The Amsterdam Executive Function Inventory (AEFI): Psychometric properties and demographically corrected normative data for adolescents aged between 15 and 18 years

Wim Van der Elst¹, Carolijn Ouwehand², Greetje van der Werf³, Hans Kuyper³, Nikki Lee², and Jelle Jolles²

¹Faculty of Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands
²LEARN! Research Institute and Faculty of Psychology & Education, VU Universiteit Amsterdam, Amsterdam, The Netherlands
³GION, Institute for Educational Research, University of Groningen, Groningen, The Netherlands

The Amsterdam Executive Function Inventory (AEFI) is a newly developed brief self-report questionnaire to assess three important components of the executive aspects of daily-life behavior—that is, Attention, Self-Control and Self-Monitoring, and Planning and Initiative. In a population-based study, the AEFI was administered to \( N = 6,730 \) Dutch adolescents aged between 15 and 18 years. Psychometric analyses showed that the construct validity and the reliability of the AEFI were adequate. Educational level and gender affected the different AEFI scale scores. Regression-based normative data that took these demographic influences into account were established, so that the AEFI can be used by clinicians and researchers who need to assess executive aspects of daily-life behavior.

Keywords: Executive functioning; Adolescents; Norms; Psychometric analyses.

“Executive functions” is an umbrella term that refers to the cognitive abilities that are needed for goal-directed behavior (Lezak, Howieson, & Loring, 2004). Various authors have argued that the assessment of executive functions is difficult (Anderson, Anderson, Northam, Jacobs, & Catroppa, 2001; Jurado & Rosselli, 2007; Klenberg, Korkman, & Lahti-Nuuttila, 2001; Lezak et al., 2004; Saboya, Coutinho, Segenreich, Ayrão, & Mattos, 2009; Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2008). For example, research has shown that “traditional” performance-based executive function tests—such as the Stroop Color Word Test (Stroop, 1935)—are stronger associated with cognitive abilities (e.g., intelligence) than with executive aspects of daily-life behavior (e.g., inattention; Biederman et al., 2008). Moreover, performance on most traditional executive function tests involves multiple lower level cognitive abilities (such as processing speed and expressive language) in addition to higher level executive components (Anderson et al., 2001). This hampers a straightforward interpretation of the test results in terms of executive functioning. Most traditional executive function tests were also developed in the field of adult (neuro)psychology, and not all of these tests can be used to assess the executive functions of children and adolescents. For example, assessing
the inhibition component of executive functioning with the Stroop Color–Word Test is only possible when reading has become a highly automatized behavior in an individual.

More recently, neuropsychological tests have been developed that were explicitly intended to maximize the prediction of real-world behavior (the so-called “verisimilitude” tests, such as the Behavioral Assessment of the Dysexecutive Syndrome for Children, BADS-C; Emslie, Wilson, Burden, Nimmo-Smith, & Wilson, 2003). The empirical evidence that verisimilitude tests are more predictive for daily-life functioning than traditional executive function tests is, however, inconsistent and rather limited (Chaytor & Schmitter-Edgecombe, 2003; Wood & Liossi, 2006). Moreover, some of the problems that arise with traditional executive function tests also occur with verisimilitude tests. For example, the Water subtest of the BADS-C requires children to remove a cork from a tube by using equipment that is provided to introduce water into the tube to float the cork. This subtest requires a complex sequence of behaviors that consists of both executive (e.g., planning) and nonexecutive (e.g., visuomotor coordination skills) components. The unequivocal interpretation of the test results in terms of executive abilities is thus hampered (Henry & Bettenay, 2010). Another problem with verisimilitude tests such as the BADS-C is that these tests are typically intended to be administered only once, as they are based on the philosophy that executive functions should be assessed by novel tasks. Research has indeed shown that the test–retest correlations for various BADS-C subtests were low (Henry & Bettenay, 2010). As a result, instruments such as the BADS-C cannot be used to address longitudinal research questions (e.g., how do the executive functions of children develop over time?) or questions with respect to clinical change (e.g., did the executive abilities of this child improve after initiating a treatment?).

In view of these problems with behavior-based tests, it has been suggested that questionnaires may provide a useful complementary approach to assess executive functions (Chaytor, Schmitter-Edgecombe, & Burr, 2006; Isquith, Crawford, Espy, & Gioia, 2005; Jurado & Rosselli, 2007). Research has indeed indicated that questionnaire-based measures were more strongly related to the executive aspects of daily-life behavior than to behavior-based executive function measures (Biederman et al., 2008). Moreover, questionnaires do not confound lower and higher level cognitive abilities, and they can be repeatedly administered in the same individual. Questionnaires also have the advantage that they can be used in both children and adults, by relying on self-report, informant report, or a combination of both.

Existing executive function questionnaires, such as the Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000), have good psychometric properties (e.g., Huizinga & Smidts, 2011), but these instruments typically contain a large number of items (e.g., 86 items for the BRIEF), which makes their administration time relatively long. It would be advantageous to have a short psychometrically sound screening instrument available that assesses the executive functions of children and adolescents in a straightforward and time-efficient way. The aim of the present study was to present such an instrument—that is, the Amsterdam Executive Function Inventory (AEFI).

In the AEFI, executive functions were conceptualized as comprising three separable—but integrated—factors (see Anderson et al., 2001). The first executive component can be verbally labeled “Attention” and consists of cognitive abilities such as selective and sustained attention (measured in the AEFI by items such as “I am not able to focus on the same topic for a long period of time” and “My thoughts easily wander”). The second executive function factor can be verbally labeled “Self-Control and Self-Monitoring” and consists of abilities such as working memory and self-monitoring (measured by items such as “I often lose things” and “I often react too fast. I’ve done or said something before it is my turn”). The third factor can be verbally labeled “Planning and Initiative” and consists of abilities such as the initiating and planning of behavior (measured in the AEFI by items such as “I can make fast decisions, e.g., in lessons” and “I am well organized. For example, I am good at planning what I need to do during a day”).

In the present study, the psychometric properties of the AEFI were evaluated as based on the data of a large sample (N = 6,730) of adolescents (aged between 15.0 and 18.0 years). Normative data were also established, because it is difficult to interpret raw test scores in a meaningful way without an empirical frame of reference (Capitani, 1997; Van der Elst, 2006). Norms are indeed useful tools, which allow clinicians and researchers to address questions such as: “Are the Attention, Self-Control and Self-Monitoring, and Planning and Initiative AEFI scale scores of this child “normal” as compared to the scores of his or her demographically matched peers, or are these scores deviant (i.e., exceptionally low or high)?” Previous studies in children and early adolescents (aged below 15 years) have shown that age (Ardila,
Rosselli, Matute, & Guajardo, 2005; Blakemore & Choudhury, 2006; Klenberg et al., 2001; Levin et al., 1991; Rebok, Smith, Pascualvaca, Mirsky, Anthony, & Kellam, 1997; Welsh, Pennington, & Groisser, 1991), gender (Anderson et al., 2001; Ardila et al., 2005; Berlin & Bohlin, 2002; Brocki & Bohlin, 2004), and educational level (Ardila et al., 2005; Klenberg et al., 2001; Noble, Norman, & Farah, 2005) affect executive functions. It is conceivable that similar demographic effects occur among older adolescents who are aged above 15 years. Thus we also evaluated the impact of age, gender, and educational level on the AEFI scale scores, so that the normative data could be appropriately corrected for the relevant demographic variables.

**METHOD**

**Participants**

Data were derived from the COOL study (Cohort Onderzoek Onderwijs Loopbanen; in English: Cohort Study of Educational Trajectories), a large-scale study into the determinants of the cognitive and social–emotional development of children and adolescents (Kuyper, Keuning, & Zijsling, 2010). A total of $N = 7,793$ participants (from $N = 75$ schools) participated in the COOL study and filled in a questionnaire, which included the AEFI. Respondents were excluded from the data analyses if they had missing values on one or more items of the AEFI ($n = 983$). The age range of the remaining $N = 6,810$ participants was between 14.1 and 20.4 years, but the majority of participants were aged between 15.0 and 18.0 years ($n = 6,730$; 98.8%). The $n = 80$ participants who were younger than 15.0 years or older than 18.0 years were excluded from the analyses to avoid a disproportionate influence of the outlying predictor values on the fitted regression model (Kutner, Nachtsheim, Neter, & Li, 2005).

The demographic characteristics of the sample are shown in Table 1. The mean (standard deviation) age of the children who were included in the analyses was 16.0 ($SD = 0.6$) years. Level of education (LE) was measured by classifying the formal schooling of the participants into one of three groups—that is, those with prevocational secondary education (LE low; 46.3% of the sample), those with higher general secondary education (LE average; 27.3% of the sample), and those with preuniversity education (LE high; 26.4% of the sample). There were slightly more female than male participants (51.0% females).

**Procedure and instruments**

The initial version of the AEFI contained 14 items, which represented three a priori expected dimensions of executive functioning—that is, Attention, Self-Control and Self-Monitoring, and Planning and Initiative (see Introduction). The responses for the AEFI items were presented on a 3-point Likert scale with the choice options 1 = “not true,” 2 = “partly true,” and 3 = “true.” Items 1, 4, 5, 6, 7, 8, 11, 12, and 13 were reverse coded, so that higher scores were indicative of better executive abilities.

**Statistical analyses**

Confirmatory factor analysis (CFA) was used to evaluate the factor structure of the AEFI. An a priori model was specified, in which three factors underlay the item responses and in which each item loaded on only one factor. Due to the categorical nature of the item responses and the non-normal score distributions of a number of items, the diagonally weighted least squares method for polychoric correlation matrices was used (instead of the standard maximum likelihood estimation method). Polychoric coefficients and an asymptotic covariance matrix were generated in PRELIS (Jöreskog & Sörbom, 1996). Prior to the CFA, pilot testing of the items to be factor-analyzed was conducted to ensure that the items that were designed to measure a common construct correlated at least moderately with each other (i.e., correlations of at least .20; see Floyd & Widaman, 1995). Kendall tau-b correlation coefficients were calculated (rather than Pearson correlation coefficients) in view of the ordinal measurement level of the items of the AEFI.

The fit of the three-factor AEFI model was evaluated with the root mean square error of approximation (RMSEA; <.08 acceptable, <.05 excellent; Browne & Cudeck, 1993), the comparative fit index (CFI; >.90 acceptable, >.95 excellent; Bentler, 1990; Bentler & Bonett, 1980), and the normed fit index (NFI; >.90 acceptable; Bentler & Bonnett, 1980). Item descriptives (means, $SD$s, and item response distributions) were calculated. The internal consistency and reliability of the established AEFI scales were estimated with Cronbach’s alpha coefficients (which should be $\geq 6$–7; Clark & Watson, 1995; Dekovic, Janssens, & Gerris, 1991; Holden, Fecken, & Cotton, 1991) and with corrected item–scale correlations (i.e., the correlations between items and scale scores that did not include
the items being evaluated, which should be ≥.30; Ferketich, 1991). For shorter scales, the corrected item–scale values provide a better index of internal consistency and reliability than Cronbach’s alpha, because Cronbach’s alpha values are not only a function of the height of the intercorrelations between the items of a scale, but also a function of the number of items on that scale (Clark & Watson, 1995).

The effects of demographic variables on the established AEFI scale scores were evaluated by regressing the scale scores on age, age2, sex, educational level, and all two-way interactions between these variables. As noted above, the participants were sampled from N = 75 different schools. This introduces a hierarchical structure in the dataset (i.e., children are clustered within schools), and thus the necessity to use hierarchical rather than single-level statistical models was examined. This was done by computing design effects—that is, 1 + [(average cluster size – 1) × intraclass correlation]—which should be below 2 in order to appropriately use single-level models (such as multiple linear regression models) rather than hierarchical models (such as linear mixed models; Muthén & Satorra, 1995). In all statistical models, age was centered (age = calendar age – 16.0) before quadratic terms and interactions were calculated to avoid multicollinearity (Kutner et al., 2005). Sex was coded as male = 1 and female = 0. Level of education was dummy coded with two dummies (LE average and LE high) and LE low as the reference category. Nonsignificant predictors (p > .01) were excluded from the full models (i.e., the models that included the main effects and all two-way interactions), but no predictor was removed as long as it was also included in a higher order term in the model (Aiken & West, 1991). The assumptions of regression analysis were tested for each model. Homoscedasticity was evaluated by grouping the participants into quartiles of the predicted scores and applying the Levene test. Normality of the residuals was investigated by visual inspection of the histograms of the residuals. The occurrence of multicollinearity was checked by calculating variance inflation factors (VIFs), which should not exceed 10 (Belsley, Kuh, & Welsch, 1980). Potential influential cases were identified by calculating Cook’s distances.

The AEFI scale scores were normed by means of a four-step procedure (Van Breukelen & Vlaeyen, 2005; Van der Elst, 2006; Van der Elst, Van Bokxel, Van Breukelen, & Jolles, 2006a, 2006b). In this procedure, the user of the normative data calculates first the testee’s predicted AEFI scale scores by means of the final regression models: predicted AEFI scale score = \( B_0 + B_1X_1 + \ldots + B_nX_n \), with \( B_0 \) = the intercept, \( B_n \) = the regression weight(s), and \( X_n \) = the predictor value(s). Second, the residuals are calculated: \( e_i = \text{observed AEFI scale score} – \text{predicted AEFI scale score} \). Third, the residuals are standardized: \( Z_i = e_i / SD(\text{residual}) \), with \( SD(\text{residual}) \) = the standard deviation of the residuals in the normative sample. Fourth, the standardized residuals are converted into percentiles via the standard normal cumulative distribution function (if the model assumption of normality of the residuals was met in the normative sample), or via the empirical cumulative distribution function of the standardized residuals (if the standardized residuals were not normally distributed in the normative sample).

All analyses were conducted with R 2.10.1 for Linux (R Development Core Team, 2010) and PRELIS/LISREL 8.8 for Windows (Jöreskog & Sörbom, 1996). An alpha level of .01 was used in all analyses.

### RESULTS

#### Factor structure and psychometric properties of the AEFI

Prior to conducting the CFAs, pilot testing of the items to be factor-analyzed was conducted. The Kendall tau-b correlation coefficients between Item 4 (“I think it is difficult to do multiple things at the...
same time, e.g., studying and listening to music”) and the other items that were designed to measure the Attention construct were below .20. After exclusion of Item 4, all Kendall tau-b intercorrelations of the items that were designed to measure Attention were above .20. The mean Kendall tau-b intercorrelation coefficients between the items that were designed to measure Self-Control and Self-Monitoring and between the items that were designed to measure Planning and Initiative were also all above .20.

The three-factor CFA model is shown in Table 2. The model adequately fitted the data (i.e., RMSEA = .06; CFI = .95; NFI = .95), and all items loaded significantly on the a priori expected factors. The average standardized factor loadings of the items of the Attention, Self-Control and Self-Monitoring, and Planning and Initiative scales equaled .69, .60, and .54, respectively. Internal consistency and reliability were adequate (Cronbach’s alpha values equaled .64, .65, and .60 for the Attention, Self-Control and Self-Monitoring, and Planning and Initiative scales, respectively, and all corrected item–scale correlations were ≥.30, all ps < .01; see Table 2). The correlations between the AEFI scale scores were all significantly positive—that is, r(Attention, Self-Control and Self-Monitoring) = .58, r(Attention, Planning and Initiative) = .15, r(Self-Control and Self-Monitoring, Planning and Initiative) = .10 (all ps < .01).

The high correlation between the Attention and the Self-Control and Self-Monitoring scale scores may be indicative (a) that a single higher order factor underlies the two first-order factors, or (b) that the items of the Attention and the Self-Control and Self-Monitoring scales load on a single first-order factor (rather than on two separate first-order factors). Additional analyses showed that this was not the case—that is, the fit of these two alternative models was worse than the fit of the a priori expected model (the p-values of the scaled χ² difference tests; Satorra & Bentler, 2001).

Note also that the item–scale correlations of the items that were part of the scale at hand (as identified with the three-factor CFA; see Table 2) were substantially higher than the item–scale correlations of items that were not part of the scale. Indeed, the average item–scale correlations for the items of the Attention, Self-Control and Self-Monitoring, and Planning and Initiative scales were .68, .55, and .54, respectively, whilst the average correlations between the items that were not part of these scales and the Attention, Self-Control and Self-Monitoring, and Planning and Initiative scale scores were only .20, .17, and .08, respectively.

The effects of age, sex, and level of education on the AEFI scale scores

The design effects were all below 2 (i.e., 1.02, 1.03, and 1.02 for the Attention, Self-Control and Self-Monitoring, and Planning and Initiative outcomes, respectively), so multiple linear regression models were used to analyze the data rather than linear mixed models. The final multiple linear regression models for the AEFI scale scores are presented in Table 3. There was no serious influence of outliers (the maximum Cook’s distance value equaled .002) or multicollinearity (maximum VIF < 2.730) observed for any of these models. The Levene test suggested that there was no heteroscedasticity for any of the models (all ps > .01). The standardized residuals were not normally distributed for the AEFI scale scores. Thus, empirical cumulative distributions of the standardized residuals were used to convert the standardized residuals into percentile values.

A significant sex by LE interaction effect was found for the Attention scale score. As shown in Figure 1a, the interaction suggested that self-reported Attention in lower educated boys was higher than that for girls, but the sex difference decreased as a function of educational level. Higher levels of the Self-Control and Self-Monitoring score were reported by higher educated adolescents (see Figure 1b). Being male and having a higher educational level was associated with higher Planning and Initiative scale scores (see Figure 1c). None of the AEFI scale scores were significantly affected by age (or age²).

Regression-based normative data

Norms for the AEFI scale scores were established by means of the four-step procedure described above. For example, suppose that a 16-year-old low-educated boy obtained a raw Attention scale score of 4. The user of the regression-based norms first calculates the expected Attention scale score for this person—that is, 6.105 [= 5.795 + (1 × 0.310) + (0 × 0.031) + (0 × 0.517) + (0 × −0.108) + (0 × −0.366)]. Secondly, the residual is calculated—that is, −2.105 (= 4 – 6.105). Thirdly, the residual is standardized—that is, −1.339 (= −2.105/1.572). Fourthly, the standardized residual is converted into a percentile value by means of the empirical cumulative distribution of the standardized residuals. A standardized residual that equals −1.339 corresponds with a percentile value of 8. Thus, 8% of the population of 16-year-old low-educated boys obtain an Attention scale score that
<table>
<thead>
<tr>
<th>AEFI item</th>
<th>Attention</th>
<th>Planning and Initiative</th>
<th>Self-Control and Self-Monitoring</th>
<th>Item descriptives</th>
<th>% responses in each category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I am not able to focus on the same topic for a long period of time</td>
<td>.51</td>
<td>.33*</td>
<td>2.03 (0.64)</td>
<td>0.18 0.60 0.22</td>
</tr>
<tr>
<td>5</td>
<td>I am easily distracted</td>
<td>.76</td>
<td>.53*</td>
<td>1.99 (0.72)</td>
<td>0.27 0.48 0.25</td>
</tr>
<tr>
<td>7</td>
<td>My thoughts easily wander</td>
<td>.80</td>
<td>.51*</td>
<td>2.00 (0.71)</td>
<td>0.25 0.50 0.25</td>
</tr>
<tr>
<td>6</td>
<td>I often react too fast. I’ve done or said something before it is my turn</td>
<td>.53</td>
<td>.35*</td>
<td>2.11 (0.70)</td>
<td>0.20 0.49 0.31</td>
</tr>
<tr>
<td>8</td>
<td>It is difficult for me to sit still</td>
<td>.66</td>
<td>.42*</td>
<td>2.13 (0.77)</td>
<td>0.25 0.38 0.37</td>
</tr>
<tr>
<td>11</td>
<td>It takes a lot of effort for me to remember things</td>
<td>.60</td>
<td>.41*</td>
<td>2.39 (0.66)</td>
<td>0.10 0.42 0.48</td>
</tr>
<tr>
<td>12</td>
<td>I often forget what I have done yesterday</td>
<td>.58</td>
<td>.39*</td>
<td>2.60 (0.62)</td>
<td>0.07 0.26 0.67</td>
</tr>
<tr>
<td>13</td>
<td>I often lose things</td>
<td>.62</td>
<td>.44*</td>
<td>2.23 (0.74)</td>
<td>0.18 0.41 0.41</td>
</tr>
<tr>
<td>2</td>
<td>I can make fast decisions (e.g., in lessons)</td>
<td>.54</td>
<td>.34*</td>
<td>2.34 (0.59)</td>
<td>0.06 0.54 0.40</td>
</tr>
<tr>
<td>3</td>
<td>I am well-organized. For example, I am good at planning what I need to do during a day</td>
<td>.64</td>
<td>.33*</td>
<td>2.19 (0.71)</td>
<td>0.18 0.46 0.36</td>
</tr>
<tr>
<td>9</td>
<td>It is easy for me to come up with a different solution if I get stuck when solving a problem</td>
<td>.67</td>
<td>.41*</td>
<td>2.21 (0.57)</td>
<td>0.08 0.64 0.28</td>
</tr>
<tr>
<td>10</td>
<td>I am full of new ideas</td>
<td>.48</td>
<td>.40*</td>
<td>2.19 (0.64)</td>
<td>0.13 0.55 0.32</td>
</tr>
<tr>
<td>14</td>
<td>I am curious, I want to know how things work</td>
<td>.37</td>
<td>.30*</td>
<td>2.34 (0.64)</td>
<td>0.09 0.48 0.43</td>
</tr>
</tbody>
</table>

*Note. AEFI = Amsterdam Executive Function Inventory; \( r_{is} \) = corrected item–scale correlation. 
*\( p < .01. \)
### TABLE 3
Final multiple linear regression models of the Amsterdam Executive Function Inventory scale scores following a step-down hierarchical procedure

<table>
<thead>
<tr>
<th>AEFI scale score</th>
<th>Variable</th>
<th>B</th>
<th>B SE</th>
<th>t</th>
<th>SD (residual)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Constant)</td>
<td>5.795</td>
<td>0.042</td>
<td>138.296*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>0.310</td>
<td>0.058</td>
<td>5.3467*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE average</td>
<td>0.031</td>
<td>0.068</td>
<td>0.452</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE high</td>
<td>0.517</td>
<td>0.066</td>
<td>7.871*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sex × LE average</td>
<td>−0.108</td>
<td>0.095</td>
<td>−1.133</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sex × LE high</td>
<td>−0.366</td>
<td>0.097</td>
<td>−3.767*</td>
<td>1.572</td>
</tr>
<tr>
<td>Attention</td>
<td>(Constant)</td>
<td>11.264</td>
<td>0.041</td>
<td>273.841*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE average</td>
<td>0.165</td>
<td>0.067</td>
<td>2.452</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE high</td>
<td>0.614</td>
<td>0.068</td>
<td>8.984*</td>
<td>2.233</td>
</tr>
<tr>
<td>Self-Control and Self-Monitoring</td>
<td>(Constant)</td>
<td>10.946</td>
<td>0.044</td>
<td>251.066*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>0.134</td>
<td>0.049</td>
<td>2.771*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE average</td>
<td>0.319</td>
<td>0.058</td>
<td>5.470*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE high</td>
<td>0.664</td>
<td>0.059</td>
<td>11.221*</td>
<td>1.928</td>
</tr>
<tr>
<td>Planning and Initiative</td>
<td>(Constant)</td>
<td>10.946</td>
<td>0.044</td>
<td>251.066*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE average</td>
<td>0.165</td>
<td>0.067</td>
<td>2.452</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE high</td>
<td>0.614</td>
<td>0.068</td>
<td>8.984*</td>
<td>2.233</td>
</tr>
</tbody>
</table>

**Note.** AEFI = Amsterdam Executive Function Inventory. The full model included age, age², LE average, LE high, sex, and all two-way interactions. The SD (residual) values correspond with standard errors of the estimates of the regression models. LE = level of education. Coding of the predictors: Sex: male = 1, female = 0; LE average: low or high education = 0, average education = 1; LE high: low or average education = 0, high education = 1.

* *p < .01.

---

**DISCUSSION**

The results of the present study confirmed that three factors underlay the item responses on the AEFI—that is, Attention, Self-Control and Self-Monitoring, and Planning and Initiative. The a priori expected three-factor model had an adequate fit with the data, and the internal consistency and reliability of the established scales were adequate. Demographic variables affected all three AEFI scale scores. Higher educated participants had higher Self-Control and Self-Monitoring scale scores and higher Planning and Initiative scale scores.

---

**Figure 1.** Mean scores (and standard errors of the means) for the (a) Attention, (b) Self-Control and Self-Monitoring, and (c) Planning and Initiative AEFI scale scores, as a function of gender and level of education.

is lower than the score that was obtained by this testee.

The four-step normative procedure provides accurate norms but lacks user-friendliness because the users of the norms have to actively make the required computations. Therefore, we also provided a normative table that was derived from the four-step normative procedure (see the Appendix). The use of this table is straightforward. For example, the Appendix immediately shows that a raw Attention scale score equal to 4 corresponds to a percentile value equal to 8 (for a 16-year-old low-educated male).
scores than their lower educated counterparts. The effect of educational level on the Attention scale score was moderated by gender: Boys had higher levels of self-reported Attention scores than girls in the group of lower educated participants, but the gender differences in the Attention scale scores decreased as the educational level of the children increased (see Figure 1a). Previous studies in which the moderating effects of (parental) level of education on performance-based measures of executive functioning were evaluated in children and early adolescents (aged below 14 years) showed similar results (Ardila et al., 2005; Klenberg et al., 2001; Noble et al., 2005). Moreover, males obtained higher Planning and Initiative scale scores than females. This finding is in line with previous studies (Ardila et al., 2005; Brocki & Bohlin, 2004), though not all studies have reported better executive functions in males (female advantages have also been reported, especially when the executive function tests used rely strongly on verbal abilities; Anderson et al., 2001; Berlin & Bohlin, 2002).

The specific patterns of association between gender, level of education, and the different components of executive functioning as measured with the AEFI are unclear, especially because gender and level of education are proxies of multiple factors that may affect executive abilities—including environmental and social variables, personality traits (such as cautiousness), and verbal abilities (Klenberg et al., 2001; Noble et al., 2005). One way to evaluate these patterns of association in more detail would be through the explicit measurement and modeling of these proxy variables together with the demographical variables and the AEFI scale scores in a series of structural equation models. This is beyond the scope of the present paper (for a discussion on the mechanisms underlying the effects of gender and educational level on aspects of executive functioning, see, e.g., Klenberg et al., 2001, and Noble et al., 2005), but from a psychometric viewpoint the finding that level of education and gender affect the AEFI scale scores implies that demographically corrected normative data for the AEFI scale scores are required to adequately interpret the scale scores (see the Appendix).

The Attention and the Self-Control and Self-Monitoring AEFI scale scores were highly correlated, which could be indicative (a) that a single higher order factor underlay these two first-order factors, or (b) that the items of these two scales loaded on a single first-order factor (rather than on two separate factors), but additional analyses showed that this was not the case. The pattern of results is in agreement with previous studies in which it was concluded that executive functions comprise multiple components, which are related but yet separable. Note also that the differential effects of gender and educational level on the different AEFI scale scores (see Table 3) further indicate that the different components of executive functioning as measured with the AEFI are indeed separable (i.e., if it were the case that a single unitary executive function factor would underlie all item responses, then it would also be expected that all three AEFI scale scores would be influenced in a similar way by gender and/or level of education, which was not the case).

Some critical remarks and limitations of the present study can be given. First, the results of the CFA provided support for the construct validity of the AEFI (i.e., the obtained factor structure was consistent with the a priori expected model; Floyd & Widaman, 1995), but other types of validity were not evaluated. Future studies are needed to establish other types of validity of the AEFI, especially its discriminant, predictive, and ecological validity. For example, it has been suggested that executive function measures may allow for a better prediction of school performance than do more general cognitive indices (such as IQ; Best, Miller, & Jones, 2009). Future studies should evaluate this issue.

Second, after exclusion of a poorly fitting item (Item 4), the Attention scale of the AEFI consisted of only three items. There are both advantages and disadvantages of shorter and longer scales. For example, longer scales usually have better psychometric properties (e.g., higher Cronbach’s alpha values), whilst shorter scales take up less precious space on a survey and increase compliance of test-takers. Three items per scale is usually considered to be the minimum number of items in a multidimensional instrument (Raubenheimer, 2004), at least when the scale has an adequate level of reliability—which was the case for all scales of the AEFI.

Third, age and age² did not affect any of the AEFI scale scores. The age range that was considered in the present study was, however, quite narrow (i.e., 15–18 years), which raises the possibility that the insignificant results may (partly) be attributable to the restricted age variability in the sample. This was, however, probably not the case, because our results are in line with earlier studies that showed that improvements in executive functioning mainly occur in younger children and in early adolescents (aged below 14 years), with little or no improvement afterwards (Anderson et al., 2001; Ardila et al., 2005; Blakemore & Choudhury, 2006; Brocki & Bohlin, 2004; Levin et al., 1991; Welsh et al., 1991). Nevertheless, future studies should administer the AEFI in a sample that has
a broader age range (e.g., 10–25 years) to further examine this issue. Note that such a study would also allow for an evaluation of the AEFI’s validity to detect the developmental improvements in executive functions during childhood and early adolescence.

Fourth, the valid use of self-report questionnaires requires that the respondents have sufficient reading skills, have insight into their own problems, have perspective-taking skills, and have sufficient cognitive abilities to understand and reflect upon the questions (Finlay & Lyons, 2001; Theunissen et al., 1998). These conditions may not be fulfilled in, for example, younger children or in children with cognitive impairments. It would thus be advantageous to have an informant-based AEFI version in addition to the self-report version. At present, a parental-report version of the AEFI is being developed.

Fifth, the distribution of educational level in our sample deviated from the population distribution (i.e., 46.3%, 27.3%, and 26.4% of the children in the sample had low, average, and high levels of education, respectively, as compared to 54.9%, 22.1%, and 23.0% in the general population of 15–18-year-old children; Kuyper et al., 2010). This difference is, however, not a problem, because all analyses and the derived normative data were corrected for LE. On a related note, the imbalance in the number of participants in the educational categories (i.e., n = 2,948, n = 1,742, and n = 1,676 for low-, average-, and high-educated participants, respectively) is not a problem when a regression-based normative procedure is used, because any imbalance in the sample does not bias the estimation of the regression weights as long as the model is properly specified (i.e., as long as educational level is included in the model, if it is truly associated with the outcomes; Kleinbaum, Kupper, Muller, & Nizam, 1998). The use of imbalanced data may lead to a loss of statistical power (because the standard error of the regression weight is proportional to the \(\sqrt{VIF}\) of the predictor at hand), but this was not a problem in the present study because LE was significant in all models.

In conclusion, the AEFI is a newly developed brief self-report questionnaire to assess different executive aspects of daily-life behavior. Based on a large sample of adolescents (aged between 15 and 18 years), we found that gender and educational level affected the different AEFI scale scores. Normative data that took these demographical variables into account were established, so that the AEFI can be used by clinicians and researchers who need to assess executive aspects of daily-life behavior. Additional psychometric analyses supported the reliability and the construct validity of the AEFI.

Original manuscript received 15 April 2011
Revised manuscript accepted 6 July 2011
First published online 24 November 2011

REFERENCES


## APPENDIX

Normative data for the Attention, Self-control and Self-monitoring, and Planning and Initiative scale scores stratified by the relevant demographic variables

<table>
<thead>
<tr>
<th>Raw score</th>
<th>LE low</th>
<th>LE average</th>
<th>LE high</th>
<th>LE low</th>
<th>LE average</th>
<th>LE high</th>
<th>Raw score</th>
<th>LE low</th>
<th>LE average</th>
<th>LE high</th>
<th>Raw score</th>
<th>LE low</th>
<th>LE average</th>
<th>LE high</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>95</td>
<td>97</td>
<td>94</td>
<td>99</td>
<td>98</td>
<td>93</td>
<td>15</td>
<td>99</td>
<td>96</td>
<td>94</td>
<td>15</td>
<td>97</td>
<td>96</td>
<td>95</td>
</tr>
<tr>
<td>8</td>
<td>84</td>
<td>88</td>
<td>83</td>
<td>91</td>
<td>89</td>
<td>81</td>
<td>14</td>
<td>91</td>
<td>86</td>
<td>82</td>
<td>14</td>
<td>91</td>
<td>88</td>
<td>81</td>
</tr>
<tr>
<td>7</td>
<td>68</td>
<td>72</td>
<td>66</td>
<td>77</td>
<td>75</td>
<td>62</td>
<td>13</td>
<td>78</td>
<td>72</td>
<td>64</td>
<td>13</td>
<td>80</td>
<td>76</td>
<td>74</td>
</tr>
<tr>
<td>6</td>
<td>42</td>
<td>48</td>
<td>39</td>
<td>56</td>
<td>52</td>
<td>35</td>
<td>12</td>
<td>64</td>
<td>57</td>
<td>47</td>
<td>12</td>
<td>65</td>
<td>60</td>
<td>57</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>26</td>
<td>20</td>
<td>30</td>
<td>28</td>
<td>18</td>
<td>11</td>
<td>40</td>
<td>40</td>
<td>32</td>
<td>11</td>
<td>46</td>
<td>41</td>
<td>38</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>11</td>
<td>7</td>
<td>15</td>
<td>13</td>
<td>6</td>
<td>10</td>
<td>25</td>
<td>25</td>
<td>18</td>
<td>10</td>
<td>26</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td>14</td>
<td>14</td>
<td>10</td>
<td>9</td>
<td>13</td>
<td>8</td>
<td>8</td>
</tr>
</tbody>
</table>

Note. The raw scale scores and their corresponding percentiles are presented. LE = level of education.