Confirmatory Factor Analysis of the Strengths and Difficulties Questionnaire in a Community Sample of French-Speaking Adolescents

Mathieu d’Acremont¹ and Martial Van der Linden²

¹Swiss Centre for Affective Sciences, ²Cognitive Psychopathology and Neuropsychology Unit, both University of Geneva, Switzerland

Abstract. The Strengths and Difficulties Questionnaire (SDQ) assesses adaptive and problematic behavior in children and adolescents (Goodman, 1997). The aim of this study was to test the construct validity of a French translation of the scale. Teachers completed the SDQ for their pupils (279 girls and 278 boys, 13–18 years). Confirmatory factor analysis (CFA) of the SDQ supported the original distinction between Conduct problems, Hyperactivity/Inattention, Peer problems, Emotional symptoms, and Prosocial behavior. Multi-group CFA revealed invariance of factor measurement across gender. In addition, boys had higher factor scores for Conduct problems, Hyperactivity/Inattention, and Peer problems whereas girls were more Prosocial. The internal reliability of the subscales ranges from acceptable to very good. These results indicate that the French version of the SDQ has a reliable factor organization among adolescent boys and girls.

Keywords: adolescence, behavior assessment, scale validation, construct validity, measurement invariance, confirmatory factor analysis

The Strengths and Difficulties Questionnaire (SDQ) was developed by Goodman (1997) to assess behavior in childhood and adolescence. More specifically, it allows the evaluation of four types of problematic behavior in youth: (1) Conduct problems, which reflect antisocial, aggressive, and oppositional behavior; (2) Hyperactivity/Inattention, which corresponds to impulsive behavior reflected by agitation and distraction; (3) Peer problems, which reflect poor relationships with other children such as loneliness or victimization; and (4) Emotional symptoms, which reflect anxiety and sadness. A fifth dimension of the scale evaluates an adaptive type of behavior: (5) Prosocial behavior, which corresponds to empathy and kindness. The scale is being used increasingly often in clinical applications and research (for a review, see Vostanis, 2006). For instance, the SDQ has been used in an educational context to better understand the behavior of children who are excluded at school or victimized or bullied (Ripley & Yuill, 2005; Smith, Talamelli, Cowie, Naylor, & Chauhan, 2004). Another interesting development of the SDQ concerns etiological research highlighting the genetic, environmental, or psychological factors underlying problematic behavior in youth (e.g., Scourfield, Van den Bree, Martin, & McGuffin, 2004; d’Acremont & Van der Linden, 2007).

The SDQ has several advantages that may explain its popularity. It takes into account the positive aspects of behavior (strengths) and not just the negative ones (difficulties). Moreover, it is a short scale that can be completed in about 5 min by a parent or teacher. The scale was initially written in English but has since been translated into more than 60 languages. Several studies on the original and translated versions have supported the five-factor construction of the scale for children and adolescents (Goodman, 2001; Woerner, Becker, Rothenberger, 2004; Muris, Meesters, van den Berg, 2003). However, Smedje, Broman, Hetta, and von Knorring (1999) found a five-factor solution in boys and girls, but some loadings differed between gender. It should also be noted that two studies supported a three-factor organization of the scale (Van Leeuwen, Meerschaert, Bosmans, De Medts, & Braet, 2006; Dickey & Blumberg, 2004).

Given the numerous advantages of the SDQ, the aim of the present study was to evaluate the construct validity (Cronbach & Meehl, 1955) of the French SDQ in a community sample of adolescents. In a validation study coordinated by Capron in France, the preliminary results showed that the French version of the SDQ was able to discriminate between a clinical and a community sample.
of adolescents (cited in Marzocchi et al., 2004, pp. 43–44). The construct validity of the French SDQ was also explored in a study located in Congo-Kinshasa where teachers completed the SDQ for school children (Kashala, Elgen, Sommerfelt, & Tylleskar, 2005). However, results of an exploratory factor analysis showed problematic loadings for several items and a very low reliability for the Peer problem subscale ($\alpha = .35$). A better reliability was recently reported for this subscale when teachers completed the French SDQ for young adolescents ($\alpha = .64$), but an exploratory factor analysis again showed problematic loadings (Capron, Théond, & Duyme, 2007). This latter result could be explained by the use of exploratory factor analysis that allows one item to load on several factors (Beauducel & Wittman, 2005). With confirmatory factor analysis (CFA), it is possible to use a priori knowledge to test more specific models (Jöreskog, 1969). Two studies have used CFA to test the five-factor structure of the SDQ. Ronning, Handegaard, Sourander, and Morch (2004) found a “somewhat variable and questionable fit” with the Norwegian self-reported version of the SDQ in adolescents (p. 76). Mellor and Stokes (2007) used the Australian self-reported and informant-rated versions of the SDQ and concluded that the model was not supported. These discrepancies could be explained by the use of different versions of the SDQ, but also by the strategy followed to interpret model fit. In the present study, model fit was evaluated with the mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR). These fit indices are recommended because they are less sensitive to small misspecifications of factor structure, which are very common in the domain of personality research (Beauducel & Wittmann, 2005). As indicated in the literature review, the measurement invariance of the SDQ across gender has been challenged. Therefore, the measurement invariance of the SDQ was tested by mean of multigroup CFA. Recently, the French SDQ has been revised to correct inconsistencies with the original English version (Fombonne et al., 2005). This revised French version was used in the present study (available since March 2005 on www.sdqinfo.com). Responses were given on a Likert scale ranging from 0 (not true) to 1 (somewhat true) and 2 (certainly true).

Strengths and Difficulties Questionnaire

The original English SDQ is a 25-item questionnaire that assesses Conduct problems, Hyperactivity/Inattention, Peer problems, Emotional symptoms, and Prosocial behavior (Goodman, 1997). The SDQ is available in two versions: an informant-rated version that can be completed by parents or teachers and a self-report version that can be completed by adolescents (Goodman, Meltzer, & Bailey, 1998). For this study, the informant-rated version was used and completed by the teacher. Recently, the French version of the SDQ was revised to correct inconsistencies with the original English version by means of a back-translation (Fombonne et al., 2005). This revised French version was used in the present study (available since March 2005 on www.sdqinfo.com).

Statistical Analysis

First of all, the group of participants was compared to a large British community sample to see if the behavior of the present sample fell in the normal range. Mean difference between the two samples was estimated by an effect size correlation ($r_{effect size}$) developed by Rosenthal, Rosnow, and Rubin (2000). A correlation of between .10 and .30 (absolute value) corresponds to a small effect, between .30 and .50 to a moderate effect, and above .50 to a large effect. A correlation < .10 (absolute value) is considered to be negligible. Estimates are marked with an asterisk (*) if the 0 was not included in their 95% confidence interval (CI). Equations used to calculate $r_{effect size}$ are reported in the appendix. Cohen’s $\delta$ are reported along with $r_{effect size}$ (Cohen, 1988). These calculations were done with R (R Development Core Team, 2006).

In order to assess the construct validity of the French SDQ and to test its measurement invariance across gender, we computed CFA and multigroup CFA with Mplus 4.2 (Muthén & Muthén, 2006). The data were clustered by teacher in order to control for the variability of their evaluations. We first used the weighted least square estimator (WLSMV), which is suited for the analysis of ordinal variables. However, no pupil received the highest score (2) on Item 22 “Stealing at school, at home, or elsewhere.” As a consequence, the estimation of the polychoric correlations, which are necessary for the WLSMV method, were unreliable between Item 22 and other items. This was because of the absence of observation in cells of the bivariate tables. To get a correct estimation of the polychoric correlations, a larger sample size would be necessary in order to fill in these empty cells. Without a larger sample size, the use of
a maximum likelihood estimator is recommended (see the Mplus discussion on www.statmodel.com). Thus, the default robust maximum likelihood MLR estimator for clustered continuous data was used. In structural equation modeling, the $\chi^2$ is known to increase with sample size and model complexity. The $\chi^2$ is, therefore, inappropriate to confirm the factor organization of questionnaires (Raykov, 1998). For these reasons, the fit of the CFA was tested by examining other indices that depend on a conventional cut-off. Based on simulations, Hu and Bentler (1998) recommend the use of the RMSEA and the SRMR. A RMSEA between 0 and 0.05 indicates a good fit and between 0.05 and 0.08 an acceptable fit. A SRMR between 0 and 0.05 indicates a good fit and between 0.05 and 0.10 an acceptable fit (Schermelleh-Engel & Moosbrugger, 2003).

Several degrees of invariance were tested successively across gender. According to the terminology used by Widaman and Reise (1997), the same pattern of zero and nonzero loadings between groups indicates configural invariance. If in addition, the same factor loadings are observed, this indicates weak factorial invariance. If in addition, the same intercepts for manifest variables are observed, this indicates strong factorial invariance. If in addition, the same residuals for manifest variables are found, this indicates strict factorial invariance. If in addition, the same factor variances are observed, this indicates variance invariance. If in addition, the same factor covariances are found, this indicates covariance invariance. Finally, if the same factor means are also observed, this indicates mean invariance. Configural and factor invariances are related to the psychometric properties of the questionnaire and are desirable. Without these degrees of invariance, at least weak factorial invariance, group differences in factors variance, covariance, and mean are interpretable (Widaman & Reise, 1997). It should be noted that the use of $\chi^2$ difference to test measurement invariance is problematic because it depends on sample size and model complexity (Kelloway, 1995). Based on simulations, Cheung and Rensvold (2002) recommended the use of the RMSEA to test configural invariance, and the change in CFI to test all other degrees of invariance. A cut-off is given for each degree of invariance. For configural invariance, a RMSEA> 0.093 indicates that the invariance hypothesis should be rejected at $\alpha = .01$ (Cheung & Rensvold, 2002, Table 4). For the other degrees of invariance, a CFI decrease > 0.009 (weak factor invariance), >0.008 (strong factor invariance), > 0.009 (strict factor invariance), > 0.006 (variance invariance), > 0.005 (covariance invariance), and > 0.006 (mean invariance) indicates that the hypothesis of measurement invariance should be rejected at $\alpha = .01$ (Cheung & Rensvold, 2002, Table 5).

### Results

#### Preliminary Analysis

The SDQ was completed for 560 pupils. Three pupils whom their teachers had known for less than 6 months were removed from the sample. In the remaining 557 questionnaires, 384 had no missing values, 59 had one missing value, 40 had two, and 74 had three or more missing values. It appeared that missing values were particularly frequent for Items 17 and 22. Responses to Item 17 were missing from 15.44% of the questionnaires and responses to Item 22 from 27.83% of them (< 10% on all other items).

In order to calculate sample means, missing values were imputed with the mean obtained by the subject on the subscale to which the missing value belonged. This imputation method is reasonable when subscales are reliable (Schafner & Graham, 2002, p. 158; see below for the reliabilities of the subscales). Means and SDs of the SDQ are presented in Table 1. These results were compared to the teacher SDQ norms reported for a British national survey of 3,407 adolescents aged from 11 to 15 years old (Meltzer, Gatward, Goodman, & Ford, 2000). The mean difference between the Swiss and British samples for Conduct problems corresponded to a negligible effect, $r_{\text{effect size}} = .05^*$. This was also the case for Hyperactivity/Inattention, $r_{\text{effect size}} = .00$, Peer problems, $r_{\text{effect size}} = .07^*$, Emotional symptoms, $r_{\text{effect size}} = .09^*$, and Prosocial behavior, $r_{\text{effect size}} = -.03$. Thus, even though some of the differences between the Swiss and the British sample were significant, they were all negligible in terms of effect size. In sum, we can consider that the observed behavior in the Swiss sample fell within the range of the normative British sample.

#### CFA for Boys and Girls

For the CFA and multiple group CFA, missing values were not imputed and were handled with the maximum likelihood procedure available in Mplus (Schafer & Graham, 2002). The original five-factor model of the SDQ was first

| Table 1. Comparison of Swiss (n = 557) and British (n = 3407) samples on the SDQ |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| SDQ subscales               | Swiss mean (SD)             | British mean (SD)           | $r_{\text{effect size}}$ (95% CI) | Cohen’s $\delta$          |
| 1. Conduct problems         | 1.14 (1.77)                 | 0.9 (1.7)                   | .05* (.02 .08)               | 0.14                        |
| 2. Hyperactivity/Inattention| 2.61 (2.47)                 | 2.6 (2.7)                   | .00 (–.03 .03)               | 0.00                        |
| 3. Peer problems            | 1.75 (1.84)                 | 1.4 (1.8)                   | .07* (.04 .10)               | 0.19                        |
| 4. Emotional symptoms       | 1.81 (2.00)                 | 1.3 (1.9)                   | .09* (.06 .12)               | 0.27                        |
| 5. Prosocial behavior       | 6.92 (2.67)                 | 7.1 (2.4)                   | –.03 (–.06 .01)              | –0.07                       |

*Note. Data from the British sample are taken from Meltzer, Gatward, Goodman, & Ford (2000). *0 not included in the 95% confidence interval.*
tested separately for girls and boys. Five latent factors were defined by their respective items and were allowed to correlate. The model had a RMSEA of 0.089 that fell above the 0.080 cut-off and a SRMR of 0.117 that fell above the 0.100 cut-off. These results suggested that the model did not fit the data well (Table 2, Model 1). A very high modification index was found between Items 2 and 10, suggesting that freeing the covariance between residuals of the 2 items would improve the fit. Item 2 is “Restless, overactive, cannot stay still for long” and was translated into “Agité(e), hyperactif(ve), ne tient pas en place.” Item 10 is “Constantly fidgeting or squirming” and was translated into “Ne tient pas en place ou se tortille constamment.” In the previous version of the French SDQ, “Fidgeting” of Item 10 was translated into “A la bougeotte.” However, this expression had to be explained to many of the young adolescents filling in the self-reported version of the scale (C. Capron, personal communication). Therefore, the more familiar expression “Ne tient pas en place” was preferred. It follows that in the revised French SDQ, “Fidgeting” of Item 10 was translated into “Ne tient pas en place ou se tortille constamment.” In the previous version of the French SDQ, “Fidgeting” of Item 10 was translated into “A la bougeotte.” However, this expression had to be explained to many of the young adolescents filling in the self-reported version of the scale (C. Capron, personal communication). Therefore, the more familiar expression “Ne tient pas en place” was preferred. It follows that in the revised French SDQ, the expression “Ne tient pas en place” appears in Items 2 and 10. Thus, it is justified to let the residuals of the two items covary. The modified model had a RMSEA of 0.078 and a SRMR of 0.099, which indicated that the modified model was acceptable (Model 2). Thus, the five-factor model of the SDQ was confirmed in girls. For boys, results showed a RMSEA of 0.083 and a SRMR of 0.101, suggesting that the model did not fit the data very well (Model 3). A very high modification index was also found between Items 2 and 10. As in the case of girls, we let the two residuals covary. Results showed a RMSEA of 0.074 combined with a SRMR of 0.092, suggesting an acceptable fit (Model 4). Thus, the five-factor model of the SDQ was confirmed in boys.

Multigroup CFA on the Whole Sample

In order to test configurual invariance, we computed a two-group CFA with no constraint concerning the equality of loadings, intercepts, and residuals between boys and girls. Factor variances were set to 1 and factor means were set to 0 in both groups (for testing measurement invariance with Mplus, see Muthén & Muthén, 2006: p. 345). We let the residuals of Items 2 and 10 covary, but we put no constraint on their equality between groups. The RMSEA was 0.076 and fell below the 0.093 cut-off, suggesting configural invariance (Model 5). We then added a constraint on the equality of loadings. The decrease in CFI was 0.004 and fell below the 0.009 cut-off, indicating weak factor invariance (Model 6). The next step was to add a constraint on the equality of intercepts. The decrease in CFI was 0.004 and fell below the 0.008 cut-off, indicating strong factor invariance (Model 7). We next added a constraint on the equality of item residuals. The covariance between residuals of Items 2 and 10 was also equated be-

---

**Table 2.** Fit indices of the CFA for girls, boys, and the whole sample

<table>
<thead>
<tr>
<th>Model</th>
<th>Type</th>
<th>df</th>
<th>RMSEA</th>
<th>SRMR</th>
<th>Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>Girls</td>
<td>265</td>
<td>0.089</td>
<td>0.117</td>
<td>Problematic</td>
</tr>
<tr>
<td>Model 2</td>
<td>Girls + error cov.</td>
<td>264</td>
<td>0.078</td>
<td>0.099</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Model 3</td>
<td>Boys</td>
<td>265</td>
<td>0.083</td>
<td>0.101</td>
<td>Problematic</td>
</tr>
<tr>
<td>Model 4</td>
<td>Boys + error cov.</td>
<td>264</td>
<td>0.074</td>
<td>0.092</td>
<td>Acceptable</td>
</tr>
<tr>
<td>Model 13</td>
<td>Whole sample</td>
<td>265</td>
<td>0.074</td>
<td>0.100</td>
<td>Problematic</td>
</tr>
<tr>
<td>Model 14</td>
<td>Whole sample + error cov.</td>
<td>264</td>
<td>0.063</td>
<td>0.087</td>
<td>Acceptable</td>
</tr>
</tbody>
</table>

**Table 3.** Fit indices and invariance between gender for the multigroup CFA

<table>
<thead>
<tr>
<th>Model</th>
<th>Invariance type</th>
<th>df</th>
<th>RMSEA (cut-off)</th>
<th>CFI</th>
<th>CFI decrease (cut-off)</th>
<th>Invariance hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 5</td>
<td>Configural</td>
<td>528</td>
<td>0.076 (0.093)</td>
<td>0.811</td>
<td>–</td>
<td>Accepted</td>
</tr>
<tr>
<td>Model 6</td>
<td>Weak factorial</td>
<td>548</td>
<td>0.076</td>
<td>0.807</td>
<td>0.004 (0.009)</td>
<td>Accepted</td>
</tr>
<tr>
<td>Model 7</td>
<td>Strong factorial</td>
<td>568</td>
<td>0.075</td>
<td>0.803</td>
<td>0.004 (0.008)</td>
<td>Accepted</td>
</tr>
<tr>
<td>Model 8</td>
<td>Strict factorial</td>
<td>594</td>
<td>0.071</td>
<td>0.816</td>
<td>–0.013 (0.009)</td>
<td>Accepted</td>
</tr>
<tr>
<td>Model 9</td>
<td>Variance</td>
<td>599</td>
<td>0.071</td>
<td>0.814</td>
<td>0.002 (0.006)</td>
<td>Accepted</td>
</tr>
<tr>
<td>Model 10</td>
<td>Covariance</td>
<td>609</td>
<td>0.071</td>
<td>0.811</td>
<td>0.003 (0.005)</td>
<td>Accepted</td>
</tr>
<tr>
<td>Model 11</td>
<td>Mean</td>
<td>614</td>
<td>0.072</td>
<td>0.803</td>
<td>0.008 (0.006)</td>
<td>Accepted</td>
</tr>
<tr>
<td>Model 12</td>
<td>On age</td>
<td>649</td>
<td>0.071</td>
<td>0.802</td>
<td>–</td>
<td>Accepted</td>
</tr>
</tbody>
</table>

Note. Error cov. = Errors of item 2 and 10 are allowed to covary. RMSEA < 0.050 and SRMR < 0.050 indicate a good fit. RMSEA < 0.080 and SRMR < 0.100 indicate an acceptable fit (Schermelleh-Engel & Moosbrugger, 2003).
tween groups. In addition, factor means were set to 0 for girls and made free for boys. The CFI increased by 0.013 and, thus, fell below the cut-off, showing strict factor invariance (Model 8). It should be noted that CFI can increase when constraints are added to a model (see Cheung & Rensvold, 2002). Together, these results indicate that the SDQ has similar psychometric properties in boys and girls.

The next step was to add a constraint on the equality of factor variances. The increase in CFI was 0.002 and fell below the cut-off, indicating comparable factors variance across gender (Model 9). Then we added a constraint on the equality of the factor covariances. The decrease in CFI was 0.003 and fell below the cut-off, indicating comparable relationships between factors for boys and girls (Model 10). Finally, we added constraint on the equality factor means. The decrease in CFI was 0.008 and fell above the 0.006 cut-off, indicating that the means differed between boys and girls (Model 11). Because strict factorial invariance was found, it was meaningful to compare the factor means of boys in Model 10. The means for boys were .29*, CI = (.14, .43) on Conduct problems, .36*, CI = (.24, .47) on Hyperactivity/Inattention, .31*, CI = (.15, .46) on Peer problems, −.17, CI = (−.38, .03) on Emotional symptoms, and −.65*, CI = (−.87, −.42) on Prosocial behavior (standardized solution). Thus, boys had higher scores for Conduct problems, Hyperactivity/Inattention, and Peer problems (moderate effect size), while girls were more Prosocial (large effect size). Then age was added to Model 10 as an independent variable (Model 12). For girls, Peer problems increased with age, .23*, CI = (.04, .42) and Prosocial behavior deceased with age, −.25*, CI = (−.49, −.02) (standardized solution). No significant effect of age was found for boys (Table 4).

<table>
<thead>
<tr>
<th>Table 4. Factors regressed on age for girls and boys for the multigroup CFA (model 12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latent variables</td>
</tr>
<tr>
<td>1. Conduct problems</td>
</tr>
<tr>
<td>2. Hyperactivity/Inattention</td>
</tr>
<tr>
<td>3. Peer problems</td>
</tr>
<tr>
<td>4. Emotional symptoms</td>
</tr>
<tr>
<td>5. Prosocial behavior</td>
</tr>
</tbody>
</table>

Note. 95% CI given in parentheses. Standardized solution. *0 not included in the 95% CI

<table>
<thead>
<tr>
<th>Table 5. Correlations between factors and reliability of the CFA for the whole sample (model 14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Conduct</td>
</tr>
<tr>
<td>2. Hyper</td>
</tr>
<tr>
<td>3. Peer</td>
</tr>
<tr>
<td>4. Emotional</td>
</tr>
<tr>
<td>5. Prosocial</td>
</tr>
</tbody>
</table>

Note. Reliability given in brackets. 95% CI given in parentheses. *0 not included in the 95% CI.

CFA for the Whole Sample

Because the five-factor model of the SDQ was tested and accepted separately for girls and boys, it is reasonable to merge their data. The five latent factors were defined by their respective items and were allowed to correlate. Results showed a RMSEA of 0.074 and a SRMR of 0.100. Thus, SRMR reached the 1.00 cut-off, suggesting that the
model did not fit the data very well (Table 2, Model 13). A very high modification index was found between Items 2 and 10. As a consequence, we let their residuals covary. The RMSEA of 0.063 combined with the SRMR of 0.087 indicated an acceptable fit (Model 14). Thus, the five-factor model of the SDQ was confirmed for the whole sample. Internal reliability of the latent factors was calculated based on the estimated model (for the formula, see Raines-Eudy, 2000: p. 126). It equals .77 for Conduct problems, .83 for Hyperactivity/Inattention, .64 for Peer problems, .78 for Emotional symptoms, and .90 for Prosocial behavior (diagonal of Table 5). Thus, the Peer problems subscale had acceptable reliability (> .60) and the other subscales had good (> .70) or very good reliability (> .80). Inter correlations between factors showed that the four problematic subscales were positively correlated to each other, but negatively related to the Prosocial subscale (Table 5). All standardized loadings were moderately high (> .30), except for Item 23, which had a lower standardized loading of 0.25* (Table 6). Item 23 had also a relatively high residual and was, thus, partly responsible for the lower reliability of the Peer problem factor.

Discussion

The aim of this study was to test the construct validity of the revised French SDQ (Fombonne et al., 2005) by mean of a CFA. Preliminary analyses revealed that missing values were more frequent on Items 17 and 22. The presence of missing values suggests that these two items are more difficult for teachers to complete. Item 17 is “Kind to younger children” and indeed it is rare that teachers see their pupils interacting with younger children. Similarly, Item 22 is “Stealing at school, at home, or elsewhere” and indeed it is rare that teachers see their pupils interacting with younger children. In this regard, it would be valuable to ask parents to complete the SDQ in addition to teachers in order to get an evaluation from outside the school.

Previous studies of the psychometric properties of the SDQ have not tested measurement invariance across gender. In the present study, multigroup CFA for the whole sample revealed strict factorial invariance and this constitutes an asset of the French SDQ. Indeed, this measurement invariance makes it possible to compare the variances/covariances of the factors as well as their means. No significant difference was found for the factor variances and covariances across gender. However, the factor means showed that teachers reported more Conduct problems, Hyperactivity/Inattention, and Peer problems for boys. Girls were evaluated as being more Prosocial. Further studies are needed to replicate this measurement invariance using the English or other versions of the SDQ.

The CFA for boys, girls, and the whole sample showed that the original five-factor model of the SDQ had an acceptable fit (after residuals of two items containing the same French expression were allowed to covary). The internal reliability of the subscales was good or very good, except for Peer problems, which had an acceptable reliability. A high correlation was found between Conduct problems and Hyperactivity/Inattention. Similarly, Goodman (2001) reported a high correlation between these two subscales when the SDQ was completed by teachers. These results are in accordance with studies showing that conduct disorder and attention deficit with hyperactivity disorder are often comorbid and share common risk factors (Patterson, DeGarmo, & Knutson, 2000).

It should be noted that our study has several limitations. First of all, our sample size is moderate. In the present study, we used a maximum likelihood estimator to fit models. However, our results suggest that a larger sample is necessary to allow the use of an estimator especially developed for ordinal variables. Larger samples are also needed to provide normative data for the French SDQ. Furthermore, interrater and test-retest reliability were not tested. This could be achieved by asking parents and teachers to complete the French SDQ at several points in time. In addition, the convergent validity of the French version has not been verified. To do this, it will be necessary to use the SDQ along with scales assessing the same types of behavior, such as the Child Behavior Checklist (Achenbach, 1991). Despite these limitations, it can be concluded that the French SDQ offers a reliable factor organization and this should encourage researchers and clinicians to use it.

Acknowledgments

We wish to thank Amanda Bernabeu, Michelle Bishoff, Laura Bless, Mena Dell’Armin, Cécile Grangier, Catherine Rouge, Claudia Salcuni, and Claudia Szalai for their help in collecting the data.

References


Mathieu d’Acremont
Swiss Centre for Affective Sciences
University of Geneva
Rue des Battoirs 7
CH-1205 Geneva
Switzerland
Tel. +41 22 379-9820
Fax +41 22 379-9844
E-mail mathieu.dacremont@pse.unige.ch
Appendix

Equations to calculate the effect size correlation ($r_{\text{effect size}}$) for the mean difference between two groups of unequal sample size (Rosenthal, Rosnow, & Rubin, 2000):

\[ S_{\text{within}}^2 = \left( \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{df_{\text{within}}} \right); \]  
[Equation 2.2]

\[ t = \frac{(M_1 - M_2)}{S_{\text{within}} \cdot N^{1/2} / 2 \cdot (n_h / n_{\tilde{h}})^{1/2}}; \]  
[Equation 2.20]

\[ r = \left[ \frac{t^2}{(t^2 + df_{\text{within}})} \right]^{1/2}. \]  
[Equation 2.3]

with:

- $n_1, M_1, S_1$ the size, mean, and $SD$ of the Swiss sample;
- $n_2, M_2, S_2$ the size, mean, and $SD$ of the British sample;
- $N = n_1 + n_2$; $df_{\text{within}} = N - 2$;
- $n_{\tilde{h}} = N / 2$; $n_h = 1 / [1 / 2 * (1 / n_1 + 1 / n_2)]$.

Note that $r$ given by equation 2.3 is always positive. To reflect the direction of the mean difference, we calculated:

\[ r_{\text{effect size}} = -r, \text{ if } t < 0; \]
\[ r_{\text{effect size}} = r, \text{ otherwise.} \]

The confidence interval was calculated with Fisher’s “z to r” transformation (Fisher, 1915).