between parents and teachers, but, when examined, are found to be relatively high, higher than those

between parents and teachers (McAuley & Trew, 2000; Seiffge-Krenke & Kollmar, 1998), although it depends on the mental health status of one or both parents (Berg-Nielsen, et al., 2003; Wals et al., 2001). Of interest is that maternal ratings tend to correlate higher with self-reports than paternal ratings (Seiffge-Krenke & Kollmar, 1998).

Yet, these general considerations are not reportedly true for all types of youth. For example, in some special populations, such as immigrant children, teachers were observed to report higher rates of internalizing problems than either parents or children did (Stevens, Vollebergh, Pels, & Crijnen, 2005). Immigrant families have been found to present particularly low agreements on mental and behavioral problems in offspring (Montgomery, 2008). For instance, the rates of agreement between refugee parents and their children were particularly low on externalizing problems (Rousseau & Drapeau, 1998; Wahlsten, Ahmad, & Von Knorring, 2002). In addition, differential patterns of agreement were reported for different minority groups (e.g., Asian/Pacific Islander and Hispanic vs. Caucasians and African Americans) for externalizing, but not internalizing behavior (Lau et al., 2004).

Thus, the field is in agreement on the importance of using multiple informants in addressing issues of psychopathology in youths to avoid various biases in reports of mental and behavior problems, but is also aware that these biases might vary asystematically depending on the nature of the assessed sample. Still, having virtually unanimously established the need to utilize multiple informants for assessing psychopathology in children and adolescents, the field is quite divergent in its approaches to how the data from various informants should be treated. Given the diversity of data obtained from different informants, the field offers a discussion of the validity of different instruments based on the notion that meaningful trait-related variance should be separated from measurement error, so that the measures reflect only what they are expected to assess (Messick, 1995). Burns and Haynes (Burns & Haynes, 2006) discussed a number of possible sources of variance in clinical psychology that can bias trait-related variance and jeopardize the validity of clinical assessments. In the literature, these sources of variance are typically referred to as “method variance,” although this concept has been criticized for its lack of precision (Sechrest, Davis, Stickle, & McKnight, 2000) and context-dependency (Courvoisier, Nussbeck, Eid, Geiser, & Cole, 2008).

Situations in which a single “target” (e.g., child) is evaluated by multiple raters on multiple traits (so-called *multitrait-multimethod* or *MTMM* design) have been of substantial interest (Funder & West, 1993). In multi-informant contexts, other informants (e.g., parents and teachers) often are not randomly selected out of a pool of uniform (i.e., interchangeable) raters, but are “fixed” for each target (i.e., for each assessed child or youth). In such a design, each type of rater has a unique perspective and access to only partially overlapping information about the target (e.g., parents have limited information concerning their child’s behavior at school, whereas teachers tend to not have access to a child’s behavior at home). Consequently, this is typically referred to as a situation with *structurally different raters* as opposed to situations using interchangeable raters [i.e., (i.e., randomly selected children from the class or peer group to which a target belongs (Eid et al., 2008)]. When data are collected from structurally different raters, specific modeling approaches should be utilized to compare and contrast the data from different informants and to appropriately separate

measurement error from true trait and method variance (Eid, Lischetzke, & Nussbeck, 2006; Eid, et al., 2008).

Confirmatory factor analysis (CFA) is nowadays the most widely used approach to analyzing MTMM data (Eid et al., 2006). The CFA approach is flexible as it provides models for different types of MTMM measurement designs (i.e., for studies employing different types of raters, see Eid et al., 2008). Further, CFA-MTMM models allow for (1) a separation of measurement error from true convergent validity and true method specificity; (2) the analysis of variance components (trait, method, error); (3) the analysis of trait-specific method effects; and (4) the inclusion of covariates (e.g., person characteristics) to explain method effects.

In 1992, Fiske and Campbell (Fiske & Campbell, 1992) noted that although their influential 1959 *Psychological Bulletin* article on the MTMM matrix (Campbell & Fiske, 1959) was among the most highly cited papers in psychology, and although sophisticated methods (like CFA) were now available and widely applied, there was still considerable lack of progress in psychology on the question of why method effects occur at all and how they can be explained. That is, although it is a common finding that the convergent validity of different sources is rather low (while the method-specificity is high) for many psychological traits, only few studies have attempted to explain why this is so. In the present study, we not only investigated convergent and discriminant validity, but also explored the causes of rater-specific effects by relating these effects to child characteristics.

To do so, here a community sample of Russian youths who were each assessed by four structurally different informants—themselves and their mothers, fathers, and teachers—is considered. This sample is selected for a variety of reason. First, it is highly important to strengthen and saturate our understanding of psychopathology in children and youth by sampling from countries other than developed Western countries. An overwhelming amount of the relevant theories and data were generated in the developed world and primarily in Western countries. Given that the majority of children and adolescents (<http://www.unicef.org/sowc05/english/index.html>) live in the developing rather than in the developed world, and in other than Western parts of the world, diversifying the sources of data is crucially important<sup>1</sup>. Second, overall, Russian children appear to report more problems than children from other countries (Carter, Grigorenko, & Pauls, 1995; Gartstein, Slobodskaya, & Kinsht, 2003; Knyazev, Slobodskaya, Safronova, & Kinsht, 2002; Knyazev, Zupancic, & Slobodskaya, 2008; Kuznetsova, Grigorenko, & Voronkova, 1996; Slobodskaya, 1999). Moreover, it appears that parents and teachers of Russian children also report higher levels of problems (Hellinckx, Grietens, & De Munter, 2000; Rescorla et al., 2007). For example, Goodman and colleagues (Goodman, Slobodskaya, & Knyazev, 2005) showed that the problem levels, as assessed by parents and teachers, were 1.5–3 times higher in Russia than in Britain. These findings, although quite replicable as shown by the literature, are still not understood. There are multiple hypotheses that might be relevant here.

---

<sup>1</sup>A relevant illustration here comes from studies of reading. While the overwhelming majority of theories and data on the acquisition of reading has been generated through studies of English, it is now clear that many theories and empirical findings do not generalize, completely or partially, to other languages (Share, 2008).

These hypotheses include lack of experience, in Russian children and youths, in answering quite personal questions, thus leading to uncertainty as to how to gauge their responses to such questions that address typical and atypical behaviors. There is also a suggestion that these elevations can be explained by a lack of community-based mental-health services; correspondingly, most often only major psychopathological conditions get diagnosed and treated, whereas relatively “minor” symptoms remain undetected and unaddressed. Moreover, there are also hypotheses suggesting that the Russian cultural environment in general, and the environment in Russian family and classroom settings in particular, appear to be more adverse and, correspondingly, might overemphasize the negative and underemphasize the positive aspects of child and adolescent development. All in all, no systematic epidemiological studies of mental health issues among children and adolescents have been conducted in Russia and, at this point, these suggestions remain hypotheses and no more. Regardless of the causality of these elevations in psychopathological behaviors, these elevations are of interest because all of these hypotheses (lack of experience in reporting on psychopathology, lack of psychopathology-related services, and specific cultural peculiarities), separately and together, may be related to other understudied populations, such as children and adolescents in other developing countries. Thus, an investigation of this and other samples from the developing world will contribute to the field’s understanding of how generalizable (or not) the theories and data obtained from children in Western countries are to the majority of the world’s children.

Thus, the objectives of these analyses are to investigate the specific biases that characterize the data obtained from various informants and to place this investigation on Russian children within other available international data (whether from the developing or developed countries) on such biases. Regardless of what part of the world the data are coming from, there is a substantial amount of disagreement between multiple informants. Here, we used modern approaches of CFA to separate true convergent validity and method (rater) specificity from random measurement error. These models also allowed us to test whether rater biases were related to characteristics of the child, that is, whether these biases could be predicted by the sex and age of the child. In addition, we wanted to find out to what degree rater biases generalized across the different psychopathological traits or were trait-specific. In other words, we investigated whether and in what way rater biases differed for different internalizing and externalizing psychopathological traits. We applied a specific CFA-MTMM approach in which trait-specific method factors can be analyzed for each construct [so-called *multiple indicator MTMM model* (Eid, Lischetzke, Nussbeck, & Trierweiler, 2003; Eid, et al., 2008; Marsh, 1993; Marsh & Hocevar, 1988)]. In summary, this work is method-oriented as it contributes, using a novel methodology, to the investigation of the problem of rater biases while appraising problematic feelings and behaviors in children and adolescents. To illustrate relative issues, we use a community-based sample of Russian children and adolescents, whose feelings and behaviors have been appraised by multiple informants—the youth themselves, their parents (both mothers and fathers), and their teachers.

## Method

### Participants

Youths ( $n = 841$ ) from two urban centers in Russia (one in Central Russia and one in Siberia<sup>2</sup>) were invited to participate in the study. The sample included 484 (57.6% of the sample) girls and 357 boys, ranging in age from 8 to 17 (mean age = 13.17,  $sd = 2.51$ ). In both cities, a number of representative public schools (grades 1–10) were recruited through local educational agencies. Both cities are large regional centers, with populations of ~1.5 and ~1 mln people. Although the income differentiation in Russia has increased tremendously over the last 15 years, the district differentiation within urban centers is still minimal. As public schools serve the population of children residing in catchment areas defined geographically, typically, public schools educate children from a variety of social-economical backgrounds, who happen to reside in a particular area of the city. This was the case for our sample, where the range of parental education was from grade 8 (the minimal level of compulsory education in Russia) to advanced degrees, with a substantial portion of the sample (49.5% of women and 39.6% of men) possessing a professional degree, obtained either through a vocational school or a college<sup>3</sup>. The majority of the families who participated in the study were complete families (77.6%), that is, included both mother and father<sup>4</sup>. Upon the agreement of a school's administration to participate in the study, flyers describing the study were distributed at parent meetings and were sent home with children; families were offered monetary compensation for participation. The compensation amounted to ~0.05% if the median household income in Russia at the time of the study<sup>5</sup>. The overall response rate was 87%, although it varied from school to school.

### Behavior Assessment

To investigate the biases of four types of structurally different informants (mothers, fathers, teachers, and self-reports) in reporting on child and adolescent psychopathology, the *Child Behavior Checklist* [CBCL (Achenbach, 1991a)], *Teacher's Report Form* [TRF (Achenbach, 1991b)], and *Youth Self-Report* [YSR (Achenbach, 1991c)] were used<sup>6</sup>.

There is a great deal of published literature on these scales and their use, and the majority of these publications have been carefully collated (Bérubé & Achenbach, 2002). The latest versions of the CBCL, TRF, and YSR forms are part of the Achenbach System of Empirically Based Assessment, ASEBA (Achenbach & Rescorla, 2001). In sum, previous research has shown that the CBCL, TRF, and YSR are reliable and valid instruments for assessing psychopathological symptoms in youths, and that this is also true for the Russian

<sup>2</sup>The selection of the cities was determined by the availability of research support and the location of the Russian investigators.

<sup>3</sup>The level of education is the best possible proxy for SES in Russia, because it is still not culturally appropriate to make inquiries about family income.

<sup>4</sup>Both variables, the educational attainment and the family status, were utilized in mean profile analyses. The educational attainment variable (professional degree vs. not) was not associated with any mean differences for either of the instruments (CBCL, TRF, or YSR) for either mothers or fathers. The family status variable (complete vs. incomplete family) generated 2 (out of 32 total comparisons) mean differences, but these differences were not consistent: teachers reported children from incomplete families ( $n = 91$ ) as more aggressive ( $p < .001$ ), and single fathers ( $n = 12$ ) reported their children to be more delinquent, but these differences were not corroborated by other reporters.

<sup>5</sup>To illustrate, in 2007, according to the Census Bureau ([http://www.census.gov/Press-Release/www/releases/archives/income\\_wealth/012528.html](http://www.census.gov/Press-Release/www/releases/archives/income_wealth/012528.html)), the median annual household income adjusted for inflation was \$50,233.00. If the family were to be paid 0.05% of their annual income, it would have amounted to ~\$25.

versions of the scales<sup>7</sup> (Carter, et al., 1995; Knyazev, et al., 2002; Kuznetsova, et al., 1996; Slobodskaya, 1999). Despite this success, the CBCL, TRF, and YSR have been criticized [e.g., (Drotar, Stein, & Perrin, 1995; Macmann, Barnett, & Lopez, 1993)], and the validity of these empirically based syndromes of taxonomy is periodically re-examined [e.g., (Heubeck, 2000)].

## Procedure

For the present study, scores were obtained on the eight empirically based cross-informant syndromes scored using the CBCL (separately for mothers and fathers), TRF, and YSR. All of the parents who completed the CBCL were asked permission to have the youth's teacher complete the TRF, if the youth attended school. Only parents residing in the same place (i.e., having the same physical address), as their children, were included in the study. If the youth had multiple teachers, the parent was asked to select the teacher who knew the youth best<sup>8</sup>. The administration of the CBCL, YSR, and TRF was carried out in individual and small group settings at schools attended by the children and supervised by research assistants. The CBCL and TRF were collected from children of all ages, whereas the YSR was collected only from children 11 years of age or older.

## Data-analytic Strategy

**Multitrait-multimethod modeling**—Model fitting was carried out with Mplus 5.2 (Muthén & Muthén, 2006). As explained above, the present study employed a multi-

<sup>6</sup>The CBCL includes 112 items sampled from emotional and problematic feelings, thoughts, and behaviors in youths. The CBCL items are rated by respondents on a scale of 0–1–2 to indicate “not true” (0), “somewhat or sometimes true” (1), and “very true or often true” (2). The TRF and YSR are similar to the CBCL, but assess problematic feelings, thoughts, and behaviors from, respectively, the teacher's and child's point of view. Some items on the YSR are “I act too young for my age,” “I argue a lot,” and “I have trouble concentrating or paying attention.” The wording of corresponding items on the three forms is adapted to the type of respondent, with the YSR items stated in the first person, the CBCL items worded from parents' perspectives, and the TRF items worded from teachers' perspectives. The YSR and CBCL respondents are asked to base their ratings on the preceding 6 months. The TRF respondents are requested to base their ratings on the preceding 2 months. All three instruments are well studied and have undergone multiple revisions and transformations (Achenbach, 1966, 1978, 1991a, 1991b, 1991c, 1993; Achenbach & Rescorla, 2001). The signature of these instruments is the argument that they are empirically based (Achenbach, 1966), although there is a secondary correspondence with DSM-based classifications (Achenbach & Dumenci, 2001). Early implementations of the instruments classified problem feelings, thoughts, and behaviors into psychopathological factors that were sex- and age-specific [cf. (Achenbach & Edelbrock, 1983)]. Subsequent revisions established factors that are common across sex and age groups (Achenbach, 1991a, 1991b, 1991c) and stressed the importance of the convergence of ratings from at least two different informants (Achenbach, et al., 1987). These revised assessments resulted in eight “cross-informant syndromes” that form the core of the current empirical taxonomy (Achenbach, 1993). The syndromes were designated as *Withdrawn*, *Somatic Complaints*, *Anxious/Depressed*, *Social Problems*, *Thought Problems*, *Attention Problems*, *Delinquent Behavior*, and *Aggressive Behavior*. This taxonomy has been praised for its applicability to a wide range of ages, both sexes, three rater perspectives, and multiple cultures [e.g., (de Groot, Koot, & Verhulst, 1996; DeGroot, Koot, & Verhulst, 1994; Erol, Simsek, Öner, & Munir, 2008; Rescorla, et al., 2007)]. Extensive reliability and validity data for the syndrome scales have been published (Achenbach & Rescorla, 2001).

<sup>7</sup>Since the validation of the Russian version predated the 2001 revision, we used the Russian adaptation of the 1991 version in this research. The adaptation of the 1991 version in Russian was carried out following international recommendations on translation and adaptation of a foreign instrument in a new culture/society; the steps and results of this adaptation have been published in international journals (see above).

<sup>8</sup>For older students, when they have multiple subject teachers (multi-teacher education starts in Russia in grade 4, when children are, on average, 9 years of age), there is one “class teacher” for every group of children. Unlike in the USA, middle- and high-school students in Russia stay together in the same “formation” (so-called *class*, which typically includes ~25 students). Although they are taught by different teachers, they remain with the same peers and the same “class teacher” throughout their schooling years. This class teacher is typically also a subject teacher (e.g., Russian Language and Literature) and she/her will teach this subject in grades 4–11, moving with her/his class throughout their years of education. In the majority of cases, class teachers know students in their classes (and their families) quite well. These class teachers also talk to other subject teachers to monitor the educational progress of the students in their classes. They also run parental meeting and, in general, communicate with their students' families on behalf of the school. In this work, the majority of the selected teachers were class teachers, but in a number of instances, the parents nominated particular subject teachers, if they felt that a particular teacher had a better understanding of their child.

informant measurement design with four types of structurally different raters. Eid et al. (2003; 2008) have presented a special CFA model that is particularly well-suited for analyzing data obtained from multi-informant designs with structurally different raters. This model is known in the literature as the *Correlated Trait-Correlated (Methods Minus 1)* or *CT-C(M-1) model* (Eid, 2000; Eid, et al., 2003; Eid, et al., 2008). The name refers to the model structure, which allows for correlated trait factors as well as  $m - 1$  correlated method factors. The number of method factors is one less than the number of methods considered, given that the model includes no method factors for one of the methods, the so-called *reference method* (see detailed description below).

One of the major advantages of the CT-C(M-1) model is that it contains clearly interpretable method factors that can be related to external variables (e.g., age and gender) to explain method effects. Further, the multiple indicator version of this model applied in the present study allows for the analysis of trait-specific method effects. In this way, method effects can be studied separately for each construct and the degree to which method effects generalize across different constructs—for example, due to Halo effects—can be analyzed.

The main assumption contextualizing the CT-C(M-1) model is that, due to the fact that each observed variable (i.e., each test/subtest, scale/subscale, or item) signifies a trait-method blend (Campbell & Fiske, 1959), a trait cannot be measured independently of the method. Yet, the convergent validity of the methods can be investigated if the methods are contrasted with each other. In the CT-C(M-1) model, such a comparison is achieved by selecting one method as the *reference method*. The reference method is selected based on prior evidence in the literature, theoretical considerations, or the clarity of the results (Geiser, Eid, & Nussbeck, 2008).

In the present study, the mother report was chosen as the reference method for all traits, and the remaining raters served as non-reference methods that were contrasted against the mother report. This decision was made based on (1) our knowledge of the literature, indicating that, although present, maternal biases tend to be smaller than those of other raters (Biederman, Mick, & Faraone, 1998); and (2) the lesser number of missing maternal reports (compared to father and child reports) in our dataset.

Figure 1 shows a path diagram of the multiple indicator CT-C(M-1) model with the indicator-specific trait factors<sup>9</sup> (Eid, et al., 2008) used in the present study. For clarity and in order to save space, Figure 1 illustrates a reduced model for only two traits (*Withdrawn* and *Somatic Complaints*); it is important to note that in the actual analysis carried out in the present study, we estimated larger models with more than two traits.

---

<sup>9</sup>“Multiple indicator” refers to the fact that in this model, each trait-method-unit (TMU) is represented by two rather than only one observed variable (indicator). Most traditional CFA approaches to multi-method data employ only a single indicator per TMU (Eid, et al., 2006). However, single indicator models have been criticized because they imply the unrealistic assumption that method (rater) effects generalize perfectly across all traits (Marsh, 1993; Marsh & Hocevar, 1988). Multiple indicator approaches overcome this serious limitation of single indicator models (Eid, et al., 2003; Eid, et al., 2008). As we wanted to study to what degree rater effects were trait-specific, we used the multiple indicator version as recommended by Eid and colleagues (2003, 2008). Indicator-specific trait factors were used to account for the fact that indicators of the same TMU may capture slightly different facets of a construct and thus may not be perfectly homogeneous (Eid, et al., 2008).



In the CT-C(M-1) model shown in Figure 1, each observed variable loads onto a factor that represents the particular trait as measured by the reference method (therefore labeled “reference factor” in Figure 1). In the picture, then, the reference factors represent “*Withdrawn* measured by mother report” and “*Somatic Complaints* measured by mother report”, respectively. As the model assumes indicator-specific reference factors, there are two reference factors for each construct (one for each indicator). Indicator-specific reference factors account for the fact that each indicator might represent a slightly different facet of each trait (i.e., the true score variables underlying the indicators of the same trait may not be perfectly homogeneous).

Furthermore, each type of rater (father, teacher, and self) is compared against the reference method (mother). For this purpose, each indicator—except those belonging to the reference method (i.e., mother report)—additionally loads onto a rater-specific factor, called a *method factor*. These method factors capture reliable variance that is specific to a particular type of rater and is not shared with the reference method (i.e., reliable variance components not shared with the mother report). Technically, the method factors are residual factors that account for the systematic residual covariation in the non-reference indicators that is not shared with the reference method.

Note that the rater-specific deviation of fathers, teachers, and children from the mother report may not be the same for the constructs *Withdrawn* and *Somatic Complaints*, but may differ across these constructs. Therefore, we modeled method factors as *trait-specific*. That is, there was a separate method factor for each rater per construct in our model. For example, the model in Figure 1 includes two method factors for the father report: One for *Withdrawn* and one for *Somatic Complaints*. Hence, in contrast to models with only a single general method factor per rater, the current model has an advantage in allowing method effects to be trait-specific. Fathers might, for instance, overestimate a child’s withdrawal but underestimate a child’s somatic problems relative to the mother report. In order to analyze rater effects as trait-specific, at least two indicators are needed per trait-method combination (TMC; see Figure 1). We therefore summed the items belonging to the same construct and the same rater up to two scores (instead of a single score), respectively. This procedure resulted in two test halves per TMC.

The necessity of using two indicators per TMC was supported by preliminary analyses. These analyses showed that a single indicator CT-C(M-1) model did not fit the data at all. Finally, the model included an error variable for each indicator that captures unsystematic (measurement error) influences. In order to identify the metric of the latent reference factors, the loadings of the indicators pertaining to the reference method were fixed to 1.00. The metric of the method factors was identified by fixing the loadings of the first indicators on these factors to 1.00, respectively (see Figure 1). Fixing one loading per factor to assign a scale to this factor is standard practice in CFA.

**Variance decomposition**—In the CT-C(M-1) model, the reference factors are uncorrelated with all method factors within the same TMU. Furthermore, error variables are uncorrelated with all latent factors and with all other error variables. As a consequence, the model permits decomposing the variances of the nonreference indicators into variance

components that are shared with the reference method (convergent validity relative to the reference method), rater-specificity, and measurement error influences (unreliability). The proportion of variance shared with the reference method (*consistency coefficient*) is indicated by the squared standardized loadings of the non-reference indicators onto the reference factor. The proportion of variance that is rater-specific (not shared with the reference method; *method-specificity coefficient*) is indicated by the squared standardized loadings of the non-reference indicators onto the method factors (Eid, et al., 2003; Eid, et al., 2008). The proportion of total reliable variance of an indicator is indicated by the sum of the consistency coefficient plus the method specificity coefficient.

When, for example, the variance explained by a method factor for the father report is small in comparison with the variance accounted for by the reference factor, this situation supports the convergent validity of mother and father ratings. In contrast, a large proportion of method-specific variance signals a lack of convergent validity.

**Relation to external variables**—In order to illustrate the use of covariates and to find out which variables might explain interindividual differences in problem behaviors as well as rater-specific deviations from the reference method, the model was extended by including age and sex as predictor variables in the structural model. In particular, we were interested in whether these variables explained variance in the problem behaviors as rated by mothers, as well as the variance due to rater-specific effects. Consequently, we regressed all reference and all method factors on age and sex. This was done by directly including age and sex as exogenous variables in the structural model and estimating all of the regression paths from these variables to the latent factors.

**Goodness-of-fit assessment**—Goodness of fit was assessed using the  $\chi^2$  test, the *Comparative Fit Index* [*CFI*; (Bentler, 1990)], the *Root Mean Square Error of Approximation* [*RMSEA*; (Steiger, 1990)], and the *Standardized Root Mean Square Residual* (*SRMR*). A good fit of a model is indicated by a non-significant  $\chi^2$  value, which is interpreted as an indication that the assumption of an exact fit of the model in the population is not rejected. The *CFI* compares the fit of the target model with the fit of a baseline model, where the baseline model is a null model that assumes zero covariation among the observed variables. To indicate a good fit, the *CFI* should take on a value greater than 0.97, but it should at least be larger than 0.95 (Hu & Bentler, 1998, 1999; Schermelleh-Engel, Moosbrugger, & Müller, 2003). The *RMSEA* coefficient is a measure of approximate correspondence between the model and the data, with values smaller than .05 pointing to an acceptable fit. *SRMR* is a summary measure of the standardized model residuals (observed minus model-estimated correlations). *SRMR* values smaller than .05 indicate that on average the model reproduces the observed correlations well.

**Handling of missing data**—Given that the collected data were from an age- and family-diverse sample, the resulting dataset was incomplete. Specifically, of the 847 participants, 600 were commented on by their mothers, 409 by their fathers, 745 by their teachers, and 502 by themselves. According to Schafer and Graham (2002) as well as Enders (2010), the use of model-based statistical approaches such as full information maximum likelihood estimation (Arbuckle, 1996; Little & Rubin, 2003; Wothke, 2000) is generally

recommended to handle missing data. In FIML, the model parameters are estimated based on all available data points. FIML leads to unbiased estimates if the data are missing at random and has been shown to outperform ad hoc strategies like listwise deletion, pairwise deletion, or mean substitution.

## Results

### Descriptive Statistics

Descriptive statistics for the cross-informant symptoms (*Withdrawn*, *Somatic Complaints*, *Anxious/Depressed*, *Social Problems*, *Thought Problems*, *Attention Problems*, *Delinquent Behavior*, and *Aggressive Behavior*) as they were reported by different raters are presented in Table 1.

A number of observations can be made by examining Table 1; these observations pertain to informant-, symptom-, and gender-differences. First, on average, as expected based on the literature, youths themselves endorse more problem behaviors than any of the evaluating adults. This pattern is characteristic of all symptoms but *Withdrawn*, and is especially pronounced for *Anxious/Depressed*, *Thought Problems*, *Attention Problems*, *Delinquent Problems*, and *Aggressive Behavior*. Second, there are clear differences in ratings for different symptoms, so that certain symptoms appear to be rated with less variability (*Withdrawn* and *Social Problems*) between informants, whereas other symptoms (*Anxious/Depressed*, *Thought Problems*, *Attention Problems*, *Delinquent Problems*) are evaluated with a substantial degree of disagreement between the informants. Finally, there are interesting gender-related differences. There are particular symptoms (*Somatic Complaints* and *Anxious/Depressed*) that are consistently rated, by all informants, as observed more in girls than in boys. However, there are other symptoms (*Attention Problems*, *Delinquent Problems*, and *Aggressive Behavior*) that are consistently endorsed at higher rates in boys by all informants but the youths themselves; girls endorse these behaviors at rates similar to those of the boys.

Table 2 presents correlations between cross-informant scales from the CBCL (mother and father separately), TRF, and YSR.

These patterns of pairwise correlations are also informative. Specifically, the results illustrate the presence of a high level of agreement between the ratings of the parents of the targeted youths (on all symptoms); a substantial level of agreement between the ratings of the mothers and the teachers (on all symptoms) and between the fathers and the teachers (on all symptoms but one, *Thought Problems*); a moderate level of agreement between the ratings of the mothers and their children (on 5 out of 8 symptoms, with the exception of *Withdrawn*, *Social Problems*, and *Aggressive Behavior*, where agreements are at borderline levels of significance, e.g.,  $> .05$ , but  $< .20$ ); a small level of agreement between the ratings of the fathers and their children (for 3 out of 8 symptoms, with agreement on *Somatic Complaints*, *Anxious/Depressed*, and *Attention Problems*); and a minimal level of agreement between the ratings of the teachers and the youths themselves (for 2 out of 8 symptoms, *Somatic Complaints* and *Delinquent Behavior*).

In summary, the analyses of the descriptive statistics and first-order correlations between different symptoms as appraised by different raters for boys and girls indicate a general convergence between parental and teacher ratings and a substantial divergence between the self-reports from the youths and the adult reports. However, one has to take into account that these correlations are not corrected for measurement error, potentially leading to bias in the estimates of convergent and discriminant validity. We applied the CT-C(M-1) model to take measurement error into account, to obtain more accurate estimates of convergent and discriminant validity, and to learn more about the specific view of each type of rater.

### Multitrait-multimethod Analyses

In applying MTMM methodology to data generated by multiple informants, the CT-C(M-1) model presents a good choice when structurally different methods (i.e., different but not mutually replaceable informants) are used to measure the same trait (e.g., cross-informant symptoms) and when there is interest in evaluating the degree of convergent validity of the different methods (Eid, et al., 2008; Geiser, et al., 2008).

In this application, each TMC was represented by 16 indicators (i.e., two indicators per each of the eight cross-informant symptoms) and 4 informants (mother, father, teacher, and youth). Thus, there were 64 observed variables. A model for all 64 indicators would have comprised a total of 972 free model parameters and thus would have had more parameters than the sample size. Therefore, we reduced the model size by specifying two separate models: One model for the internalizing traits (*Withdrawn, Somatic Complaints, Anxious/Depressed, Thought Problems, and Attention Problems*) and one model for the externalizing traits (*Social Problems, Delinquent Behavior, and Aggressive Behavior*). Although in the original conceptualization internalizing problems combine *Withdrawal, Somatic Complaints, and Anxiety/Depression* scales, while externalizing traits combine *Delinquent Behavior and Aggressive Behavior* scales, the inclusion of additional scales into these two broad categories seemed reasonable based on the pattern of manifest correlations and in correspondence to the aims of this work (the intention was to evaluate rater biases, not the validity of second-order scales of internalizing and externalizing problems). The adequacy of this grouping was supported by the obtained indicators of fit for both models.

The estimated models had exactly the same structure as the model presented in Figure 1 (i.e., with trait-specific method factors and indicator-specific reference factors), except that we considered more than two constructs simultaneously. Both, the CT-C(M-1) model for internalizing problems [ $\chi^2(440) = 790.73$ ;  $CFI = .96$ ;  $RMSEA = .03$ ;  $SRMR = .04$ ] and the model for externalizing problems [ $\chi^2(147) = 238.88$ ;  $CFI = .98$ ;  $RMSEA = .03$ ;  $SRMR = .04$ ] showed a good approximate fit to the data.

### Convergent Validity and Method Specificity

Table 3 shows the estimated intercepts, factor loadings and variance components for the observed variables. The estimated intercepts reflect the model-implied means of the observed variables. The unstandardized estimates are in concert with the data presented in Table 1, when the variance related to age- and gender-differences is taken into account. Of interest are the standardized intercepts because they allow a comparison of the most and

least frequently endorsed symptoms by different raters when the variability in the scales is taken into account. Specifically, parents (both mothers and fathers) endorse mostly items on *Attention Problems* and *Aggressive Behavior* scales; teachers on *Anxious/Depressed*; and youths themselves items on *Withdrawn* and *Aggressive Behavior*.

The trait and method factor loadings reveal that, on average, factor loadings are larger for method than for trait latent variables, especially for the TRF (teacher report) and YSR (self-report). This supports the observations obtained from Table 2: Although there is a substantial amount of agreement between the parents, this agreement is only partially shared with teachers, and, it appears, not shared with the youths themselves. These interpretations are further confirmed by a review of the consistency and method specificity coefficients. To repeat, the consistency coefficient captures the proportion of the observed variance of a nonreference indicator (father, teacher, or self) that is shared with the reference method (mother report). The method-specificity coefficient represents the proportion of observed variance that is also reliable, but specific to a particular nonreference method (i.e., not shared with the mother report). The reliability shows the proportion of variance of an observed variable that is not due to measurement error.

Similar to previously discussed results, these indicators show the presence of disagreements between parent and teacher ratings and, especially, between ratings of adults and the youths themselves. The consistency coefficients for the father report scales show us that for most traits, 40–50% of the observed variance of the father report is shared with the mother report. This indicates that convergent validity is relatively high for the father report across most traits. An exception is the trait *Thought Problems*. For this trait, only 20–22% of the father report variance is shared with the mother report.

The teacher report scales show considerably lower convergent validity with respect to the mother report that is below 10% shared variance for most traits. The only trait for which the TRF scales show somewhat higher convergent validity is *Attention Problems* (20–25% shared variance).

The consistency coefficients for the youth are the smallest of all raters (below 10% shared variance for all traits). Some of the self-report indicators did not even show statistically significant factor loadings on the reference factor. Non-significant loadings on the reference factor indicate that there is no convergent validity at all. The self-report indicators on average also show the lowest reliability estimates of all raters. Hence, we can conclude that the self-report shows very low or no convergence at all with the mother report, implying that the self-report measures something almost completely different. Furthermore, the degree of measurement error seems to be highest for the self-report measures.

### Latent Correlations

Table 4 shows the estimated correlations of the reference factors. These correlations indicate the associations of the problem behaviors for the mother report corrected for measurement error. It can be seen that the correlations between indicator-specific factors pertaining to the *same* trait (e.g., between *Withdrawn 1* and *Withdrawn 2*) range between  $r = .52$  and  $.90$ . This simply shows that the test halves pertaining to the same construct are not perfectly

homogenous, as all correlations are considerably smaller than 1.00. This can be explained by the fact that most of the scales consist of relatively heterogeneous items. Of note also is that some between-construct correlations (e.g., correlations between *Thought Problems* and *Anxious/Depression*) are higher than the correlations between different halves of the same scales (e.g., the *Thought Problems* scale). The highest correlations are between the halves of the *Attention Problems* scale ( $r = .83$ ) and the *Aggressive Behavior* scale ( $r = .90$ ), indicating that the items forming these scales appear to be more homogeneous than the items forming the other scales.

Substantively more interesting are the correlations of reference factors pertaining to *different* traits. These types of correlations tell us something about the degree of discriminant validity of the behaviors as measured by mother report. These correlations range between  $r = .07$  and  $.69$  and, thus, indicate a high to moderate level of discriminant validity of the traits. Relatively high correlations (low discriminant validity) are found between the reference factors representing *Withdrawn* and *Anxious/Depressed* as well as between factors representing *Delinquent Behavior* and *Aggressive Behavior*. These high correlations are expected, however, given that these concepts are closely linked theoretically.

An inspection of the correlations between the method factors (see Table 5) permits us (1) to find out to what degree method effects generalize across different traits (for the same rater), and (2) to analyze to what degree different raters share a common perspective that deviates from the mother's view. Concerning (1), we found that several method factors pertaining to the self and father report were very highly correlated with other method factors belonging to the same rater. One correlation was even estimated to equal 1.00 (this can happen when method effects generalize perfectly across two traits).

For the self-report, very high correlations occurred between the method factors for *Anxious/Depressed* and *Withdrawn* ( $r = 1.00$ ), *Thought Problems* and *Withdrawn* ( $r = .96$ ), *Attention Problems* and *Withdrawn* ( $r = .95$ ), *Attention Problems* and *Anxious/Depressed* ( $r = .98$ ), and for *Aggressive Behavior* and *Delinquent Behavior* ( $r = .95$ ). For the father report, the method factors for *Anxious/Depressed* and *Withdrawn* ( $r = .91$ ), *Attention Problems* and *Anxious/Depressed* ( $r = .92$ ) as well as for *Aggressive Behaviors* and *Delinquent Behaviors* ( $r = .96$ ) were almost perfectly correlated. This finding indicated that rater-specific effects generalized (almost) perfectly across some traits for the self- and father report. For example, fathers who over- (or under-)estimated a child's aggressive behavior also tended to over- (or under-)estimate the child's delinquent behavior relative to mother's evaluation. A substantive explanation for a high convergence of method-specific effects across different constructs can be the occurrence of Halo effects, where raters consistently over- or underestimate (and fail to discriminate between) different behaviors.

However, almost all of the remaining method factor intercorrelations were substantially lower, ranging from  $r = .12$  to  $.84$ . This indicated that for most traits, method effects did *not* generalize perfectly but were trait-specific. On average, the lowest correlations of this kind were found for the teacher report, indicating that teachers' biases differed more strongly across different traits than other raters'.

Another interesting question for the present study was whether fathers, teachers, and youths share a common view about symptoms/problem behaviors that is not shared with the mothers. This question can be investigated by looking at the correlations between method factors that belong to the same construct but different raters: The higher the correlations, the greater the degree to which rater-specific deviations from the reference method (mother) are shared across raters. In the present case, we found that these types of correlations were generally low ( $r < .25$ ) and that many of them were even statistically non-significant. This shows us that all methods are rather divergent in their view of children's problem behaviors and that there is almost no "common bias" (no common deviation from the mother report).

In the next step, we extended the models by including age and gender as predictors of all reference factors and all method factors to reflect the age- and gender-specificity of the obtained ratings (see Tables 4 and 5). The extended models also fit the data well,  $\chi^2(470) = 844.00$ ;  $CFI = .96$ ;  $RMSEA = .03$ ;  $SRMR = .04$  for internalizing problem behaviors and  $\chi^2(165) = 271.40$ ;  $CFI = .98$ ;  $RMSEA = .03$ ;  $SRMR = .03$  for externalizing problem behaviors.

Table 4 contains the regression coefficients and  $R^2$  values for the regression of the reference factors (mother reported problem behavior) on age and gender. A positive regression coefficient for age means that the problem behavior tends to be reported more frequently by mothers as age increases, whereas a negative coefficient indicates a decline with age in the problem behavior as reported by mothers. A positive regression coefficient for gender (coded 0 = girls, 1 = boys) indicates higher than average problem behavior in the group of boys (as reported by mothers).

These coefficients have a different meaning for the regression of the method factors on age and gender (see Table 5). For the method factors, the regression coefficients indicate whether the over- or underestimation of a given problem behavior by a specific rater (i.e., rater bias relative to the mother report) is related to the age and/or sex of the children. That is, these coefficients tell us whether the bias of a specific rater (deviation from the mother report) for a specific trait can be explained by a child's age or sex.

As expected from the literature and the data presented in Table 1, there is substantial age- and gender-related variability in the latent variables. Consistently, a unit increase in age (here age is measured in years) corresponds to a specific increase in a given symptom. Of particular interest are the relationships between age and method latent variables. Interestingly, several of the self-report method factors were significantly related to age with rather high standardized regression weights. In particular, the self-report method factors for *Withdrawn*, *Thought Problems*, *Attention Problems*, *Delinquent Behavior*, and *Aggressive Behavior* were significantly positively related to age. This shows us that the overestimation of these problem behaviors by youths relative to the mother report increases with increasing age. That is, the older the children, the more their own view seems to diverge from their mothers' view.

Furthermore, several method factors were significantly related to gender. Negative regression coefficients indicate that the over- or underestimation of a problem behavior relative to the mother rating was stronger for girls, whereas positive regression coefficients

indicate that the over- or underestimation relative to mother reports was stronger for boys. Significant coefficients were mainly found for method factors pertaining to teacher and self-ratings. For the traits *Somatic Complaints* and *Anxious/Depressed*, teachers tended to deviate more strongly from the mother's view for girls than for boys. In contrast, for *Attention Problems* and *Delinquent Behavior*, teachers tended to deviate more from mothers if the child was a boy.

Significant gender-specific method effects were also found for the self-ratings of several traits. For the traits *Somatic Complaints*, *Anxious/Depressed*, *Thought Problems*, *Attention Problems*, *Social Problems*, and *Aggressive Behavior* the self-report method factors were negatively related to gender, implying that girls tended to deviate more strongly from their mothers' view for these traits than did boys. In sum, these findings further underscore the need for a sophisticated modeling approach to the assessment of problem behaviors in children that takes the trait-specificity of method effects into account.

## Discussion

In this article, the distribution of and correlations among scores on cross-informant symptoms from three known assessment instruments, the CBCL, TRF, and YSR, are reported. The convergent and divergent validity of these instruments using an MTMM modeling approach is evaluated in a community-based Russian sample of children and adolescents.

The analyses revealed a set of interesting findings with regard to the convergent and divergent validity of the CBCL, TRF, and YSR, as they are exemplified in these data. In general, the results indicate a fairly high degree of convergent validity between parent ratings, a notable degree of validity between parent and teacher ratings, and a low degree of validity between the youths' self- and other-ratings. This lack of convergence becomes more obvious when one takes a look at the variance components (see Tables 3 and 4). It can be seen that YSR indicators show very low consistency but high method-specificity coefficients. This is the case for all but one trait, *Somatic Complaints*. Teacher ratings appear to be more in sync with parent ratings and, yet, there is much teacher-specific variability in these data, as indicated by higher modification indices from the fitted models. The specificity of teacher ratings is especially high for externalizing problems such as *Attention Problems*, *Delinquent Behavior*, and *Aggressive Behavior*. Ratings on the scale of *Delinquent Behaviors* are consistent for all adults, but inconsistent with the youths' reports. It is important to mention here that the literature indicates the limited validity of youth self-report ratings for such symptoms as attention problems (Schwab-Stone et al., 1996). The situation is different for delinquent behaviors; it might be that youth are the best informants of those behaviors because they tend to hide them from adults and, thus, adults might not be aware of them. These interesting variations in the consistency/method specificity for specific symptoms (e.g., *Anxious/Depressed*) lead to a hypothesis that different informants might view this trait very differently, almost as different traits. A similar observation was made by researchers who investigated self-reported and adult (parent and teacher) reports of anxiety and depression using different assessments (Eid, et al., 2008; Geiser, 2009). They speculated that different ratings might capture different facets of anxiety and depression (e.g., school-



related anxiety as compared to fear of novel experiences). The latent trait that demonstrates a low level of consistency for all informants is *Social Problems*. Yet, this result, perhaps, is not surprising given that similar observations were made with regard to both English (Heubeck, 2000) and Dutch (de Groot, et al., 1996) versions of these assessments. Of interest also is the finding that measurement error appears to play a minor role in these assessments; this is indicated by the high reliability coefficients for almost all of the observed variables. The exceptions are the YSR-based indicators of *Thought Problems*, *Attention Problems*, and *Delinquent Behavior* and the TRF-based scale for *Thought Problems*.

It appears that discriminant validity is variable for different latent traits. Specifically, based on the results presented in Table 4, there are different patterns of clustering. Most of the latent traits are differentiated well, and, although correlated, the correlations between them are low (e.g., .161 for *Somatic Complaints* and *Thought Problems*) or moderate (e.g., .264 for *Withdrawn* and *Somatic Complaints*), even if they conceptually belong to the same cluster of internalizing problems. Yet, the latent traits that are typically co-labeled as a set of externalizing problems (*Social Problems*, *Attention Problems*, *Delinquent Behavior*, and *Aggressive Behavior*) are highly correlated with each other, posing a question on the specificity of these scales when they are considered separately. Because the literature attests to the “separateness” of these constructs in the US, it is possible that the implicit distinctions between these types of behaviors and the items that endorse them are less demarcated by Russians linguistically and culturally. Clearly, further investigations are needed to explore the discriminant validity between these traits in Russian language and society.

It is also important to comment on the method effects. The high correlation between the CBCLs completed by the father and mother suggest an overall generalizability of their ratings. Thus, it is likely that, when only one parent is available as an informant, his or her ratings will be generalizable to those of the other parents. Yet, the latent correlations between the TRF and YSR, between themselves and with parent ratings are either moderate or non-existent. Thus, it is important to sample from all three sources, parents, teachers, and youths themselves, to generate a more accurate picture of the pathological symptomatology in a community-based sample of Russian children and adolescents. Similarly, average correlations between the observed variables are low: The mean CBCL-TRF correlation in our sample is .260 (sd = .086), the CBCL-YSR correlation is .148 (sd = .083), and the TRF-YSR correlation is .047 (sd = .071). These findings are all either quite consistent with or lower than the observations in the literature. The correlations of .25, .27, and .20, for respective pairs, were found in the 1987 meta-analysis (Achenbach, et al., 1987). There are higher (Achenbach, Dumenci, & Rescorla, 2002) and lower (Muris, Merckelbach, & Walczak, 2002) correlations presented in the literature, but the bottom line remains the same. These informants are structurally different; they contribute rather diverse information and, correspondingly, cannot be substituted in gathering and interpreting information on psychopathological symptoms in a normative sample. The importance of including multiple informants in studies of psychopathology in children and adolescents is also supported by studies from clinical populations (Achenbach & Dumenci, 2001). Specifically, there is a variability in the profiles of comorbidity of symptoms when ratings from multiple

informants are considered (McConaughy & Achenbach, 1994). There is also inconsistency in how the CBCL and YSR correlate with DSM diagnoses (Krol, De Bruyn, van Aarle, & van den Bercken, 2001; Lengua, Sadowski, Friedrich, & Fisher, 2001).

Fiske and Campbell (1992) pointed out that our understanding of “method effects” is still limited in many areas of psychology. In contrast to many other multi-method investigations, in the present study, we analyzed rater-specific effects in more detail by regressing these effects on the children’s age and gender. A notable finding was that rater-specific deviations from the mother report (which served as reference method) were related to age and gender for some traits and some types of raters. In particular, we observed that the “bias” of self-ratings (relative to mother’s ratings) increased with age for some traits. Furthermore, a consistent finding was that for many traits, the self-report bias was stronger for girls than for boys.

Although these results are of interest and correspond to the available literature, they formulate a number of possibilities for future work. For example, in these analyses, based on the constellation of the data and the evidence in the literature, the maternal report was chosen as the reference method. Here we analyzed only the psychopathology-related portion of the Achenbach instruments, but it is possible that for the competency-based analyses, especially for those questions that comment on school achievement, teacher reports should be selected as the reference method. Further investigation is needed with regard to developing guidelines of what informant is the best for what traits and in what situations. Additionally, in future studies, other explanatory variables (e.g., peer popularity for self-ratings or parental psychopathology for parent ratings) should be investigated; these variables could enhance our understanding of rater-specific effects in the assessment of psychopathological traits in children and adolescents. Furthermore, more research on the importance of these biases for both diagnosing and formulating prognoses concerning the course of psychopathology is desired. Indeed, might these observed biases be informative for diagnostic and treatment practices? This is a question that has a practical value and needs to be investigated further.

In summary, this work contributes to our understanding of the dynamics of the data on youths’ psychopathological symptoms, when such data are obtained from different informants, by employing a sophisticated latent variable modeling approach that allows us to consider the specifics of each trait and each (trait-specific) rater effect in detail. This work also contributes to furthering the investigation of psychopathological symptoms in community-based samples of children and adolescents by considering a Russian sample, a sample from a country that is viewed as a developing country by the World Bank standards. Because the majority of the world’s children and adolescents live in developing rather than developed countries, engaging youth growing up in the developing world in research is important.

Although informative, this study presents a number of weaknesses. First, although the Achenbach instruments have been used in Russia before, no norms have been developed just yet and, thus, it is difficult to scale the degree of psychopathology in this particular sample compared to a community-based sample of Russian youth. However, in the present study,

we were primarily interested in analyzing cross-informant agreement. Hence, our focus was not on norms or the identification of abnormal cases in this sample, and the use of raw scores instead of norm (e.g. T) scores was thus appropriate for the current analyses. Note that relationships with covariates such as age may be obscured if norm scores rather than raw scores are used. Consequently, CTC(M-1) analyses including covariates should always be conducted using the raw scores rather than norm scores.

Second, it is also of note that, when presented in Russian, the scales appear to be characterized by some weaknesses. Specifically, it appears that there is a lot of heterogeneity between items forming the same scale, whereas there appears to be an overlap between items forming different scales. Third, the YSR demonstrated considerably lower levels of reliability in this study, compared to both the CBCL and TRF. Further methodological explorations are needed to understand the underlying structure of these instruments when translated into Russian. Finally, the power of the analyses could be enhanced by building a more systematically structured sample of children and youths that is representative of the Russian population.

Third, in our models, we used just two observed variables per TMC. Although two indicators are sufficient to separate convergent validity, trait-specific method effects, and measurement error in the CTC(M-1) model, having three or more indicators per latent variable is generally preferred. The reason is that in general, models with three or more indicators per factor have been found to be more stable and less prone to improper solutions (i.e., out-of-range parameter estimates such as negative variances; see, e.g., Marsh, Hau, Balla, & Grayson, 1998). Although we could have constructed three indicators per TMC, this would have further increased the complexity of our already large models. Furthermore, in the present analyses, no estimation problems occurred with just two indicators, and we obtained reasonable parameter estimates and standard errors. Therefore, we can be confident that the results are “clean” even with just two indicators. Nonetheless, we recommend that researchers using latent variable analysis use at least three indicators per factor if possible.

## Conclusions

To conclude, the data presented here affirm two observations that are already present in the literature. First, our results underscore the importance of including multiple informants in assessing psychopathology in children and adolescents, and of using modern statistical approaches to data to analyze multimethod data. Second, it confirms that Russian children and adolescents, their parents and their teacher tend to endorse more problematic feelings and behaviors than children elsewhere. It is rather obvious that informants introduce systematic biases in interpreting items that sample from psychopathological feelings and behaviors. In addition, it appears that the degree of accuracy in assessing specific psychopathological behaviors varies for different symptoms and different methods. Correspondingly, a careful consideration of these biases should take place in interpreting data obtained through such checklists as the CBCL, TRF, and YSR, or other comparable instruments.

## Acknowledgments

Preparation of this article was supported in part by research grants from the National Institutes of Health (R01 DC007665, P50 HD052120, and P50 MH81756) and from the National Council for Eurasian and East European Studies. Grantees undertaking such projects are encouraged to express their professional judgment freely. Therefore, this article does not necessarily reflect the position or policies of the National Institutes of Health and the National Council for Eurasian and East European Studies, and no official endorsement should be inferred. The authors are thankful to the research teams in Russia for their invaluable efforts in data collection and data entry, to Ms. Mei Tan for her editorial assistance, and to Martin Schultze for his assistance with data analyses.

## References

- Achenbach TM. The classification of children's psychiatric symptoms: A factor analytic study. *Psychological Monographs*. 1966; 80(615)
- Achenbach TM. The Child Behavior Profile: I. Boys aged 6–11. *Journal of Consulting and Clinical Psychology*. 1978; 46:478–488. [PubMed: 670491]
- Achenbach, TM. Manual for the Child Behavior Checklist/4-18 and 1991 profile. Burlington, VT: Department of Psychiatry, University of Vermont; 1991a.
- Achenbach, TM. Manual for the Teacher's Report Form and 1991 profile. Burlington, VT: Department of Psychiatry, University of Vermont; 1991b.
- Achenbach, TM. Manual for the Youth Self-Report and 1991 profile. Burlington, VT: Department of Psychiatry, University of Vermont; 1991c.
- Achenbach, TM. Empirically based taxonomy: How to use syndromes and profile types derived from the CBCL/4-18, TRF, and YSR. Burlington, VT: University of Vermont, Department of Psychiatry; 1993.
- Achenbach TM. As others see us: Clinical and research implications of cross-informant correlations for psychopathology. *Current Directions in Psychological Science*. 2006; 15:94–98.
- Achenbach TM, Dumenci L. Advances in empirically based assessment: Revised cross-informant syndromes and new DSM-oriented scales for the CBCL, YSR, and TRF: Comment on Lengua, Sadowksi, Friedrich, and Fisher. *Journal of Consulting and Clinical Psychology*. 2001; 69:699–702. [PubMed: 11550735]
- Achenbach TM, Dumenci L, Rescorla LA. Ten-year comparisons of problems and competencies for national samples of youth: Self, parent and teacher reports. *Journal of Emotional and Behavioral Disorders*. 2002; 10:194–203.
- Achenbach, TM.; Edelbrock, C. Manual for the Child Behavior Checklist and Revised Child Behavior Profile. Burlington, VT: University of Vermont, Department of Psychiatry; 1983.
- Achenbach TM, Krukowski RA, Dumenci L, Ivanova MY. Assessment of adult psychopathology: Meta-analyses and implications of cross-informant correlations. *Psychological Bulletin*. 2005; 131:361–382. [PubMed: 15869333]
- Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*. 1987; 10:213–232. [PubMed: 3562706]
- Achenbach, TM.; Rescorla, LA. Manual for the ASEBA school-age forms and profiles. Burlington, VT: University of Vermont Research Center for Children, Youth, & Families; 2001.
- Arbuckle, JL. Full information estimation in the presence of incomplete data. In: Marcoulides, GA.; Schumacker, RE., editors. *Advanced structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates; 1996. p. 243-277.
- Becker A, Hagenberg N, Roessner V, Woerner W, Rothenberger A. Evaluation of the self-reported SDQ in a clinical setting: Do self-reports tell us more than ratings by adult informants? *European Child and Adolescent Psychiatry*. 2004; 13:17–24.
- Begovac I, Rudan V, Skocic M, Filipovic O, Szirovicza L. Comparison of self-reported and parent-reported emotional and behavioral problems in adolescents from Croatia. *Collegium Antropologicum*. 2004; 28:393–401. [PubMed: 15636099]
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990; 107:238–246. [PubMed: 2320703]

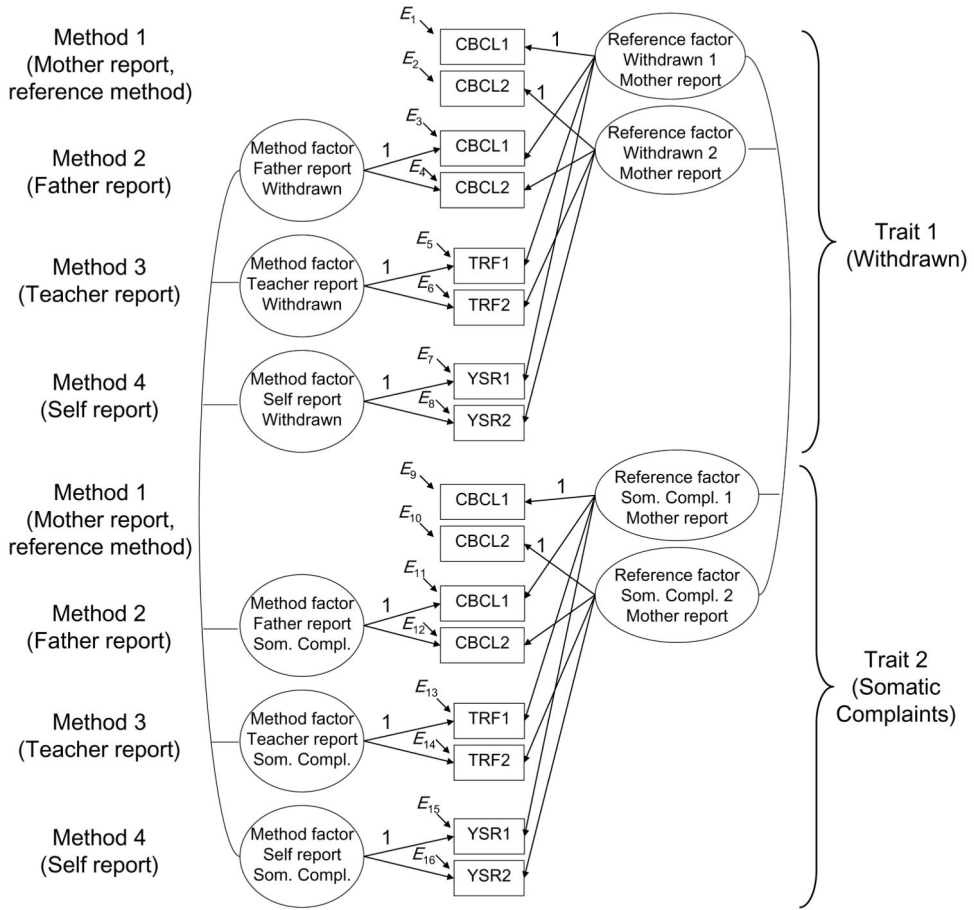
- Berg-Nielsen TS, Vika A, Dahl AA. When adolescents disagree with their mothers: CBCL-YSR discrepancies related to maternal depression and adolescent self-esteem. *Child: Care, Health & Development*. 2003; 29:207–213.
- Bérubé, RL.; Achenbach, TM. Bibliography of published studies using the Achenbach System of Empirically Based Assessment (ASEBA): 2002 edition. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families; 2002.
- Biederman J, Mick E, Faraone SV. Biased maternal reporting of child psychopathology? *Journal of the American Academy of Child & Adolescent Psychiatry*. 1998; 37:10–12. [PubMed: 9444889]
- Burns, GL.; Haynes, SN. Clinical psychology: Construct validation with multiple sources of information and multiple settings. In: Eid, M.; Diener, E., editors. *Handbook of multimethod measurement in psychology*. Washington, DC: American Psychological Association; 2006. p. 401-418.
- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*. 1959; 56:81–105. [PubMed: 13634291]
- Carlson GA, Kashani JH, Thomas MD, Vaidya A, Daniel AE. Comparison of two structured interviews on a psychiatrically hospitalized population of children. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1987; 26:645–648. [PubMed: 3667493]
- Carter AS, Grigorenko EL, Pauls DL. A Russian adaptation of the Child Behavior Checklist: Psychometric properties and associations with child and maternal affective symptomatology and family functioning. *Journal of Abnormal Child Psychology*. 1995; 23:663–686.
- Courvoisier DS, Nussbeck FW, Eid M, Geiser C, Cole DA. Analyzing the convergent and discriminant validity of states and traits: Development and applications of multimethod latent state-trait models. *Psychological Assessment*. 2008; 20:270–280. [PubMed: 18778163]
- de Groot A, Koot HM, Verhulst FC. Cross-cultural generalizability of the Youth Self-Report and Teacher's Report Form cross-informant syndromes. *Journal of Abnormal Child Psychology*. 1996; 24:651–664. [PubMed: 8956089]
- De Los Reyes A, Kazdin AE. Informant discrepancies in the assessment of childhood psychopathology: a critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*. 2005; 131:483–509. [PubMed: 16060799]
- DeGroot A, Koot HM, Verhulst FC. Cross-cultural generalizability of the Child Behavior Checklist cross-informant syndromes. *Psychological Assessment*. 1994; 6:225–230.
- Drotar D, Stein REK, Perrin EC. Methodological issues in using the Child Behavior Checklist and its related instruments in clinical child psychology research. *Journal of Clinical Child Psychology*. 1995; 24:184–192.
- Duhig AM, Renk K, Epstein MK, Phares V. Interparental agreement on internalizing, externalizing, and total behavior problems: A meta-analysis. *Clinical Psychology Science and Practice*. 2000; 7:435–453.
- Eid M. A multitrait-multimethod model with minimal assumptions. *Psychometrika*. 2000; 65:241–261.
- Eid, M.; Lischetzke, T.; Nussbeck, FW. Structural equation models for multitraitmultimethod data. In: Eid, M.; Diener, E., editors. *Handbook of multimethod measurement in psychology*. Washington, DC: American Psychological Association; 2006. p. 283-299.
- Eid M, Lischetzke T, Nussbeck FW, Trierweiler LI. Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple indicator CTC(M–1) model. *Psychological Methods*. 2003; 8:38–60. [PubMed: 12741672]
- Eid M, Nussbeck FW, Geiser C, Cole DA, Gollwitzer M, Lischetzke T. Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*. 2008; 13:230–253. [PubMed: 18778153]
- Erol N, Simsek Z, Öner O, Munir K. Epidemiology of attention problems among Turkish children and adolescents: a national study. *Journal of Attention Disorders*. 2008; 11:538–545. [PubMed: 18192617]
- Fabrega H, Ulrich R, Loeber R. Adolescent psychopathology as a function of informant and risk status. *Journal of Nervous and Mental Disease*. 1996; 184:27–34. [PubMed: 8551286]
- Fiske DW, Campbell DT. Citations do not solve problems. *Psychological Bulletin*. 1992; 112:393–395.

- Funder DC, West SG. Consensus, self-other agreement, and accuracy in personality judgment: An introduction. *Journal of Personality*. 1993; 61:457–476. [PubMed: 8151499]
- Gartstein MA, Slobodskaya HR, Kinsht IA. Cross-cultural differences in temperament in the first year of life: United States of America (US) and Russia. *International Journal of Behavioral Development*. 2003; 27:316–328.
- Geiser, C. Multitrait-multimethod-multioccasion modeling. München, Germany: AVM; 2009.
- Geiser C, Eid M, Nussbeck FW. On the meaning of the latent variables in the CT-C(M-1) model: A comment on Maydeu-Olivares and Coffman (2006). *Psychological Methods*. 2008; 13:49–57. [PubMed: 18331153]
- Goodman R, Meltzer H, Bailey V. The strengths and difficulties questionnaire: A pilot study on the validity of the self-report version. *European Child and Adolescent Psychiatry*. 1998; 7:125–130. [PubMed: 9826298]
- Goodman R, Slobodskaya H, Knyazev GG. Russian child mental health: A cross-sectional study of prevalence and risk factors. *European Child & Adolescent Psychiatry*. 2005; 14:28–33. [PubMed: 15756513]
- Hellinckx, W.; Grietens, H.; De Munter, A. Parent-reported problem behavior in 12–16-year-old American and Russian children: A cross-national comparison. In: Singh, N.; Leung, JP.; Singh, A., editors. *International perspectives on child and adolescent mental health: Selected Proceedings of the First International Conference on Child & Adolescent Mental Health*. Kidlington, UK: Elsevier; 2000. p. 205-222.
- Herjanic B, Reich W. Development of a structured psychiatric interview for children: agreement between child and parent on individual symptoms. *Journal of Abnormal Child Psychology*. 1982; 10:307–324. [PubMed: 7175040]
- Heubeck BG. Cross-cultural generalizability of CBCL syndromes across three continents: from the USA and Holland to Australia. *Journal of Abnormal Child Psychology*. 2000; 28:439–450. [PubMed: 11100918]
- Hu L, Bentler PM. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*. 1998; 3:424–453.
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55.
- Jensen PS, Rubio-Stipec M, Canino G, Bird HR, Dulcan MK, Schwab-Stone ME, et al. Parent and child contributions to diagnosis of mental disorder: are both informants always necessary? *Journal of the American Academy of Child & Adolescent Psychiatry*. 1999; 38:1569–1579. [PubMed: 10596258]
- Knyazev GG, Slobodskaya HR, Safronova MV, Kinsht IA. School adjustment and health in Russian adolescents. *Psychology, Health & Medicine*. 2002; 7:143–155.
- Knyazev GG, Zupancic M, Slobodskaya HR. Child personality in Slovenia and Russia: Structure and mean level of traits in parent and self-ratings. *Journal of Cross-Cultural Psychology*. 2008; 39:317–334.
- Krol NPCM, De Bruyn EEJ, van Aarle E, JM, van den Bercken JHL. Computerized screening for DSM classifications using CBCL/YSR extended checklists: A clinical try-out. *Computers in Human Behavior*. 2001; 17:315–337.
- Kuznetsova IV, Grigorenko EL, Voronkova OI. Initial diagnostics of abnormal development in school settings using the CBCL. *Shkola Zdorovia*. 1996; 3(2):48–60.
- Lahey BB, Flagg EW, Bird HR, Schwab-Stone ME, Canino G, Dulcan MK, et al. The NIMH Methods for the Epidemiology of Child and Adolescent Mental Disorders (MECA) Study: background and methodology. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1996; 35:855–864. [PubMed: 8768345]
- Lau AS, Garland AF, Yeh M, McCabe KM, Wood PA, Hough RL. Race/ethnicity and inter-informant agreement in assessing adolescent psychopathology. *Journal of Emotional and Behavioral Disorders*. 2004; 12:145–156.
- Lengua LJ, Sadowski CA, Friedrich WN, Fisher J. Rationally and empirically derived dimensions of children's symptomatology: Expert ratings and confirmatory factor analyses of the CBCL. *Journal of Consulting and Clinical Psychology*. 2001; 69:683–698. [PubMed: 11550734]

- Little, RJA.; Rubin, DB. Statistical analysis with missing data. New York, NY: Wiley; 2003.
- Loeber R, Green SM, Lahey BB, Stouthamer-Loeber M. Optimal informants on childhood disruptive behaviors. *Development and Psychopathology*. 1989; 1:317–337.
- Macmann GM, Barnett DW, Lopez EJ. The Child Behavior Checklist/4-18 and related materials: Reliability and validity of syndromal assessment. *School Psychology Review*. 1993; 22:322–333.
- Marsh HW. Multitrait-multimethod analyses: Inferring each trait/method combination with multiple indicators. *Applied Measurement in Education*. 1993; 6:49–81.
- Marsh HW, Hocevar D. A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology*. 1988; 73:107–117.
- Martin JL, Ford CB, Dyer-Friedman J, Tang J, Huffman LC. Patterns of agreement between parent and child ratings of emotional and behavioral problems in an outpatient clinical setting: When children endorse more problems. *Developmental and Behavioural Pediatrics*. 2004; 25:150–155.
- McAuley C, Trew K. Children's adjustment over time in foster care: Cross-informant agreement, stability and placement disruption. *British Journal of Social Work*. 2000; 30:91–107.
- McConaughy SH, Achenbach TM. Comorbidity of empirically based syndromes in matched general population and clinical samples. *Journal of Child Psychology and Psychiatry*. 1994; 35:1141–1157. [PubMed: 7995848]
- Messick S. Validity of psychological assessment. *American Psychologist*. 1995; 50:741–749.
- Montgomery E. Self- and parent assessment of mental health: disagreement on externalizing and internalizing behaviour in young refugees from the Middle East. *Clinical Child Psychology & Psychiatry*. 2008; 13:49–63. [PubMed: 18411865]
- Muris P, Merckelbach H, Walczak S. Aggression and threat perception abnormalities in children with learning and behavior problems. *Child Psychiatry and Human Development*. 2002; 33:147–163. [PubMed: 12462352]
- Muthén, LK.; Muthén, BO. Mplus user's guide. Los Angeles, CA: Author; 2006.
- Reich W, Herjanic B, Welner Z, Gandhi PR. Development of a structured psychiatric interview for children: agreement on diagnosis comparing child and parent interviews. *Journal of Abnormal Child Psychology*. 1982; 10:325–336. [PubMed: 7175041]
- Renk K, Phares V. Cross-informant ratings of social competence in children and adolescents. *Clinical Psychology Review*. 2004; 24:239–254. [PubMed: 15081518]
- Rescorla L, Achenbach TM, Ivanova MY, Dumenci L, Almqvist F, Bilenberg N, et al. Epidemiological comparisons of problems and positive qualities reported by adolescents in 24 countries. *Journal of Consulting and Clinical Psychology*. 2007; 75:351–358. [PubMed: 17469893]
- Rousseau C, Drapeau A. Parent-child agreement on refugee children's psychiatric symptoms: a transcultural perspective. *Journal of the American Academy of Child & Adolescent Psychiatry*. 1998; 37:629–636. [PubMed: 9628083]
- Sawyer MG, Baghurst P, Mathias J. Differences between informants' reports describing emotional and behavioural problems in community and clinic-referred children: A research note. *Journal of Child Psychology and Psychiatry*. 1992; 33:441–449. [PubMed: 1564085]
- Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research - Online*. 2003; 8:23–74.
- Schwab-Stone ME, Shaffer D, Dulcan MK, Jensen PS, Fisher P, Bird HR, et al. Criterion validity of the NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC-2.3). *Journal of the American Academy of Child & Adolescent Psychiatry*. 1996; 35:878–888. [PubMed: 8768347]
- Sechrest, L.; Davis, M.; Stickle, T.; McKnight, P. Understanding "method" variance. In: Bickman, L., editor. *Research design: Donald Campbell's legacy*. Thousand Oaks, CA: Sage; 2000. p. 63-87.
- Seiffge-Krenke I, Kollmar F. Discrepancies between mothers' and fathers' perceptions of sons' and daughters' problem behaviour: a longitudinal analysis of parent-adolescent agreement on internalising and externalising problem behaviour. *Journal of Child Psychology & Psychiatry & Allied Disciplines*. 1998; 39:687–697.

- Share DL. On the anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin*. 2008; 134:584–615. [PubMed: 18605821]
- Slobodskaya HR. Competence, emotional and behavioral problems in Russian adolescents. *European Child and Adolescent Psychiatry*. 1999; 8:173–180. [PubMed: 10550698]
- Smith SR. Making sense of multiple informants in child and adolescent psychopathology: A guide for clinicians. *Journal of Psychoeducational Assessment*. 2007; 25:139–149.
- Stanger C, Lewis M. Agreement among parents, teachers, and children on internalizing and externalizing behaviour problems. *Journal of Clinical Child Psychology*. 1993; 22:107–115.
- Stanger C, MacDonald VV, McConaughy SH, Achenbach TM. Predictors of cross-informant syndromes among children and youths referred for mental health services. *Journal of Abnormal Child Psychology*. 1996; 24:597–614. [PubMed: 8956086]
- Steiger JH. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*. 1990; 25:173–180.
- Stevens GW, Vollebergh WA, Pels TV, Crijnen AA. Predicting internalizing problems in Moroccan immigrant adolescents in The Netherlands. *Social Psychiatry & Psychiatric Epidemiology*. 2005; 40:1003–1011. [PubMed: 16341615]
- van den Ende J, Verhulst FC. Informant, gender and age differences in ratings of adolescent problem behaviour. *European Child and Adolescent Psychiatry*. 2005; 14:117–126. [PubMed: 15959657]
- Wahlsten VS, Ahmad A, Von Knorring AL. Do Kurdistanian and Swedish parents and children differ in their rating of competence and behavioural problems? *Nordic Journal of Psychiatry*. 2002; 56:279–283. [PubMed: 12470319]
- Wals M, Hillegers MH, Reichart CG, Ormel J, Nolen WA, Verhulst FC. Prevalence of psychopathology in children of a bipolar parent. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2001; 40:1094–1102. [PubMed: 11556634]
- Wothke, W. Longitudinal and multi-group modeling with missing data. In: Little, TD.; Schnabel, KU.; Baumert, J., editors. *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples*. Hillsdale, NJ: Erlbaum; 2000. p. 219-240.
- Youngstrom E, Loeber R, Stouthamer-Loeber M. Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting & Clinical Psychology*. 2000; 68:1038–1050. [PubMed: 11142538]
- Zahn-Waxler C, Klimes-Dougan B, Slattery MJ. Internalizing problems of childhood and adolescence: prospects, pitfalls, and progress in understanding the development of anxiety and depression. *Development & Psychopathology*. 2000; 12:443–466. [PubMed: 11014747]
- Zimmerman RS, Khoury E, Vega W, Gil A, Warheit G. Teacher and parent perceptions of behavior problems among a sample of African American, Hispanic, and Non-Hispanic White students. *American Journal of Community Psychology*. 1995; 23:181–197. [PubMed: 7572829]





**Figure 1.** Path diagram of a CT-C(M-1) Model with indicator-specific reference factors and trait-specific method factors for the traits *Withdrawn* and *Somatic Complaints*. Two indicators (test halves) were used for each trait-method combination to enable the analysis of trait-specific method effects. The connections between the factors indicate that all reference and all method factors can be correlated. The variables  $E_1$  to  $E_{16}$  are measurement error variables. The model is identified by fixing the first loading per factor to 1.00. The remaining loadings are freely estimated. Two traits are shown only to reduce the size and complexity of the figure. The models actually estimated in the present study considered more than two traits simultaneously.

**Table 1**

Descriptive Statistics for Problem Scales in the CBCL, TRF, and YSR

	CBCL		TRF	YSR
	Mother	Father		
<b>Problem Scales</b>	<i>Mean(sd)</i>	<i>Mean(sd)</i>	<i>Mean(sd)</i>	<i>Mean(sd)</i>
<b>Withdrawn</b>				
Boys	3.54(2.65)	3.27(2.53)	3.02(3.08)	3.68(2.27)
Girls	3.50(2.71)	3.16(2.51)	3.21(2.95)	3.49(2.22)
Total Sample	3.52(2.68)	3.21(2.52)	3.13(3.01)	3.57(2.25)
<b>Somatic Complaints</b>				
Boys	2.85(2.87)	2.06(2.62)	1.43(3.20)	2.77(3.11)
Girls	3.20(2.92)	2.45(2.65)	1.86(2.92)*	4.33(3.30)***
Total Sample	3.04(2.90)	2.27(2.64)	1.68(2.89)	3.59(3.30)
<b>Anxious/Depressed</b>				
Boys	4.59(3.56)	4.03(3.37)	4.86(4.37)	6.02(4.33)
Girls	5.50(4.15)**	4.68(3.90)	5.71(4.47)**	7.72(4.85)***
Total Sample	5.08(3.91)	4.38(3.66)	5.34(4.44)	6.98(4.70)
<b>Social Problems</b>				
Boys	2.51(2.13)	2.15(2.00)	2.49(2.93)	2.97(2.10)
Girls	2.28(2.10)	2.28(2.08)	2.09(2.93)	3.03(2.18)
Total Sample	2.39(2.11)	2.22(2.04)	2.26(2.93)	3.00(2.14)
<b>Thought Problems</b>				
Boys	0.93(1.58)**	0.57(0.99)	0.32(0.80)	2.60(2.24)
Girls	0.63(1.13)	0.75(1.30)	0.36(0.85)	3.24(2.48)**
Total Sample	0.77(1.36)	0.66(1.15)	0.34(0.83)	2.96(2.40)
<b>Attention Problems</b>				
Boys	5.82(3.76)***	5.31(3.66)**	9.17(6.81)***	10.31(8.80)
Girls	4.45(3.15)	4.26(3.14)	5.59(5.19)	11.44(8.94)
Total Sample	5.08(3.51)	4.74(3.43)	7.14(6.20)	10.95(8.89)
<b>Delinquent Behavior</b>				
Boys	2.35(2.12)***	2.36(2.32)***	2.41(2.76)***	4.52(2.79)
Girls	1.39(1.60)	1.46(1.71)	1.45(1.99)	4.62(3.00)
Total Sample	1.83(1.92)	1.88(2.06)	1.86(2.40)	4.58(2.91)
<b>Aggressive Behavior</b>				
Boys	9.23(5.86)***	8.24(5.78)**	7.56(8.41)***	8.55(5.23)
Girls	6.83(4.35)	6.76(4.80)	3.92(5.81)	9.36(5.52)

	CBCL		TRF	YSR
	Mother	Father		
Problem Scales	<i>Mean(sd)</i>	<i>Mean(sd)</i>	<i>Mean(sd)</i>	<i>Mean(sd)</i>
Total Sample	7.93(5.23)	7.45(5.33)	5.49(7.28)	9.01(5.41)

Note:

The group with the higher value is noted with \*

\* for  $p < .05$ ;

\*\* for  $p < .01$ ; and

\*\*\* for  $p < .001$ .

The corresponding  $n(s)$  for all instruments by gender are: CBCL<sub>Mother</sub>—324 girls and 276 boys; CBCL<sub>Father</sub>—219 girls and 190 boys; TRF—423 girls and 322 boys; YSR—284 girls and 218 boys.

**Table 2**  
 Manifest Correlations among Problems Scales in the CBCL (Mother and Father), TRF, and YSR

	CBCL <sub>M</sub> -CBCL <sub>F</sub>	CBCL <sub>M</sub> -TRF	CBCL <sub>M</sub> -YSR	CBCL <sub>F</sub> -TRF	CBCL <sub>F</sub> -YSR	TRF-YSR
Problem Scales	<i>r</i> ( <i>p</i> )	<i>r</i> ( <i>p</i> )	<i>r</i> ( <i>p</i> )	<i>r</i> ( <i>p</i> )	<i>r</i> ( <i>p</i> )	<i>r</i> ( <i>p</i> )
Withdrawn	.613(.000)	.261(.000)	.076(.171)	.219(.000)	.058(.367)	-.005(.918)
Somatic Complaints	.629(.000)	.299(.000)	.295(.000)	.298(.000)	.210(.001)	.142(.004)
Anxious/Depressed	.653(.000)	.259(.000)	.252(.000)	.141(.007)	.268(.000)	.046(.352)
Social Problems	.609(.000)	.286(.000)	.105(.060)	.141(.007)	.113(.078)	-.046(.351)
Thought Problems	.419(.000)	.190(.000)	.155(.005)	.080(.125)	.104(.103)	.078(.113)
Attention Problems	.710(.000)	.414(.000)	.246(.000)	.311(.000)	.172(.007)	.033(.503)
Delinquent Behavior	.690(.000)	.313(.000)	.128(.021)	.326(.000)	.060(.350)	.145(.003)
Aggressive Behavior	.692(.000)	.336(.000)	.101(.071)	.283(.000)	.032(.620)	-.016(.750)

*Note:* Table shows manifest correlations of total sum scores.

M = Mother; F = Father

**Table 3**

CT-C(M-1) Model for Multiple Informants: Intercepts, Factor Loadings, Reliability, Consistency/Method Specificity, and Explained Variance in Observed Variables

Problem Scales/Method	Intercept (US/S)	Trait factor loading (US/S)	Method factor loading (US/S)	Consistency	Method Specificity	Reliability
<i>Withdrawn</i>						
CBCL <sub>Mother</sub> 1	0.43/1.07	1.00/0.88				.77
CBCL <sub>Mother</sub> 2	0.35/1.02	1.00/0.83				.69
CBCL <sub>Father</sub> 1	0.39/1.05	0.70/0.66	1.00/0.45	.44	.20	.64
CBCL <sub>Father</sub> 2	0.31/0.93	0.80/0.68	1.01/0.50	.46	.25	.71
TRF 1	0.46/0.98	0.36/0.27	1.00/0.78	.07	.61	.68
TRF 2	0.32/0.88	0.35/0.27	0.76/0.77	.07	.59	.66
YSR 1	0.40/1.00	-0.12/-0.11(n.s.)	1.00/0.40	.01	.16	.17
YSR 2	0.45/1.43	0.19/0.18	0.93/0.47	.03	.22	.25
<i>Somatic Complaints</i>						
CBCL <sub>Mother</sub> 1	0.27/0.82	1.00/0.93				.86
CBCL <sub>Mother</sub> 2	0.42/1.09	1.00/0.91				.83
CBCL <sub>Father</sub> 1	0.21/0.65	0.67/0.63	1.00/0.37	.40	.14	.54
CBCL <sub>Father</sub> 2	0.31/0.83	0.69/0.66	1.89/0.61	.44	.37	.81
TRF 1	0.20/0.46	0.34/0.24	1.00/0.64	.06	.41	.47
TRF 2	0.23/0.63	0.30/0.28	1.01/0.76	.08	.58	.66
YSR 1	0.35/0.78	0.39/0.27	1.00/0.59	.07	.34	.41
YSR 2	0.44/1.07	0.32/0.27	1.18/0.75	.07	.56	.63
<i>Anxious/Depressed</i>						
CBCL <sub>Mother</sub> 1	0.36/1.17	1.00/0.95				.90
CBCL <sub>Mother</sub> 2	0.36/1.17	1.00/0.89				.79
CBCL <sub>Father</sub> 1	0.29/1.02	0.68/0.71	1.00/0.52	.50	.27	.77
CBCL <sub>Father</sub> 2	0.32/1.07	0.71/0.64	1.16/0.58	.41	.34	.75
TRF 1	0.27/0.95	0.25/0.26	1.00/0.85	.07	.72	.79
TRF 2	0.25/0.94	0.27/0.28	0.67/0.61	.08	.37	.45

Problem Scales/Method	Intercept (US/S)	Trait factor loading (US/S)	Method factor loading (US/S)	Consistency	Method Specificity	Reliability
YSR 1	0.41/1.35	0.23/0.22	1.00/0.67	.05	.45	.50
YSR 2	0.56/1.62	0.25/0.20	1.18/0.69	.04	.48	.52
<i>Social Problems</i>						
CBCL <sub>Mother</sub> 1	0.37/0.95	1.00/0.88				.77
CBCL <sub>Mother</sub> 2	0.29/0.97	1.00/0.89				.79
CBCL <sub>Father</sub> 1	0.34/0.93	0.68/0.63	1.00/0.45	.40	.20	.60
CBCL <sub>Father</sub> 2	0.28/0.94	0.68/0.60	0.90/0.50	.36	.25	.61
TRF 1	0.14/0.55	0.15/0.20	1.00/0.67	.04	.45	.49
TRF 2	0.22/0.72	0.28/0.24	1.25/0.71	.06	.50	.56
YSR 1	0.39/1.03	0.18/0.16	1.00/0.75	.03	.56	.59
YSR 2	0.44/1.22	0.12/0.09 (n.s.)	0.75/0.60	.01	.36	.37
<i>Thought Problems</i>						
CBCL <sub>Mother</sub> 1	0.12/0.49	1.00/0.72				.52
CBCL <sub>Mother</sub> 2	0.07/0.33	1.00/0.88				.77
CBCL <sub>Father</sub> 1	0.11/0.49	0.60/0.47	1.00/0.47	.22	.22	.44
CBCL <sub>Father</sub> 2	0.06/0.33	0.41/0.40	0.86/0.46	.20	.21	.41
TRF 1	0.04/0.28	0.27/0.31	1.00/0.92	.10	.85	.95
TRF 2	0.02/0.18	0.08/0.14	0.27/0.35	.02	.12	.14
YSR 1	0.59/1.28	0.04/0.02 (n.s.)	1.00/0.64	.00	.41	.41
YSR 2	0.42/1.18	0.17/0.09 (n.s.)	0.66/0.54	.01	.29	.30
<i>Attention Problems</i>						
CBCL <sub>Mother</sub> 1	0.48/1.29	1.00/0.89				.77
CBCL <sub>Mother</sub> 2	0.56/1.38	1.00/0.86				.74
CBCL <sub>Father</sub> 1	0.47/1.27	0.82/0.74	1.00/0.49	.55	.24	.79
CBCL <sub>Father</sub> 2	0.52/1.32	0.77/0.68	1.07/0.49	.46	.24	.70
TRF 1	0.33/0.93	0.54/0.50	1.00/0.68	.25	.46	.71
TRF 2	0.32/0.93	0.39/0.40	0.91/0.65	.20	.42	.62
YSR 1	0.72/1.93	0.17/0.15	1.00/0.65	.02	.42	.44

Problem Scales/Method	Intercept (US/S)	Trait factor loading (US/S)	Method factor loading (US/S)	Consistency	Method Specificity	Reliability
YSR 2	0.65/1.76	0.11/0.11 (n.s.)	1.11/0.74	.01	.55	.56
<i>Delinquent Behavior</i>						
CBCL <sub>Mother</sub> 1	0.17/0.79	1.00/0.78				.61
CBCL <sub>Mother</sub> 2	0.23/0.88	1.00/0.90				.81
CBCL <sub>Father</sub> 1	0.18/0.80	0.88/0.65	1.00/0.40	.42	.20	.62
CBCL <sub>Father</sub> 2	0.23/0.84	0.85/0.72	1.09/0.36	.52	.13	.65
TRF 1	0.15/0.62	0.47/0.31	1.00/0.80	.10	.64	.74
TRF 2	0.24/0.74	0.39/0.27	1.17/0.71	.07	.50	.57
YSR 1	0.57/1.56	0.26/0.12	1.00/0.60	.01	.36	.37
YSR 2	0.42/1.13	0.14/0.09 (n.s.)	1.20/0.72	.01	.52	.53
<i>Aggressive Behavior</i>						
CBCL <sub>Mother</sub> 1	0.37/1.27	1.00/0.90				.81
CBCL <sub>Mother</sub> 2	0.37/1.39	1.00/0.96				.92
CBCL <sub>Father</sub> 1	0.33/1.19	0.75/0.70	1.00/0.54	.49	.29	.78
CBCL <sub>Father</sub> 2	0.36/1.32	0.74/0.70	1.10/0.61	.49	.37	.86
TRF 1	0.19/0.73	0.36/0.37	1.00/0.79	.14	.62	.76
TRF 2	0.22/0.70	0.31/0.26	1.41/0.93	.07	.86	.93
YSR 1	0.57/2.04	0.16/0.15	1.00/0.72	.02	.52	.54
YSR 2	0.55/2.17	0.17/0.17	0.91/0.72	.03	.52	.55

Note:

Mother report = reference method. US = unstandardized parameter estimate; S = standardized parameter estimate; n.s. = not significantly different from zero ( $p > .05$ ); Consistency = proportion of variance in an observed indicator that is shared with the reference method. Method specificity = proportion of variance in an observed indicator that is specific to the particular non-reference method and not shared with the reference method (not applicable for the reference method). Reliability = proportion of reliable observed variance (sum of consistency and method specificity).

**Table 4**  
 CT-C(M-1) Model: Latent Correlations Between Problem Behaviors as Reported by Mothers and Regression on Gender and Age

	W1	W2	SC1	SC2	A/D1	A/D2	TP1	TP2	API	AP2	R <sup>2</sup>
<i>Internalizing Problem Scales</i>											
Withdrawn 1											.00 (n.s.)
Withdrawn 2	.62										.05
Somatic Complaints 1	.26	.28									.00 (n.s.)
Somatic Complaints 2	.32	.36	.68								.01 (n.s.)
Anxious/Depressed 1	.63	.65	.45	.50							.02
Anxious/Depressed 2	.45	.69	.38	.47	.74						.01 (n.s.)
Thought Problems 1	.36	.50	.32	.38	.57	.64					.01 (n.s.)
Thought Problems 2	.07 (n.s.)	.19	.14	.17	.13	.33	.52				.04
Attention Problems 1	.27	.34	.29	.29	.46	.56	.51	.21			.10
Attention Problems 2	.30	.40	.32	.30	.53	.57	.51	.30	.83		.09
<i>Regression coefficients (US\$)</i>											
Age	0.00/0.01 (n.s.)	0.03/0.23	0.00/0.03 (n.s.)	0.01/0.07 (n.s.)	-0.01/-0.11	-0.00/-0.02 (n.s.)	-0.01/-0.08 (n.s.)	-0.01/-0.13	-0.03/-0.19	-0.03/-0.22	
Gender	-0.04/-0.06 (n.s.)	-0.00/-0.00 (n.s.)	-0.02/-0.04 (n.s.)	-0.06/-0.08 (n.s.)	-0.07/-0.12	-0.04/-0.08 (n.s.)	0.02/0.05 (n.s.)	0.05/0.15	0.17/0.25	0.14/0.20	
<i>Externalizing Problem Scales</i>											
Social Problems 1		SP1		DB1	AB1	AB2					.07
Social Problems 2	.55										.01 (n.s.)
Delinquent Behavior 1	.36	.31									.05
Delinquent Behavior 2	.32	.26	.67								.07
Aggressive Behavior 1	.55	.49	.61	.55							.09
Aggressive Behavior 2	.55	.46	.63	.55	.90						.10
<i>Regression coefficients (US\$)</i>											
Age	-0.03/-0.25	-0.01/-0.10 (n.s.)	-0.00/-0.02 (n.s.)	0.01/0.10	-0.02/-0.15	-0.02/-0.14					
Gender	0.06/0.09	0.02/0.04 (n.s.)	0.07/0.21	0.12/0.25	0.14/0.26	0.14/0.28					

Note: W = Withdrawn; SC = Somatic Complaints; A/D = Anxious/Depressed; TP = Thought Problems; AP = Attention Problems. For the variable Gender, Girls are coded with 0, Boys are coded with 1. US = unstandardized; S = standardized; n.s. = not significant ( $p > .05$ ).



**Table 5**  
**CT-C(M-1) Model: Correlations Between Latent Method Factors and Regression on Gender and Age**

	WF	WT	WS	SCF	SCT	SCS	ADF	ADT	ADS	TPF	TPT	TTS	APF	APT	APS	R <sup>2</sup>
<i>Internalizing Problem Scales</i>																
Withdrawn Father (WF)																.00 (n.s.)
Withdrawn Teacher (WT)	.05 (n.s.)															.01 (n.s.)
Withdrawn Self (WS)	.23 (n.s.)	-.09 (n.s.)														.32
Somatic Complaints Father (SCF)	.46	.01 (n.s.)	-.16 (n.s.)													.01 (n.s.)
Somatic Complaints Teacher (SCT)	.06 (n.s.)	.36	-.04 (n.s.)	.15 (n.s.)												.01 (n.s.)
Somatic Complaints Self (SCS)	.17 (n.s.)	.20	.69	-.06 (n.s.)	.10 (n.s.)											.08
Anxious/Depressed Father (ADF)	.91	-.13 (n.s.)	.20 (n.s.)	.57	-.01 (n.s.)	.07 (n.s.)										.00 (n.s.)
Anxious/Depressed Teacher (ADT)	.13 (n.s.)	.84	-.24	.04 (n.s.)	.54	.11 (n.s.)	-.04 (n.s.)									.01 (n.s.)
Anxious/Depressed Self (ADS)	.13 (n.s.)	.17	1.00	-.08 (n.s.)	.05 (n.s.)	.73	.20 (n.s.)	.01 (n.s.)								.03 (n.s.)
Thought Problems Father (TPF)	.49	-.10 (n.s.)	.06 (n.s.)	.14 (n.s.)	.01 (n.s.)	.17 (n.s.)	.68	-.13 (n.s.)	-.04 (n.s.)							.00 (n.s.)
Thought Problems Teacher (TPT)	.07 (n.s.)	.12	.16 (n.s.)	.02 (n.s.)	.14	.13 (n.s.)	-.15 (n.s.)	.31	-.01 (n.s.)	.00 (n.s.)						.00 (n.s.)
Thought Problems Self (TPS)	.27	.07 (n.s.)	.96	.06 (n.s.)	.02 (n.s.)	.54	.30	.05 (n.s.)	.83	.01 (n.s.)	-.09 (n.s.)					.15 (n.s.)
Attention Problems Father (APF)	.79	-.18 (n.s.)	.06 (n.s.)	.40	-.05 (n.s.)	.02 (n.s.)	.92	-.07 (n.s.)	.11 (n.s.)	.74	-.02 (n.s.)	.09 (n.s.)				.01 (n.s.)
Attention Problems Teacher (APT)	.14 (n.s.)	.34	.07 (n.s.)	-.01 (n.s.)	.39	.06 (n.s.)	-.01 (n.s.)	.61	-.16 (n.s.)	-.11 (n.s.)	.37	-.01 (n.s.)	-.09 (n.s.)			.04
Attention Problems Self (APS)	.20 (n.s.)	.19	.95	-.12 (n.s.)	.03 (n.s.)	.70	.16 (n.s.)	.05 (n.s.)	.98	.00 (n.s.)	.06 (n.s.)	.81	.05 (n.s.)	-.04 (n.s.)		.21
<i>Regression coefficients (US\$)</i>																
Age	0.00/0.03 (n.s.)	0.02/0.11	0.05/0.56	-0.00/-0.09 (n.s.)	0.01/0.05 (n.s.)	0.01/0.11 (n.s.)	0.00/0.01 (n.s.)	-0.01/-0.07 (n.s.)	0.00/0.01 (n.s.)	0.00/0.04 (n.s.)	-0.00/-0.03 (n.s.)	0.04/0.34	-0.01/-0.07 (n.s.)	0.01/0.05 (n.s.)	0.04/0.34	
Gender	0.00/0.01 (n.s.)	-0.03/-0.05 (n.s.)	0.02/0.05 (n.s.)	-0.00/-0.02 (n.s.)	-0.05/-0.10	-0.14/-0.26	0.01/0.02 (n.s.)	-0.05/-0.10	-0.07/-0.17	0.01/0.04 (n.s.)	-0.00/-0.01 (n.s.)	-0.11/-0.18	-0.00/-0.00 (n.s.)	0.09/0.19	-0.16/-0.30	
<i>Externalizing Problem Scales</i>																
Social Problems Father (SPF)		SPT	SPS	DBF	DBT	DBS	ABF	ABT	ABS							
Social Problems Teacher (SPT)	-.13 (n.s.)															.00 (n.s.)
Social Problems Self (SPS)	.17 (n.s.)	-.17														.01 (n.s.)
Delinquent Behavior Father (DBF)	.56	.02 (n.s.)	-.07 (n.s.)													.04 (n.s.)
Delinquent Behavior Teacher (DBT)	-.06 (n.s.)	.57	.03 (n.s.)	.25												.00 (n.s.)
Delinquent Behavior Self (DBS)	-.07 (n.s.)	-.03 (n.s.)	.37	.16 (n.s.)	.05 (n.s.)											.03
Aggressive Behavior Father (ABF)	.70	-.05 (n.s.)	.02 (n.s.)	.96	.06 (n.s.)	.01 (n.s.)										.37
Aggressive Behavior Teacher (ABT)	-.05 (n.s.)	.53	-.15	.04 (n.s.)	.74	.03 (n.s.)	.14 (n.s.)									.00 (n.s.)

	WF	WT	WS	SCF	SCT	SCS	ADF	ADT	ADS	TPF	TPT	TPS	APF	APT	APS	R <sup>2</sup>
Aggressive Behavior Self (ABS)	.11 (n.s.)	-.07 (n.s.)	.55	-.02 (n.s.)	.02 (n.s.)	.95	-.08 (n.s.)	-.11 (n.s.)								.22
<i>Regression coefficients (US/S)</i>																
Age	-.00/-0.02	-0.01/-0.07 (n.s.)	-0.02/-0.12 (n.s.)	0.00/0.02 (n.s.)	0.01/0.08 (n.s.)	0.06/0.60	0.00/0.01 (n.s.)	-0.00/-0.02 (n.s.)	0.04/0.45							
Gender	-0.02/-0.06	0.03/0.08 (n.s.)	-0.10/-0.16	0.00/-0.01 (n.s.)	0.06/0.16	-0.04/-0.08 (n.s.)	-0.01/-0.04 (n.s.)	0.08/0.20	-0.05/-0.11							

Note. For the variable Gender, Girls are coded with 0, Boys are coded with 1. US = unstandardized; S = standardized; n.s. = not significant ( $p > .05$ ).