Psychological Assessment
2014, Vol. 26, No. 1, 55–68

1040-3590/14/$12.00 DOI: 10.1037/a0034475

Factorial Validity and Measurement Invariance Across Intelligence Levels and Gender of the Overexcitabilities Questionnaire-II (OEQ-II)

Wim Van den Broeck, Joeri Hofmans, Sven Cooremans, and Eva Staels
Vrije Universiteit Brussel

The concept of overexcitability, derived from Dabrowski's theory of personality development, offers a promising approach for the study of the developmental dynamics of giftedness. The present study aimed at (a) examining the factorial structure of the Overexcitabilities Questionnaire-II scores (OEQ-II) and (b) testing measurement invariance of these scores across intelligence and gender. A sample of 641 Dutch-speaking adolescents from 11 to 15 years old, 363 girls and 278 boys, participated in this study. Results showed that a model without cross-loadings did not fit the data well (using confirmatory factor analysis), whereas a factor model in which all cross-loadings were included yielded fit statistics that were in support of the factorial structure of the OEQ-II scores (using exploratory structural equation modeling). Furthermore, our findings supported the assumption of (partial) strict measurement invariance of the OEQ-II scores across intelligence levels and across gender. Such levels of measurement invariance allow valid comparisons between factor means and factor relationships across groups. In particular, the gifted group scored significantly higher on intellectual and sensual overexcitability (OE) than the nongifted group, girls scored higher on emotional and sensual OE than boys, and boys scored higher on intellectual and psychomotor OE than girls.

Keywords: overexcitabilities, intelligence, measurement invariance, giftedness, high sensitivity

Within the research domain on giftedness, it is widely acknowledged that giftedness involves more than high intelligence (Sternberg, 2004; Sternberg, Jarvin, & Grigorenko, 2011). Overexcitability (OE) is a characteristic often seen to distinguish gifted individuals from their peers (Lovecky, 2004; Mendaglio, 2003; Piechowski & Colangelo, 1984; Silverman, 1997) and is defined as a "higher than average responsiveness to stimuli, manifested either by psychomotor, sensual, emotional (affective), imaginational, or intellectual excitability, or a combination thereof" (Dabrowski, 1972, p. 303). In line with this definition, Dabrowski (1972) identified five forms of OE, that is, a higher level of physical energy (psychomotor OE), an enhanced refinement and aliveness of the senses (sensual OE), a vivid imagination (imaginational OE), an intellectual curiosity and thirst for knowledge (intellectual OE), and a heightened emotional sensitivity (emotional OE; Daniels & Piechowski, 2009; Mendaglio & Tillier, 2006).

The Factorial Structure of the Overexcitability Questionnaire

Following Dabrowski's (1972; theoretical) footsteps, Piechowski (1979, 1986) hypothesized that OE may be more prevalent among gifted individuals. To investigate this hypothesis, an instrument was developed that could measure the different OEs. This instrument was the Overexcitability Questionnaire (OEQ), an open-ended questionnaire consisting of 21 items (see Lysy & Piechowski, 1983). Researchers using the OEQ have found gifted individuals to score significantly higher on several OEs compared to nongifted individuals (Ackerman, 1997; Falk, Manzanero, & Miller, 1997; Gallagher, 1985; Lysy & Piechowski, 1983; Miller, Silverman, & Falk, 1994; Piechowski & Colangelo, 1984; Piechowski & Cunningham, 1985; Piechowski, Silverman, & Falk, 1985; Schiever, 1985; Silverman & Ellsworth, 1981). The similarity of the OE-profiles, as observed in gifted individuals, led some researchers to suggest that OEs could perhaps be used as a complementary tool to identify giftedness (Ackerman, 1997; Bouchard, 2004; Gallagher, 1985; Miller et al., 1994). To investigate this, research had to be performed on a larger scale. Because the administration and the scoring of the OEQ were labor-intensive, a new measuring instrument had to be devised. This new version of the OEQ, the OEQ-II, was released in 1999 and is a 50-item self-rating questionnaire with 10 items for each OE domain. In the original pilot study, all items had loadings of .50 or above on their respective OE domain and all OE domains showed high Cronbach's alpha coefficients (see Falk, Lind, Miller, Piechowski, & Silverman, 1999).

Whereas this pilot test seemed to support the factorial validity of the OEQ-II scores, later tests based on confirmatory factor analysis (CFA) were less positive. In particular, Tieso (2007) found the fit

---

Wim Van den Broeck, Department of Clinical and Lifespan Psychology, Vrije Universiteit Brussel, Brussels, Belgium; Joeri Hofmans, Department of Experimental and Applied Psychology, Vrije Universiteit Brussel; Sven Cooremans and Eva Staels, Department of Clinical and Lifespan Psychology, Vrije Universiteit Brussel.

The authors wish to acknowledge Fanny Ferdinande, Evy Seminckx, and Leen Vandercammen for their contribution in collecting the data. The first and the second authors contributed equally to the study.

Correspondence concerning this article should be addressed to Wim Van den Broeck, Faculty of Psychology and Educational Sciences, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium. E-mail: wivdbroe@vub.ac.be

of the data to the hypothesized model to be moderate (root-mean-square error of approximation [RMSEA] = .049; goodness-of-fit index [GFI] = .79; comparative fit index [CFI] = .85). Whereas this lack of an adequate fit is informative, it may not be surprising. The reason is that most CFA analyses rely on the independent clusters model (ICM; Marsh et al., 2009), a model in which each item loads on only one factor and where no cross-loadings are present. While an ICM is very parsimonious, it is often found to be too restrictive to model scores pertaining to measures of personality (Church & Burke, 1994; Marsh, 2007; Marsh et al., 2009, 2010). With exploratory structural equation modeling (ESEM), an interesting alternative approach became available recently (Asparouhov & Muthén, 2009). ESEM uses exploratory factor analysis (EFA) for measurement modeling, which means that it models all cross-loadings. Moreover, ESEM has the additional advantage that it gives access to all the well-known SEM parameters such as residual correlations, regressions of factors on covariates, and regressions among factors (Asparouhov & Muthén, 2009). In a study of the Big-Five factor structure of the NEO-Five Factor Inventory, Marsh et al. (2010) demonstrated that when switching from the ICM to the ESEM model, the gain in absolute fit outweighed the loss in parsimony, resulting in better approximate fit indexes for the latter model (CFI and RMSEA).

Drawing on these findings, the first goal of this study is to examine the factorial structure of the scores from the Dutch language version of the OEQ-II. Two models will be tested: the ICM and ESEM. Based on previous findings on the factorial structure of a similar measuring instrument (Marsh et al., 2010), it is hypothesized that the ESEM will yield a better model fit–model parsimony balance than the ICM.

Measurement Invariance of the OEQ-II Scores

The majority of studies that have used the OEQ-II were designed to compare the scores of gifted and nongifted people. The results of these studies varied. Whereas all studies found gifted people to score higher on the intellectual OE scale (Bouchet & Falk, 2001; Chang, 2001; Chavez, 2004; Pardo de Santayana Sanz, 2006; Siu, 2010; Tieso, 2007; Wirthwein & Rost, 2011), only two studies found that gifted scored higher than nongifted people on all OE subscales (Chang, 2001; Siu, 2010). In the pilot study, Bouchet and Falk (2001) also reported a higher score on the emotional OE scale. Two studies found gifted people to score higher on the imaginational OE scale (Pardo de Santayana Sanz, 2006; Tieso, 2007). Only one study reported gifted scoring lower than nongifted people, and this on the psychomotor OE scale (Pardo de Santayana Sanz, 2006). A recent study that compared OE-scores across different countries and cultures revealed that the Korean gifted group scored higher on the psychomotor OE scale than the U.S. group, while the U.S. group scored higher on the imaginational OE scale than the Korean group (Piirto, Montgomery, & May, 2008).

Apart from differences between gifted and nongifted people, studies on overexcitability have also reported gender differences (see Silverman, 2008). In particular, females score higher than males on the Emotional OE scale and Sensual OE scale (Bouchet & Falk, 2001; Gross, Rinn, & Jamieson, 2007; Miller, Falk, & Huang, 2009; Moon & Montgomery, 2005; Tieso, 2007; Treat, 2006; Wirthwein, Becker, Loehr, & Rost, 2011). Conversely, males were found to score significantly higher on the intellectual (Bouchet & Falk, 2001; Miller et al., 2009; Moon & Montgomery, 2005; Treat, 2006) and on the psychomotor OE scale (Moon & Montgomery, 2005; Tieso, 2007; Treat, 2006). Finally, one study found higher scores for females on the imaginational OE scale than for males (Gross et al., 2007).

At this point it is important to note that a logical requirement of valid cross-group comparisons is that the questionnaire is measuring the same construct in the same way in all of these groups. This is the general issue of measurement invariance (Byrne, Shavelson, & Muthén, 1989; Millsap, 2011). Only when the assumption of measurement invariance is valid, observed mean differences are indicative of differences on the latent factor level. As most of the studies with the OEQ-II were aimed at comparing OE-scores across different groups (gifted vs. nongifted; male versus female), it is surprising that measurement invariance has not yet been investigated.

To address this limitation, the second goal of this study is to test for measurement invariance of the OEQ-II scores across a gifted and nongifted group as well as across gender. Following the tests on measurement invariance, we explore between-group differences in factor variances, factor covariances, and factor means. This will allow us to evaluate whether the differences that were found in previous research point to true between-group differences.

Method

Participants

A sample of 641 adolescents, ranging from 11 to 15 years old ($M = 13.3$ and $SD = 0.78$) participated voluntarily in this study. There were 363 girls and 278 boys. The participants were students from the first (53.4%) or the second year (46.6%) of secondary school. The sample was drawn from seven schools, located in different regions of Flanders, the Dutch speaking part of Belgium. As an indication of socioeconomic status, for 1.6% of the participants both parents received no or only primary education, 12.6% of the participants had at least one parent who received at most secondary education, and 85.8% of the participants' had at least one parent with a higher education degree. For boys and girls, these socioeconomic status (SES) percentages were very similar (Mann-Whitney U, $p = .739$), although for the gifted group parents were more likely to fall in higher SES-categories than for the nongifted group (Mann-Whitney U, $p = .031$). For 75.2% of the participants both parents were of Belgian descent, 13.2% of the participants had one parent who was of Belgian descent, and 11.6% of the participants' parents were of foreign origin (in half of the cases of African origin). As expected, the distribution of ethnicity was similar for boys and girls (Mann-Whitney U, $p = .770$). Compared with nongifted individuals (72.5%), gifted individuals tended to be more frequently from Belgian descent (81.3%) although this difference was only marginally significant (Mann-Whitney U, $p = .052$).

Instruments

Overexcitability Questionnaire-II. The instrument used to measure the five forms of OE was the Dutch language version of the Overexcitability Questionnaire-II (OEQ-II; Falk et al., 1999). This Dutch version was obtained using back translation (Braham

& Andries, 2004). The OEQ-II is a 50-item self-rating questionnaire (see Appendix). Each OE subscale (psychomotor; sensual; imaginational; intellectual; emotional) is made up of 10 items and all items are rated on a 5-point Likert-type scale ranging from 1 (*not at all like me*) to 5 (*very much like me*). The original English version of the instrument resulted in scores that are internally consistent, with Cronbach alpha scores from .84 (emotional) to .89 (sensual; intellectual; Falk et al., 1999). For the translated version, Braham and Andries (2004) obtained Cronbach alpha scores ranging from .79 (imaginational) to .85 (psychomotor).

Raven's Standard Progressive Matrices. The Raven's Standard Progressive Matrices (SPM) was administered to measure participants' (general) intelligence. The SPM is a nonverbal test, consisting of five sets of 12 different matrices that gradually increase in difficulty. Each item shows a design with one part missing, and the respondent has to select the correct part to complete the design from six options (Raven, Raven, & Court, 1998). The Raven's Progressive Matrices is widely accepted as a measure of analytical reasoning and fluid intelligence (Carpenter, Just, & Shell, 1990). To investigate measurement invariance across intelligence, we created groups according to the score on the Raven: The first group was formed with students who had a score above P80 ($N = 144$), the second with students who had a score beneath P60 ($N = 407$). This way, the more intelligent (or gifted) students were grouped in one group, the students with an average to low intelligence in another. Although the IQ scores of gifted people often exceed P80, we chose for this cutoff as it allowed us to test for measurement invariance while still retaining a reasonably large group of intelligent students. The reliability estimate of the Raven scores based on a congeneric model was .72 (.73 for boys and .71 for girls, .52 for the nongifted group and .01 for the gifted group). The low reliability estimate for the gifted group is due to a serious range restriction in the scores of this group. This is not problematic as the precision of the IQ-scores in this group is better than that of the nongifted group and that of the entire sample (standard errors of measurement 4.48 vs. 8.57 and 7.94, respectively).

Administration Procedure

Both the OEQ-II and the Raven SPM were administered during class meetings at a single 1-hr session. Participation was anonymous, and the data were used for research purposes only. After some demographic data were gathered, and the students were instructed on the administration, they had 40 min to complete the Raven SPM, and 15 min to fill out the OEQ-II.

Data Analytic Procedure

All analyses were conducted using Mplus (Version 6.12; L. K. Muthén & Muthén, 2010). As the items of the OEQ-II are ordered-categorical measures, we used a weighted least squares mean and variance adjusted (WLSMV) estimator (Flora & Curran, 2004; B. Muthén & Kaplan, 1985), which is an estimator specifically designed for models involving ordinal, nonnormally distributed data.[1]

To examine the factorial validity of the OEQ-II scores, both an ICM and an ESEM were tested. Students' scores on the OE items were modeled as indicators of the five OE factors. Because these factors are predicted to covary according to Dabrowski's (1972) theory, factors were allowed to correlate in the ICM analysis. For purposes of comparability, in the ESEM analysis we used an oblique target rotation in which we rotated the solution toward a simple structure (i.e., the structure proposed by the ICM model). The fit of our ICM- and ESEM-models was not only assessed with the χ^2 statistic (with $0 \leq \chi^2 \leq 2df$ indicating good fit and $2df \leq \chi^2 \leq 3df$ indicating acceptable fit) but also with the CFI, the Tucker–Lewis index (TLI), and the RMSEA. For the CFI and TLI, values above .90 reflect acceptable fit and values above .95 reflect excellent fit to the data. For the RMSEA, values less than .05 indicate good model fit (Hu & Bentler, 1999; Marsh et al., 2010).

To test for measurement invariance, we first refitted the ESEM model in a CFA framework (i.e., ESEM-Within-CFA; Marsh, Nagengast, & Morin, 2013). The reason is that, despite the flexibility of the ESEM model, some aspects of traditional SEM models cannot readily be implemented. One such aspect, which is of particular relevance to the present study, is that ESEM does not allow for partial invariance of the factor loadings. For this reason, we build a CFA model with the same *df* and, within rounding error, the same chi-square value, fit statistics, and parameter estimates of the ESEM model. This was done by adding m^2 constraints to the ESEM-Within-CFA model (with m being the number of factors). To do so, we fixed the m factor variances to 1 and fixed the cross-loadings of m anchor items (i.e., items with a large loading on the factor that they are designed to measure and small cross-loadings on the other factors) to the values of the ESEM solution (see the supplemental material of Marsh et al., 2013). This procedure yields an ESEM-Within-CFA solution that is equivalent to the ESEM solution and that is comparable to traditional SEM in terms of possible parameter constraints (Marsh et al., 2013). Subsequently, invariance of the ESEM-Within-CFA model was evaluated by estimating a series of nested ESEM-Within-CFA models using the Theta parameterization of MPlus (Millsap & Yun-Tein, 2004). In the Theta parameterization, residual variances for continuous latent response variables of observed categorical dependent variables are parameterized in the model, but scale factors for continuous latent response variables are not (L. K. Muthén & Muthén, 2010).[2]

First, we examined whether the same configuration holds across both groups (i.e., *configural or pattern invariance*). To this end, we estimated an ESEM-Within-CFA model in which the factor variances were fixed to 1 in both groups, the cross-loadings of the anchor items were fixed to the values of the single-group ESEM solution in both groups, and the means were fixed to 0 in the first group (Morin, Marsh, & Nagengast, 2013). Moreover, the first and second item threshold of each item were constrained equal across groups (Millsap & Yun-Tein, 2004). The other model parameters (i.e., the factor load-

---

[1] Normality was tested using the SPSS macro of DeCarlo (1997). For 48 of the 50 items, the univariate distribution of the scores significantly differed from normality. In addition, an omnibus test for multivariate normality based on Small's statistic (DeCarlo, 1997) revealed significant deviations from multivariate normality as well, $\chi^2(100) = 3720.91$, $p < .001$.

[2] When using the theta parametrization in Mplus, all latent intercepts are fixed to zero (see Millsap & Yun-Tein, 2004).

ings of the nonanchor items, the main loadings of the anchor items, the third and fourth item thresholds, the unique item variances, the factor covariances, and the factor means in the second group) were estimated freely. Subsequently, we progressively constrained additional sets of parameters to be equal across the groups. In particular, in the second test, *weak or metric modification index (MI)* was tested by, in addition to the constraints from the configural invariance model, also constraining the factor loadings of the nonanchor items and the main loadings of the anchor items to be equal across groups. No equality constraint was put on the third and fourth item thresholds, the unique item variances, the factor covariances, and the factor means. Moreover, the factor variances in the second group were also estimated freely (see Morin et al., 2013). *Strong or scalar MI* was evaluated in the third test, where in addition to the constraints of the weak MI model, the third and fourth item threshold were constrained to be equal across both groups (unique item variances, factor variances, factor covariances, and factor means were still allowed to differ). In the fourth test, we tested for *strict MI* by extending the strong MI model with an across-group equality constraint on the unique item variances (in addition to the factor loadings and thresholds; Marsh et al., 2010; Meredith, 1993; Millsap & Yun-Tein, 2004).

Following the tests for MI, we tested for structural invariance by examining whether the factor variances, factor covariances, and factor means were invariant across groups. This was done by adding equality constraints to the strict MI model. In particular, first *factor variance invariance* was tested by constraining the factor variances to be equal across groups (in addition to the constraints of the strict MI model). Subsequently, we tested for *factor covariance invariance* by constraining the factor variances and covariances to be equal across groups. Finally, to test for *factor mean invariance*, we constrained the factor variances and covariances to be equal across groups, and fixed the factor means in the second group to 0.

From the foregoing, it is clear that we used a stepwise procedure where with each step more constraints were added to the model. At each step of the sequence, the absolute as well as the relative fit of the model, that is, the degree of model fit as well as the degree of change in fit, was evaluated. To assess absolute model fit, we used the χ^2, CFI, TLI, and RMSEA. To compare the fit between consecutive models, the DIFFTEST option was used to obtain a corrected chi-square difference test (L. K. Muthén & Muthén, 2010). If the corrected chi-squared difference test was nonsignificant, the fit of the more highly constrained model did not differ significantly from that of the less constrained one. Hence, the more highly constrained model was preferred. However, if the corrected chi-squared difference test was significant, at least one of the constrained parameters was noninvariant across groups.

To identify possible sources of misfit (i.e., the noninvariant parameters), we relied on an inspection of the MIs (see Byrne et al., 1989). MIs pertain to specific fixed parameters in the model (in this case the parameter constraints) with a MI for a particular parameter representing the lower bound estimate of the expected decrease in chi-square when this parameter would be estimated freely (i.e., when the parameter constraint would be removed). As such, the parameter with the highest MI has the largest contribution to model misfit, which implies that the data do not support the constraint on this parameter and that it has to be removed. The procedure of identifying noninvariant parameters was done in a one-by-one fashion through examination of the MIs. Specifically, the parameter constraint found to contribute most to model misfit was removed and the model subsequently reestimated and reevaluated based on the DIFFTEST. If the DIFFTEST was still statistically significant, an additional parameter was freed based on the MI's of the last model. This procedure was repeated until the *p*-value of the DIFFTEST exceeded .05.

Results

Missing Data

Initial inspection of the data revealed that 1.47% of the OEQ-II item scores were missing or invalid. To examine whether the data were missing at random (MAR), thereby justifying the use of data imputation methods, we conducted Little's missing completely at random (MCAR) test (Little, 1988). This test, for which the null hypothesis is that the data are missing completely at random, was found to be statistically significant ($p = .037$). However, missing at random (MAR) could be assumed, which means that the missingness on a given variable does not depend on the values of the variable itself but may depend on other variables that are included. MPlus provides maximum-likelihood estimation under MAR for ordered categorical variables.

Reliability

The OEQ-II total score demonstrated satisfactory internal consistency reliability ($\alpha = .897$). Moreover, the subscale scores displayed acceptable internal consistency as well, with Cronbach's alphas ranging from .75 to .83.

OEQ-II Factor Structure: ESEM Versus ICM

In a first step, all data were subjected to both an ESEM and an ICM. As judged by two of the three fit indices, the ESEM solution provided a modestly acceptable fit to the data: CFI = .913 (>.90) and RMSEA = .044 (<.05). One fit index was below the recommended cutoff value: TLI = .891 (<.90). For the ICM model (see Figure 1), all three fit indices fell below their recommended cut value: CFI = .846 (<.90); TLI = .839 (<.90); RMSEA = .053 (>.05).

Because both the ICM and the ESEM solutions were unable to meet the criteria for acceptable fit, the models were reexamined for potential improvements. Inspection of the MIs suggested that the fit indices of the ESEM model were acceptable if an error covariance were added between Items 1 ("I like to daydream") and 14 ("When I get bored, I begin to daydream"; MI = 194.544) and between Items 37 ("I am moved by beauty in nature") and 48 ("I love to listen to the sounds of nature"; MI = 139.90). Adding these error covariances also made sense from a substantive point of view as Items 1 and 14 ("daydreaming"), and 37 and 48 ("nature") refer to the same activity or object. Therefore, it could be expected that these items share unique variance that is not shared with the other items of the Imaginational and Sensual factor, respectively (as these other items concern other activities and objects). In other words, we believe that these items share 'unique' variance that is linked to the particular activity. Inclusion of the two error covariances resulted in models with the following fit indices: χ^2 = 2872.19; *df* = 1163, $p < .001$; CFI = .877; TLI = .870; RMSEA = .048 for the ICM and χ^2 = 1829.49; *df* = 983; $p < .001$; CFI = .939; TLI = .924; RMSEA = .037 for the ESEM. When comparing the fit indices of both solutions, it is clear that the ESEM solution provided a good fit on all three indices, whereas the ICM did not. Apparently, the loss of parsimony due to the less restrictive ESEM-model did not outweigh the gain in absolute fit. For this reason, we continued with the ESEM model, the factor loadings (λs), thresholds (τs), and unique item variances (δs) of which are shown in Table 1.

Measurement Invariance Across Intelligence

Table 2 presents the goodness-of-fit statistics and the DIFFTEST. The configural invariance model (i.e., Model 1) showed good fit statistics: RMSEA was well below .05, CFI and TLI were above .90, and the χ^2-value was smaller than 2*df*. In a next step, weak MI was tested (i.e., Model 2). Again, the fit indices indicated a good model fit (see Table 2). Moreover, comparison of the fit indices with the previous model showed an improvement on all three approximate fit indices, thereby supporting the more parsimonious weak MI model. However, the chi-square difference statistic (DIFFTEST) pointed to a significant difference with the previous model. To identify the noninvariant factor loadings, we inspected the MIs. After freeing the factor loading with the highest MI, the model was again compared with the configural invariance model using the DIFFTEST. If the DIFFTEST became nonsignificant, the procedure stopped; otherwise the MIs of the new model were inspected and an additional factor loading was freed. Following this procedure the loadings of Item 48 on Intellectual OE (MI = 19.52), Item 49 on Imaginational OE (MI = 18.40), Item 15 on Imaginational OE (MI = 18.59), and Item 29 on Sensual OE (MI = 16.00) were identified as noninvariant. When these factor loadings were freed across both groups, the DIFFTEST became nonsignificant (see Model 2d).

Subsequently, partial strong (or scalar) MI was tested (i.e., Model 3). As Table 2 shows, all fit indices indicated good fit, and comparison with the partial weak MI model showed an improvement of the three approximate fit indices. Moreover, the DIFFTEST was not statistically significant.

When testing for partial strict measurement invariance, model fit was good and compared to the previous model the approximate fit indices were virtually the same (see Model 4 in Table 2). However, the DIFFTEST was significant. Using the same procedure as before, we systematically examined the MIs to test which unique item variances were responsible for this departure from partial strict measurement invariance. We found that the unique variance of Items 27 (MI = 9.87), 14 (MI = 9.40), and 19 (MI = 9.38) were larger in the nongifted group

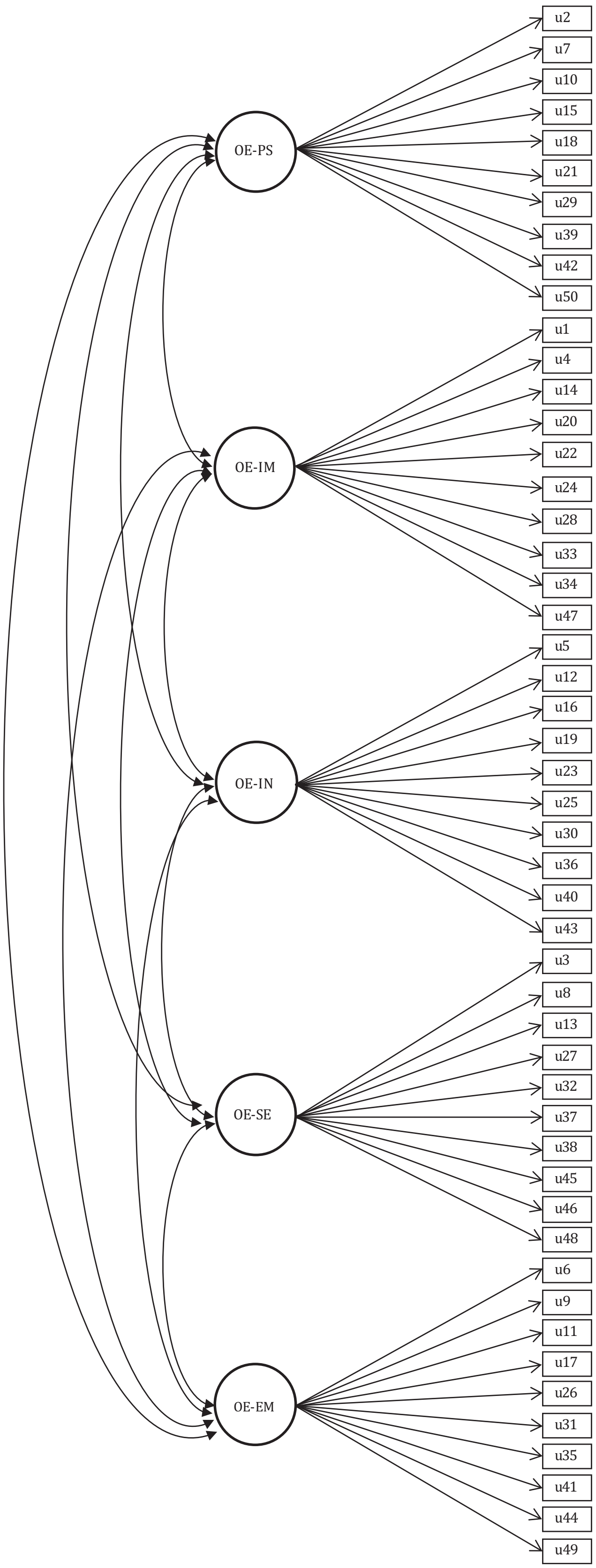

Figure 1. Independent clusters model of the five-factor structure of the Overexcitabilities Questionnaire-II (OEQ-II; Falk et al., 1999). OE = overexcitability; PS = psychomotor; IM = imaginational; IN = intellectual; SE = sensual; EM = emotional.

Table 1
Factor Loadings (λ), Thresholds (τ), and Error Variances (δ) of the ESEM-Model for the Total Sample

	Psychomotor		Imaginational		Intellectual		Sensual		Emotional		1		2		3		4		
Item	λ	*SE*	λ	*SE*	λ	*SE*	λ	*SE*	λ	*SE*	τ	*SE*	τ	*SE*	τ	*SE*	τ	*SE*	δ
1	−.08	.04	**.45**	.04	.08	.05	.00	.04	−.05	.05	−1.3	.07	−.40	.05	.58	.05	1.41	.07	.79
2	**.40**	.04	.02	.04	.27	.05	−.11	.05	−.21	.05	−1.28	.07	−.57	.05	.37	.05	1.11	.06	.71
3	.11	.04	.09	.05	.00	.04	**.46**	.04	.09	.05	−1.34	.07	−.55	.05	.42	.05	1.21	.07	.69
4	.06	.04	**.19**	.05	.15	.05	.08	.05	−.09	.05	−1.04	.06	−.37	.05	.46	.05	1.18	.06	.91
5	.06	.04	−.01	.04	**.45**	.04	.06	.05	−.10	.05	−1.43	.07	−.74	.06	.14	.05	.99	.06	.78
6	.04	.03	−.13	.07	.23	.05	−.03	.04	**.51**	.04	−1.37	.07	−.76	.06	.19	.05	1.13	.06	.67
7	**.46**	.04	−.08	.04	.30	.05	−.14	.05	−.07	.05	−1.29	.07	−.63	.05	.06	.05	.82	.06	.66
8	−.08	.04	−.08	.04	.07	.05	**.71**	.03	.03	.04	−.05	.05	.64	.05	1.16	.06	1.69	.09	.47
9	−.11	.04	.22	.07	.07	.04	−.01	.04	**.42**	.05	−1.52	.08	−.6	.05	.32	.05	1.11	.06	.69
10	**.85**	.02	−.17	.05	−.07	.03	.13	.04	−.01	.03	−2.05	.11	−1.42	.07	−.58	.05	.15	.05	.29
11	.00	.03	−.01	.04	.10	.05	.18	.05	**.42**	.04	−1.27	.07	−.38	.05	.59	.05	1.44	.07	.69
12	−.02	.04	.04	.05	**.33**	.05	−.05	.04	.10	.05	−1.42	.07	−.49	.05	.61	.05	1.49	.08	.86
13	−.02	.03	−.12	.05	.07	.04	**.73**	.04	.00	.03	−.59	.05	.30	.05	1.1	.06	1.73	.09	.47
14	−.01	.04	**.37**	.05	.08	.05	.05	.04	−.03	.05	−1.25	.07	−.56	.05	.04	.05	.80	.06	.83
15	**.68**	.03	−.03	.03	.16	.05	−.05	.04	.05	.04	−1.86	.10	−1.29	.07	−.63	.05	.18	.05	.47
16	−.04	.03	.14	.05	**.51**	.04	−.04	.04	.09	.05	−1.42	.07	−.58	.05	.24	.05	1.02	.06	.66
17	.17	.04	.22	.08	−.08	.05	.07	.04	**.43**	.06	−1.07	.06	−.37	.05	.37	.05	.98	.06	.66
18	**.73**	.02	.05	.04	.04	.03	.04	.03	−.03	.03	−1.14	.06	−.32	.05	.49	.05	1.15	.06	.44
19	.08	.04	.03	.04	**.46**	.05	.12	.05	.08	.05	−1.51	.08	−.63	.05	.35	.05	1.27	.07	.67
20	−.06	.04	**.44**	.04	.12	.05	−.02	.04	−.01	.05	−.36	.05	.40	.05	1.07	.06	1.63	.08	.78
21	**.33**	.04	.42	.05	−.02	.04	−.16	.05	.02	.05	−.95	.06	−.32	.05	.26	.05	.84	.06	.71
22	.04	.03	**.57**	.04	.10	.05	.10	.05	.00	.04	−.72	.05	.03	.05	.68	.05	1.34	.07	.57
23	−.07	.03	.09	.05	**.61**	.04	.04	.04	−.03	.04	−1.11	.06	−.16	.05	.70	.05	1.39	.07	.59
24	−.02	.03	**.72**	.04	.05	.04	.11	.05	−.13	.05	−.55	.05	.1	.05	.73	.06	1.36	.07	.46
25	−.05	.03	−.02	.04	**.57**	.04	.16	.04	.03	.04	−1.04	.06	−.13	.05	.69	.05	1.43	.07	.60
26	.18	.04	.46	.06	.04	.04	−.12	.04	**.27**	.06	−1.73	.09	−.88	.06	.13	.05	.80	.06	.61
27	.11	.04	.24	.06	−.01	.04	**.08**	.05	.28	.06	−1.76	.09	−1.04	.06	−.40	.05	.33	.05	.77
28	.05	.04	**.49**	.04	.01	.04	−.05	.05	.03	.05	−.92	.06	−.07	.05	.74	.06	1.43	.07	.75
29	**.56**	.03	.27	.05	−.04	.04	.05	.04	−.01	.04	−1.17	.06	−.33	.05	.45	.05	1.14	.06	.59
30	.06	.03	−.07	.05	**.48**	.04	.17	.05	.33	.05	−1.44	.07	−.59	.05	.26	.05	1.07	.06	.46
31	.00	.03	−.13	.08	−.01	.03	.06	.04	**.68**	.04	−1.99	.11	−1.14	.06	−.12	.05	.89	.06	.56
32	.01	.03	.17	.04	.00	.03	**.61**	.04	.01	.04	−.66	.05	.24	.05	1.07	.06	1.74	.09	.53
33	−.04	.04	**.48**	.05	−.11	.05	.30	.05	.03	.05	.03	.05	.62	.05	1.18	.06	1.76	.09	.61
34	.01	.03	**.60**	.04	−.09	.05	.13	.05	.04	.04	−.45	.05	.28	.05	.92	.06	1.65	.08	.59
35	−.07	.03	.28	.09	−.10	.05	−.02	.03	**.56**	.06	−.82	.06	−.16	.05	.40	.05	.99	.06	.55
36	−.06	.03	.15	.05	**.43**	.05	.06	.04	.21	.05	−.94	.06	−.19	.05	.60	.05	1.28	.07	.61
37	.11	.04	.09	.04	.11	.05	**.42**	.04	.03	.04	−1.06	.06	−.14	.05	.62	.05	1.34	.07	.70
38	−.13	.04	−.01	.05	−.03	.05	**.01**	.05	.04	.05	−1.19	.07	−.52	.05	.28	.05	.91	.06	.98
39	**.57**	.03	.13	.05	.12	.05	−.04	.03	.17	.05	−1.30	.07	−.58	.05	.18	.05	.77	.06	.53
40	.07	.04	.14	.05	**.32**	.05	.02	.04	.20	.05	−1.11	.06	−.19	.05	.82	.06	1.50	.08	.73
41	.00	.03	.29	.08	.06	.04	−.03	.03	**.53**	.06	−1.36	.07	−.58	.05	.28	.05	1.07	.06	.52
42	**.62**	.03	.06	.04	−.07	.04	.17	.05	.12	.04	−1.12	.06	−.44	.05	.26	.05	.89	.06	.55
43	.13	.04	.07	.04	**.51**	.04	.15	.05	.00	.03	−1.21	.07	−.25	.05	.63	.05	1.32	.07	.57
44	−.12	.05	−.01	.05	−.05	.05	.02	.05	**.18**	.05	−1.03	.06	−.55	.05	.02	.05	.53	.05	.96
45	−.02	.03	.05	.03	−.04	.03	**.80**	.03	−.06	.04	−.84	.06	.06	.05	.82	.06	1.40	.07	.39
46	.03	.03	.08	.04	.10	.04	**.55**	.04	.01	.04	−1.11	.06	−.35	.05	.42	.05	1.09	.06	.60
47	.04	.04	**.38**	.05	.12	.04	.19	.05	.08	.05	−1.40	.07	−.65	.05	.00	.05	.67	.05	.67
48	.03	.04	.01	.04	.05	.05	**.51**	.04	.03	.04	−1.04	.06	−.25	.05	.49	.05	1.13	.06	.70
49	−.02	.03	.13	.07	.08	.04	.14	.04	**.44**	.05	−1.28	.07	−.44	.05	.69	.05	1.51	.08	.64
50	**.75**	.03	−.10	.04	.06	.04	−.01	.03	−.03	.04	−1.71	.09	−1.11	.06	−.36	.05	.26	.05	.43

Note. ESEM = exploratory structural equation modeling. The principal subscale loadings are printed in bold.

than in the gifted group. After releasing the equality constraint for these unique variances, the DIFFTEST was no longer significant (see Model 4c).

Having established partial strict measurement invariance of the measurement model across intelligence groups, we are now in a position to validly compare the factor variances, factor covariances, and factor means across the gifted and nongifted group (Lievens & Anseel, 2004; Yoo, 2002). To do so, we started from the partial strict measurement invariance model and added an across-group equality constraint on the factor variances (see Model 5 in Table 2). This caused a small deterioration in model fit, which was due to the fact that the variance of Psychomotor OE was larger for the gifted than for the nongifted group (MI = 22.66). Subsequently, factor covariance invariance was demonstrated by a nonsignificant decrease in model fit when we added an additional equality constraint on the factor covariances (i.e., Model 6). Finally, we tested for invariance of the factor means by adding an equality constraint on the factor means as well. As

Table 2
ESEM-Models Testing Measurement Invariance Across Intelligence

Model	χ^2	*df*	*p*	CFI	TLI	RMSEA
1. Configural Invariance	2629.30	2011	<.001	.934	.919	.033
2. Weak Measurement Invariance						
χ^2 diff test with Model 1 = 286.18 (225); p = .0036	2743.52	2236	<.001	.946	.941	.029
2a. Partial Weak Measurement Invariance—Model 2 + $\lambda_{48\text{-int}}$ free						
χ^2 diff test with Model 1 = 278.79 (224); p = .0075	2732.85	2235	<.001	.947	.942	.028
2b. Partial Weak Measurement Invariance—Model 2a + $\lambda_{49\text{-ima}}$ free						
χ^2 diff test with Model 1 = 271.50 (223); p = .0147	2722.99	2234	<.001	.948	.943	.028
2c. Partial Weak Measurement Invariance—Model 2b + $\lambda_{15\text{-ima}}$ free						
χ^2 diff test with Model 1 = 263.22 (222); p = .0302	2711.74	2233	<.001	.949	.944	.028
2d. Partial Weak Measurement Invariance—Model 2c + $\lambda_{29\text{-sen}}$ free						
χ^2 diff test with Model 1 = 255.44 (221); p = .0558	2701.92	2232	<.001	.950	.945	.028
3. Partial Strong Measurement Invariance						
χ^2 diff test with Model 2d = 114.43 (100); p = .1535	2784.79	2332	<.001	.952	.949	.027
4. Partial Strict Measurement Invariance						
χ^2 diff test with Model 3 = 77.88 (50); p = .0070	2829.24	2382	<.001	.952	.951	.026
4a. Partial Strict Measurement Invariance—Model 4 + δ_{27} free						
χ^2 diff test with Model 3 = 72.29 (49); p = .0169	2824.35	2381	<.001	.953	.951	.026
4b. Partial Strict Measurement Invariance - Model 4a + δ_{14} free						
χ^2 diff test with model 3 = 66.09 (48); p = .0426	2818.54	2380	<.001	.953	.952	.026
4c. Partial Strict Measurement Invariance—Model 4b + δ_{19} free						
χ^2 diff test with Model 3 = 60.29 (47); p = .0923	2812.58	2379	<.001	.954	.952	.026
5. Factor variance invariance						
χ^2 diff test with Model 4c = 12.34 (5); p = .0240	2816.41	2384	<.001	.954	.952	.026
5a. Factor variance invariance—Model 5 + φ_2 free						
χ^2 diff test with Model 4c = 8.16 (4); p = .0858	2809.56	2383	<.001	.954	.953	.025
6. Factor covariance invariance						
χ^2 diff test with Model 5a = 8.23 (10); p = .6062	2731.74	2393	<.001	.964	.963	.023
7. Factor mean invariance						
χ^2 diff test with Model 6 = 43.27 (5); p < .0001	2853.83	2398	<.001	.951	.950	.026
7a. Partial Factor mean invariance—Model 7 + ξ_{int} free						
χ^2 diff test with Model 6 = 10.53 (4); p = .0324	2742.63	2397	<.001	.963	.962	.023
7b. Partial Factor mean invariance—Model 7a + ξ_{sen} free						
χ^2 diff test with Model 6 = 2.09 (3); p = .5546	2729.22	2396	<.001	.964	.964	.022

Note. ESEM = exploratory structural equation modeling; CFI = comparitive fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error of approximation; psy = psychomotor; sen = sensual; ima = imaginational; int = intellectual; emo = emotional; diff = difference.

can be seen from the significant DIFFTEST, this substantially worsened the model fit (see Model 7 in Table 2). Subsequent inspections based on the MIs revealed that the means of Intellectual (MI = 163.35) and Sensual OE (MI = 22.63) differed across groups with the gifted group scoring higher on both OE domains. The factor variances, covariances, and factor means are shown in Table 3.

Measurement Invariance Across Gender

Based on the same well-fitting ESEM-Within-CFA model that was fitted on the total sample, measurement invariance was tested across gender groups using the same succession of nested models. Table 4 presents the goodness-of-fit statistics and the DIFFTEST. The first model (i.e., the configural invariant model), showed good fit statistics: RMSEA was well below .05 and CFI and TLI were above .90.

In the next model, weak MI was tested. This resulted in better fit indices (see Model 2 in Table 4). However, the chi-square difference statistic (DIFFTEST) was significant. To identify the noninvariant factor loadings, the MIs were inspected. On the basis of this procedure the loadings of Item 43 on Emotional OE (MI = 16.18), Item 31 on Imaginational OE (MI = 12.18), Item 6 on

Table 3
Factor Means, Variances, and Covariances for Gifted and Nongifted Groups

Variable	*M* difference	Psychomotor OE	Imaginational OE	Intellectual OE	Sensual OE	Emotional OE
Psychomotor OE	.00	G = 1; NG = .70				
Imaginational OE	.00	.05	1			
Intellectual OE	$-.70^{**}$	$.19^{*}$	.15	1		
Sensual OE	$-.32^{**}$	−.06	$.25^{*}$	$.44^{**}$	1	
Emotional OE	.00	−.00	$.34^{**}$	$.21^{*}$	$.42^{**}$	1

Note. OE = overexcitability; G = gifted; NG = nongifted. Factor means are for the nongifted group (factor means for the gifted group were fixed at zero).
* $p < .05$. ** $p < .01$.

Table 4
ESEM-Models Testing Measurement Invariance Across Gender

Model	χ^2	*df*	*p*	CFI	TLI	RMSEA
1. Configural Invariance	2885.26	2011	<.001	.930	.915	.037
2. Weak Measurement Invariance						
χ^2 diff test with Model 1 = 323.54 (225); $p < .0001$	3013.56	2236	<.001	.938	.932	.033
2a. Partial Weak Measurement Invariance—Model 2 + $\lambda_{43\text{-emo}}$ free						
χ^2 diff test with Model 1 = 309.98 (224); $p = .0001$	2996.60	2235	<.001	.939	.933	.033
2b. Partial Weak Measurement Invariance—Model 2a + $\lambda_{31\text{-ima}}$ free						
χ^2 diff test with Model 1 = 300.33 (223); $p = .0004$	2982.21	2234	<.001	.940	.934	.032
2c. Partial Weak Measurement Invariance—Model 2b + $\lambda_{6\text{-psy}}$ free						
χ^2 diff test with Model 1 = 292.59 (222); $p = .0010$	2973.15	2233	<.001	.941	.935	.032
2d. Partial Weak Measurement Invariance—Model 2c + $\lambda_{17\text{-emo}}$ free						
χ^2 diff test with Model 1 = 285.60 (221); $p = .0022$	2963.41	2232	<.001	.942	.936	.032
2e. Partial Weak Measurement Invariance—Model 2d + $\lambda_{30\text{-sen}}$ free						
χ^2 diff test with Model 1 = 279.40 (220); $p = .0041$	2956.12	2231	<.001	.942	.936	.032
2f. Partial Weak Measurement Invariance—Model 2e + $\lambda_{15\text{-int}}$ free						
χ^2 diff test with Model 1 = 274.63 (219); $p = .0063$	2950.44	2230	<.001	.942	.937	.032
2g. Partial Weak Measurement Invariance—Model 2f + $\lambda_{42\text{-ima}}$ free						
χ^2 diff test with Model 1 = 268.94 (218); $p = .0106$	2943.69	2229	<.001	.943	.937	.032
2h. Partial Weak Measurement Invariance—Model 2g + $\lambda_{44\text{-int}}$ free						
χ^2 diff test with Model 1 = 264.36 (217); $p = .0155$	2938.54	2228	<.001	.943	.938	.032
2i. Partial Weak Measurement Invariance—Model 2h + $\lambda_{14\text{-psy}}$ free						
χ^2 diff test with Model 1 = 261.05 (216); $p = .0195$	2935.37	2227	<.001	.943	.938	.032
2j. Partial Weak Measurement Invariance—Model 2i + $\lambda_{47\text{-emo}}$ free						
χ^2 diff test with Model 1 = 256.20 (215); $p = .0284$	2930.64	2226	<.001	.944	.938	.031
2k. Partial Weak Measurement Invariance—Model 2j + $\lambda_{28\text{-sen}}$ free						
χ^2 diff test with Model 1 = 251.06 (214); $p = .0419$	2924.78	2225	<.001	.944	.938	.031
2l. Partial Weak Measurement Invariance—Model 2k + $\lambda_{6\text{-sen}}$ free						
χ^2 diff test with Model 1 = 245.46 (213); $p = .0629$	2918.30	2224	<.001	.945	.939	.031
3. Partial Strong Measurement Invariance						
χ^2 diff test with Model 2l = 163.51 (100); $p = .0001$	3036.63	2324	<.001	.943	.940	.031
3a. Partial Strong Measurement Invariance—Model 3 + $\tau_{49\text{-threshold3}}$ free						
χ^2 diff test with Model 2l = 151.33 (99); $p = .0006$	3026.08	2323	<.001	.944	.941	.031
3b. Partial Strong Measurement Invariance—Model 3a + $\tau_{48\text{-threshold4}}$ free						
χ^2 diff test with Model 2l = 143.71 (98); $p = .0018$	3019.49	2322	<.001	.944	.941	.031
3c. Partial Strong Measurement Invariance—Model 3b + $\tau_{38\text{-threshold4}}$ free						
χ^2 diff test with Model 2l = 137.87 (97); $p = .0041$	3014.66	2321	<.001	.945	.942	.031
3d. Partial Strong Measurement Invariance—Model 3c + $\tau_{38\text{-threshold3}}$ free						
χ^2 diff test with Model 2l = 128.94 (96); $p = .0140$	3008.33	2320	<.001	.945	.942	.030
3e. Partial Strong Measurement Invariance—Model 3d + $\tau_{48\text{-threshold3}}$ free						
χ^2 diff test with Model 2l = 122.95 (95); $p = .0284$	3002.86	2319	<.001	.945	.942	.030
3f. Partial Strong Measurement Invariance—Model 3e + $\tau_{37\text{-threshold3}}$ free						
χ^2 diff test with Model 2l = 117.99 (94); $p = .0477$	2998.96	2318	<.001	.946	.943	.030
3g. Partial Strong Measurement Invariance—Model 3f + $\tau_{37\text{-threshold4}}$ free						
χ^2 diff test with Model 2l = 111.92 (93); $p = .0883$	2994.60	2317	<.001	.946	.943	.030
4. Partial Strict Measurement Invariance						
χ^2 diff test with Model 3g = 106.65 (50); $p < .0001$	3063.85	2367	<.001	.944	.942	.030
4a. Partial Strict Measurement Invariance—Model 4 + δ_{33} free						
χ^2 diff test with Model 3g = 91.86 (49); $p = .0002$	3049.32	2366	<.001	.945	.944	.030
4b. Partial Strict Measurement Invariance—Model 4a + δ_{44} free						
χ^2 diff test with Model 3g = 79.16 (48); $p = .0031$	3034.78	2365	<.001	.947	.945	.030
4c. Partial Strict Measurement Invariance—Model 4b + δ_{37} free						
χ^2 diff test with Model 3g = 73.03 (47); $p = .0088$	3029.87	2364	<.001	.947	.945	.030
4d. Partial Strict Measurement Invariance—Model 4c + δ_{48} free						
χ^2 diff test with Model 3g = 66.43 (46); $p = .0259$	3023.38	2363	<.001	.947	.945	.030
4e. Partial Strict Measurement Invariance—Model 4e + δ_{38} free						
χ^2 diff test with Model 3g = 60.60 (45); $p = .0601$	3017.57	2362	<.001	.948	.946	.029
5. Factor variance invariance						
χ^2 diff test with Model 4e = 3.25 (5); $p = .6614$	2997.16	2367	<.001	.950	.948	.029
6. Factor covariance invariance						
χ^2 diff test with Model 5 = 26.67 (10); $p = .0029$	2985.90	2377	<.001	.951	.950	.028
6a. Partial Factor covariance invariance—Model 6 + φ_{32} free						
χ^2 diff test with Model 5 = 16.15 (9); $p = .0638$	2959.09	2376	<.001	.953	.952	.028
7. Factor mean invariance						
χ^2 diff test with Model 6a = 104.71 (5); $p < .0001$	3364.06	2381	<.001	.922	.919	.036

Table 4 (*continued*)

Model	χ^2	*df*	*p*	CFI	TLI	RMSEA
7a. Partial Factor mean invariance—Model 7 + ξ_{emo} free χ^2 diff test with Model 6a = 44.99 (4); $p < .0001$	3012.51	2380	<.001	.950	.948	.029
7b. Partial Factor mean invariance—Model 7a + ξ_{int} free χ^2 diff test with Model 6a = 30.34 (3); $p < .0001$	2992.58	2379	<.001	.951	.950	.028
7c. Partial Factor mean invariance—Model 7b + ξ_{psy} free χ^2 diff test with Model 6a = 15.00 (2); $p = .0006$	2973.87	2378	<.001	.952	.951	.028
7d. Partial Factor mean invariance—Model 7c + ξ_{sen} free χ^2 diff test with Model 6a = 2.06 (1); $p = .1511$	2960.65	2377	<.001	.953	.952	.028

Note. ESEM = exploratory structural equation modeling; CFI = comparitive fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error of approximation; psy = psychomotor; sen = sensual; ima = imaginational; int = intellectual; emo = emotional; diff = difference.

Psychomotor OE (MI = 8.77), Item 17 on Emotional OE (MI = 8.33), Item 30 on Sensual OE (MI = 8.07), Item 15 on Intellectual OE (MI = 7.68), Item 42 on Imaginational OE (MI = 6.29), Item 44 on Intellectual OE (MI = 6.03), Item 14 on Psychomotor OE (MI = 5.87), Item 47 on Emotional OE (MI = 5.79), and Items 28 (MI = 5.96) and 30 on Sensual OE (MI = 6.22) were identified as noninvariant.

In a next step, we tested for partial strong MI (i.e., Model 3). All fit indices improved in comparison to the partial weak MI model. However, the DIFFTEST was again statistically significant. An inspection of the MIs showed that this was due to the third (MI = 8.74) and fourth (MI = 5.49) threshold of Item 49, the third (MI = 4.53) and fourth (MI = 6.51) threshold of Item 38, the third threshold of Item 48 (MI = 4.71), and the third (MI = 3.44) and fourth (MI = 4.71) threshold of Item 37.

Subsequently, partial strict measurement invariance was tested (i.e., Model 4). Again, model fit was good, and compared to the partial strong MI model the approximate fit indices were identical or even better (see Table 4). However, the DIFFTEST was significant. The MIs revealed that the unique variances of Items 33 (MI = 14.42), 44 (MI = 11.45), 37 (MI = 7.08), 48 (MI = 6.91), and 38 (MI = 5.57) were smaller for boys than for girls.

Given that partial strong (and even partial strict) invariance can be upheld, differences in factor variances, factor covariances, and factor means between the gender groups were tested. To this end, we first added an equality constraint on the factor variances to the partial strict invariance model (i.e., Model 5 in Table 4). As can be seen from the DIFFTEST, this additional constraint did not worsen the model fit. Next, we further restricted the model by putting an additional invariance constraint on the factor covariances (i.e., Model 6). From this test it appeared that the covariance between the Imaginational OE and the Psychomotor OE was noninvariant across the gender groups, with both OEs being negatively related in boys but not in girls. Finally, we examined whether the factor means were equal across gender by testing whether the previous model with the partial invariant factor variances and covariances fitted the data significantly better than a model with partial invariant factor variances, covariances, and invariant factor means. As can be seen in Table 4 (Model 6), this is the case, which implies that the means are not equal for boys and girls. To exactly locate those differences, we inspected the MIs, and these revealed that the means of Emotional (MI = 198.52), Intellectual (MI = 16.13), Psychomotor (MI = 16.07), and Sensual OE (MI = 14.68) were noninvariant across gender, with girls scoring higher on Emotional and Sensual OE, and boys scoring higher on Intellectual and Psychomotor OE (see Table 5).

Conclusions and Discussion

The first purpose of this study was to assess the factorial validity of the OEQ-II scores by means of the ICM and ESEM. Only the ESEM-solution yielded fit statistics that were in support of the hypothesized five-factor structure of the OEQ-II scores. That the ICM was not able to fit the data adequately did not come as a surprise. Because the independent clusters model is too restrictive for the complex reality of personality research, researchers have repeatedly experienced difficulties with getting a good fit when personality structure models were evaluated by ICMs (Church & Burke, 1994; Marsh, 2007; Marsh et al., 2009).

The second purpose of this study was to assess measurement invariance of the OEQ-II scores across intelligence as well as across gender. Our results revealed partial strict measurement invariance across intelligence (i.e., three out of 50 items showed larger unique variances for the nongifted group, and only four out

Table 5
Factor Means, variances, and Covariances For Girls and Boys

Variable	*M* difference	Psychomotor OE	Imaginational OE	Intellectual OE	Sensual OE	Emotional OE
Psychomotor OE	$.40^{**}$	1				
Imaginational OE	.00	♀ = .11; ♂ = $-.24^{**}$	1			
Intellectual OE	$.48^{**}$	$.18^{*}$	$.37^{**}$	1		
Sensual OE	$-.34^{**}$	.05	$.23^{*}$	$.43^{**}$	1	
Emotional OE	-1.05^{**}	.16	$.33^{**}$	$.50^{**}$	$.39^{**}$	.1

Note. OE = overexcitability. Factor means are for boys (factor means for girls were fixed at zero).
* $p < .05$. ** $p < .01$.

of 250 factor loadings were noninvariant), and partial strict measurement invariance across gender (i.e., five items showed larger unique variances for girls than for boys, seven thresholds out of 200 were noninvariant, and only 12 out of 250 factor loadings were noninvariant). Because the proportion of noninvariant parameters is relatively small, these noninvariant parameters did not heavily affect the group comparisons (Byrne et al., 1989; Cheung & Rensvold, 1998; Hofmans, Pepermans, & Loix, 2009; Millsap & Kwok, 2004; Yoo, 2002).

These between-group comparisons revealed that the gifted group scored higher on Intellectual and Sensual OE than the nongifted group, a finding that has consistently been found in previous research (Bouchet & Falk, 2001; Chang, 2001; Chavez, 2004; Pardo de Santayana Sanz, 2006; Siu, 2010; Tieso, 2007; Wirthwein & Rost, 2011). As such, our study shows that these differences are real, rather than the result of measurement noninvariance of the OEQ-II. At the same time, for the other OE domains, mixed findings have been reported in the literature. Our results suggest that these inconsistencies may have been the result of the failure to distinguish genuine OE differences from noninvariance.

Regarding gender differences, our results are fully in line with the findings of previous studies, which did not took the possibility of differential item functioning into account (Bouchet & Falk, 2001; Gross et al., 2007; Miller et al., 2009; Moon & Montgomery, 2005; Tieso, 2007; Treat, 2006). In particular, girls showed higher levels of Emotional and Sensual OE, whereas boys scored higher on Psychomotor and Intellectual OE. Thus, it appears that these gender differences are robust. Given the substantial difference in factor means between gender groups, our results clearly point to the need for constructing separate norm scores for both groups.

One limitation of this study is our choice of operationalizing giftedness by using cutoff scores. Perhaps we would have found different results when working with a sample of students who had been clinically identified as being gifted prior to the study. A second limitation is that the same sample was used to obtain the structure of OE and also to test measurement invariance across gender and intellectual level. Cross-validation based on other independent samples would be needed to strengthen our conclusions. Additionally, although our sample turned out to be quite representative with respect to socioeconomic and demographic features, it was a convenience sample. Finally, we relied on the DIFFTEST to evaluate whether there was a significant difference between the nested models. Whereas this chi-square difference test is widely used, it is also known to suffer from the same problems as the traditional chi-square test, that is, it is overly sensitive to minor misspecifications of the model (Hofmans, Dries, & Pepermans, 2008; Marsh et al., 2013). For this reason, alternative decision rules based on the fit indices were developed (e.g., $\Delta\text{CFI} \leq .01$ and $\Delta\text{RMSEA} \leq .015$; Cheung & Rensvold, 2002; Chen, 2007). From Tables 2 and 4, it can be seen that use of these cutoff values would lead to the identification of less (or even no) noninvariant items. Because we chose to be very conservative when testing for measurement invariance, the reported number of noninvariant items should probably be considered as an "upper limit" of the actual number of noninvariant items.

References

Ackerman, C. M. (1997). Identifying gifted adolescents using personality characteristics: Dabrowski's overexcitabilities. *Roeper Review: A Journal on Gifted Education, 19,* 229–236. doi:10.1080/02783199709553835

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16,* 397–438. doi:10.1080/10705510903008204

Bouchard, L. L. (2004). An instrument for the measure of Dabrowskian overexcitabilities to identify gifted elementary students. *Gifted Child Quarterly, 48,* 339–350. doi:10.1177/001698620404800407

Bouchet, N. M., & Falk, R. F. (2001). The relationship among giftedness, gender, and overexcitability. *Gifted Child Quarterly, 45,* 260–267. doi:10.1177/001698620104500404

Braham, A., & Andries, C. (2004). *Overexcitability Questionnaire II: Nederlandstalige versie*. Unpublished manuscript, Vrije Universiteit Brussel, Faculty of Psychology and Educational Sciences, Department of Clinical and Lifespan Psychology, Brussels, Belgium.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456–466. doi:10.1037/0033-2909.105.3.456

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97,* 404–431. doi:10.1037/0033-295X.97.3.404

Chang, H. J. (2001). *A research on the overexcitability traits of gifted and talented students in Taiwan* (Unpublished master's thesis). National Taiwan Normal University, Taipei, Taiwan.

Chavez, R. A. (2004). *Integral evaluation of the creative personality: Phenomenology, clinical and genetics* (Unpublished doctoral dissertation). National Autonomous University of Mexico, UNAM, Mexico City, Mexico.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14,* 464–504. doi:10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (1998). Cross-cultural comparisons using non-invariant measurement items. *Applied Behavioral Science Review, 6,* 93–110. doi:10.1016/S1068-8595(99)80006-3

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233–255. doi:10.1207/S15328007SEM0902_5

Church, A. T., & Burke, P. J. (1994). Exploratory and confirmatory tests of the Big Five and Tellegen's three-and four-dimensional models. *Journal of Personality and Social Psychology, 66,* 93–114. doi:10.1037/0022-3514.66.1.93

Dabrowski, K. (1970). *Mental growth through positive disintegration*. London, England: Gryf.

Dabrowski, K. (1972). *Psychoneurosis is not an illness*. London, England: Gryf.

Daniels, S., & Piechowski, M. M. (2009). Embracing intensity: Overexcitability, sensitivity and the developmental potential of the gifted. In S. Daniels & M. M. Piechowski (Eds.), *Living with intensity: Understanding sensitivity, excitability, and emotional development in gifted children, adolescents, and adults* (pp. 3–17). Scottsdale, AZ: Great Potential Press.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods, 2,* 292–307. doi:10.1037/1082-989X.2.3.292

Falk, R. F., Lind, S., Miller, N. B., Piechowski, M. M., & Silverman, L. K. (1999). *The Overexcitability Questionnaire—Two (OEQ-II): Manual, scoring system, and questionnaire*. Denver, CO: Institute for the Study of Advanced Development.

Falk, R. F., Manzanero, J. B., & Miller, N. B. (1997). Developmental potential in Venezuelan and American artists: A cross-cultural validity study. *Creativity Research Journal, 10,* 201–206.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9,* 466–491. doi:10.1037/1082-989X.9.4.466

Gallagher, S. A. (1985). A comparison of the concept of overexcitabilities with measures of creativity and school achievement in sixth grade students. *Roeper Review: A Journal on Gifted Education, 8,* 115–119. doi:10.1080/02783198509552950

Gross, C. M., Rinn, A. N., & Jamieson, K. M. (2007). Gifted adolescents' overexcitabilities and self-concepts: An analysis of gender and grade level. *Roeper Review: A Journal on Gifted Education, 29,* 240–248. doi:10.1080/02783190709554418

Hofmans, J., Dries, N., & Pepermans, R. (2008). The Career Satisfaction Scale: Response bias among men and women. *Journal of Vocational Behavior, 73,* 397–403. doi:10.1016/j.jvb.2008.08.001

Hofmans, J., Pepermans, R., & Loix, E. (2009). Measurement invariance matters: A case made for the ORTOFIN. *Journal of Economic Psychology, 30,* 667–674. doi:10.1016/j.joep.2009.06.002

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. doi:10.1080/10705519909540118

Lievens, F., & Anseel, F. (2004). Confirmatory factor analysis and invariance of an organizational citizenship behaviour measure across samples in a Dutch-speaking context. *Journal of Occupational and Organizational Psychology, 77,* 299–306. doi:10.1348/0963179041752727

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83,* 1198–1202. doi:10.1080/01621459.1988.10478722

Lovecky, D. V. (2004). *Different minds: Gifted children with AD/HD, Asperger syndrome, and other learning deficits*. London, England: Jessica Kingsley.

Lysy, K. Z., & Piechowski, M. M. (1983). Personal growth: An empirical study using Jungian and Dabrowskian measures. *Genetic Psychology Monographs, 108,* 267–320.

Marsh, H. W. (2007). Application of confirmatory factor analysis and structural equation modeling in sport/exercise psychology. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of on sport psychology* (3rd ed., pp. 774–798). New York, NY: Wiley. doi:10.1002/9781118270011.ch35

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the Big-Five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22,* 471–491. doi:10.1037/a0019227

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling, 16,* 439–476. doi:10.1080/10705510903008220

Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Measurement invariance of Big-Five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology, 49,* 1194–1218. doi:10.1037/a0026913

Mendaglio, S. (2003). Heightened multifaceted sensitivity of gifted students: Implications for counseling. *The Journal of Secondary Gifted Education, 14,* 72–82.

Mendaglio, S., & Tillier, W. (2006). Dabrowski's theory of positive disintegration and giftedness: Overexcitability research findings. *Journal for the Education of the Gifted, 30,* 68–87.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525–543. doi:10.1007/BF02294825

Miller, N. B., Falk, R. F., & Huang, Y. (2009). Gender identity and the overexcitability profiles of gifted college students. *Roeper Review: A Journal on Gifted Education, 31,* 161–169. doi:10.1080/02783190902993920

Miller, N. B., Silverman, L. K., & Falk, R. F. (1994). Emotional development, intellectual ability, and gender. *Journal for the Education of the Gifted, 18,* 20–38. doi:10.1177/016235329401800103

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Routledge.

Millsap, R. E., & Kwok, O. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods, 9,* 93–115. doi:10.1037/1082-989X.9.1.93

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Mutivariate Behavioral Research, 39,* 479–515. doi:10.1207/S15327906MBR3903_4

Moon, J., & Montgomery, D. (2005). Profiles of overexcitabilities for Korean high school gifted students according to gender and domain of study. *Journal of Gifted/Talented Education, 15,* 1–10.

Morin, A. J. S., Marsh, H. W., & Nagengast, B. (2013). Chapter 10: Exploratory structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed.). Charlotte, NC: Information Age.

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38,* 171–189. doi:10.1111/j.2044-8317.1985.tb00832.x

Muthén, L. K., & Muthén, B. (2010). *Mplus user's guide 6.1.* Los Angeles, CA: Muthén & Muthén.

Pardo de Santayana Sanz, R. (2006). *The gifted student and learning disabilities: Validation of the OEQ-II as a diagnosis test*. Madrid, Spain: Universidad Complutense de Madrid.

Piechowski, M. M. (1979). Developmental potential. In N. Colangelo & R. T. Zaffrann (Eds.), *New voices in counseling the gifted* (pp. 25–57). Dubuque, IA: Kendall/Hunt.

Piechowski, M. M. (1986). The concept of developmental potential. *Roeper Review: A Journal on Gifted Education, 8,* 190–197. doi:10.1080/02783198609552971

Piechowski, M. M., & Colangelo, N. (1984). Developmental potential of the gifted. *Gifted Child Quarterly, 28,* 80–88. doi:10.1177/001698628402800207

Piechowski, M. M., & Cunningham, K. (1985). Patterns of overexcitability in a group of artists. *The Journal of Creative Behavior, 19,* 153–174. doi:10.1002/j.2162-6057.1985.tb00655.x

Piechowski, M. M., & Miller, N. (1995). Assessing developmental potential in gifted children: A comparison of methods. *Roeper Review: A Journal on Gifted Education, 17,* 176–180. doi:10.1080/02783199509553654

Piechowski, M. M., Silverman, L. K., & Falk, R. F. (1985). Comparison of intellectually and artistically gifted on five dimensions of mental functioning. *Perceptual and Motor Skills, 60,* 539–549. doi:10.2466/pms.1985.60.2.539

Piirto, J., Montgomery, D., & May, J. (2008). A comparison of Dabrowski's overexcitabilities by gender for American and Korean high school gifted students. *High Ability Studies, 19,* 141–153. doi:10.1080/13598130802504080

Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual: Standard progressive matrices*. Oxford, England: Oxford Psychologists Press.

Schiever, S. W. (1985). Creative personality characteristics and dimensions of mental functioning in gifted adolescents. *Roeper Review: A Journal on Gifted Education, 7,* 223–226. doi:10.1080/02783198509552901

Silverman, L. K. (1997). Family counseling with the gifted. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (2nd ed., pp. 382–397). Boston, MA: Allyn & Bacon.

Silverman, L. K. (2008). The theory of positive disintegration in the field of gifted education. In S. Mendaglio (Ed.), *Dabrowski's theory of positive disintegration* (pp. 157–173). Scottsdale, AZ: Great Potential Press.

Silverman, L. K., & Ellsworth, B. (1981). The theory of positive disintegration and its implications for giftedness. In N. Duda (Ed.), *Theory of positive disintegration: Proceedings of the third international conference* (pp. 179–194). Miami, FL: University of Miami.

Siu, A. F. Y. (2010). Comparing overexcitabilities of gifted and non-gifted school children in Hong Kong: Does culture make a difference? *Asia Pacific Journal of Education, 30,* 71–83. doi:10.1080/02188790903503601

Sternberg, R. J. (2004). Introduction to definitions and conceptions of giftedness. In S. M. Reis & R. J. Sternberg (Eds.), *Definitions and conceptions of giftedness* (pp. xxiii–xxvi). Thousand Oaks, CA: Corwin Press.

Sternberg, R. J., Jarvin, L., & Grigorenko, E. L. (2011). *Explorations in giftedness*. New York, NY: Cambridge University Press.

Tieso, C. L. (2007). Patterns of overexcitabilities in identified gifted students and their parents: A hierarchical model. *Gifted Child Quarterly, 51,* 11–22. doi:10.1177/0016986206296657

Treat, A. R. (2006). Overexcitability in gifted sexually diverse populations. *Journal of Secondary Gifted Education, 17,* 244–257.

Wirthwein, L., Becker, C. V., Loehr, E., & Rost, D. H. (2011). Overexcitabilities in gifted and non-gifted adults: Does sex matter? *High Ability Studies, 22,* 145–153. doi:10.1080/13598139.2011.622944. doi:10.1080/13598139.2011.622944

Wirthwein, L., & Rost, D. H. (2011). Focussing on overexcitabilities: Studies with intellectually gifted and academically talented adults. *Personality and Individual Differences, 51,* 337–342. doi:10.1016/j.paid.2011.03.041

Yakmaci-Guzel, B., & Akarsu, F. (2006). Comparing overexcitabilities of gifted and non-gifted 10th grade students in Turkey. *High Ability Studies, 17,* 43–56. doi:10.1080/13598130600947002

Yoo, B. (2002). Cross-group comparisons: A cautionary note. *Psychology & Marketing, 19,* 357–368. doi:10.1002/mar.10014

Appendix

OEQ-II Inventory

Directions: Please rate how much each statement fits you. Respond on the basis of what you are like now, *not* how you would like to be or how you think you should be. *Circle the number* under the statement that most accurately reflects the way you see yourself.

	Not at All Like Me	Not Much Like Me	Some-what Like Me	A Lot Like Me	Very Much Like Me
1. I like to daydream.	1	2	3	4	5
2. I am a competitive person.	1	2	3	4	5
3. The varieties of sound and color are delightful.	1	2	3	4	5
4. My pretend world is very real to me.	1	2	3	4	5
5. I am an independent thinker.	1	2	3	4	5
6. I feel other people's feelings.	1	2	3	4	5
7. If an activity is physically exhausting, I find it satisfying.	1	2	3	4	5
8. Viewing art is a totally absorbing experience.	1	2	3	4	5
9. I worry a lot.	1	2	3	4	5
10. I love to be in motion.	1	2	3	4	5
11. It makes me sad to see a lonely person in a group.	1	2	3	4	5
12. I can take difficult concepts and translate them into something more understandable.	1	2	3	4	5
13. I get great joy from the artwork of others.	1	2	3	4	5
14. When I get bored, I begin to daydream.	1	2	3	4	5
15. When I have a lot of energy, I want to do something really physical.	1	2	3	4	5
16. I question everything--how things work, what things mean, why things are the way they are.	1	2	3	4	5
17. I can be so happy that I want to laugh and cry at the same time.	1	2	3	4	5
18. I am more energetic than most people my age.	1	2	3	4	5
19. I can form a new concept by putting together a number of different things.	1	2	3	4	5
20. Sometimes I pretend I am someone else.	1	2	3	4	5
21. The longer that I have to sit still, the more restless I get.	1	2	3	4	5
22. Things that I picture in my mind are so vivid that they seem real to me.	1	2	3	4	5
23. I observe and analyze everything.	1	2	3	4	5
24. I find myself mixing truth and fantasy in my thoughts.	1	2	3	4	5
25. Theories get my mind going.	1	2	3	4	5
26. I have strong feelings of joy, anger, excitement, and despair.	1	2	3	4	5
27. I feel music throughout my whole body.	1	2	3	4	5
28. I enjoy exaggerating reality.	1	2	3	4	5
29. I feel like my body is constantly in motion.	1	2	3	4	5
30. I love to solve problems and develop new concepts.	1	2	3	4	5
31. I am deeply concerned about others.	1	2	3	4	5
32. I delight in colors, shapes, and textures of things more than other people do.	1	2	3	4	5
33. I believe that dolls, stuffed animals, or the characters in books are alive and have feelings.	1	2	3	4	5
34. Words and sounds create unusual images in my mind.	1	2	3	4	5
35. My strong emotions move me to tears.	1	2	3	4	5
36. I like to dig beneath the surface of issues.	1	2	3	4	5
37. I am moved by beauty in nature.	1	2	3	4	5
38. I am not sensitive to the color, shape, and texture of things like some people are.	1	2	3	4	5
39. When I am nervous, I need to do something physical.	1	2	3	4	5
40. I try to analyze my thoughts and actions.	1	2	3	4	5
41. I can feel a mixture of different emotions all at once.	1	2	3	4	5
42. I am the type of person who has to be active—walking, cleaning, organizing, doing something.	1	2	3	4	5

(*Appendix continues*)

Appendix (*continued*)

	Not at All Like Me	Not Much Like Me	Some-what Like Me	A Lot Like Me	Very Much Like Me
43. I like to play with ideas and try to think about how to put them to use.	**1**	**2**	**3**	**4**	**5**
44. I am an unemotional person.	**1**	**2**	**3**	**4**	**5**
45. I enjoy the sensations of colors, shapes, and designs.	**1**	**2**	**3**	**4**	**5**
46. The difference in aromas is interesting.	**1**	**2**	**3**	**4**	**5**
47. I have a talent for fantasy.	**1**	**2**	**3**	**4**	**5**
48. I love to listen to the sounds of nature.	**1**	**2**	**3**	**4**	**5**
49. I take everything to heart.	**1**	**2**	**3**	**4**	**5**
50. I thrive on intense physical activity, e.g. fast games and sports.	**1**	**2**	**3**	**4**	**5**

Note. From *The Overexcitability Questionnaire–Two (OEQ-II): Manual, Scoring System, and Questionnaire*, by R. F. Falk, S. Lind, N. B. Miller, M. M. Piechowski, & L. K. Silverman, 1999, Denver, CO: Institute for the Study of Advanced Development. Copyright 1999 by Institute for the Study of Advanced Development. Reprinted with permission.

Received February 1, 2012
Revision received June 19, 2013
Accepted July 5, 2013 ■