Running head: Factor Validity of the Adult Symptom Rating Scale

Factorial Validity of the ADHD Adult Symptom Rating Scale in a French community sample: Results from the ChiP-ARD study

Alexandre J. S. MORIN[1*], Antoine TRAN[2], Hervé CACI[2]

[*]Corresponding author
[1] Centre for Positive Psychology and Education
University of Western Sydney
Bankstown Campus, Building 19
Locked Bag 1797, Penrith, NSW 2751, Australia
Phone : +61 2 9772 6130
Fax : +61 2 9772 6193
http://www.uws.edu.au/cppe
A.Morin@uws.edu.au

[2]Hôpitaux Pédiatrique de Nice – CHU Lenval
57, avenue de la Californie
06200 Nice

This is the prepublication version of the following manuscript:

Morin, A.J.S., Tran, A., & Caci, H. (2016). Factorial Validity of the ADHD Adult Symptom Rating Scale in a French community sample: Results from the ChiP-ARD study. *Journal of Attention Disorders*, *20* (6), 530-541. doi: 10.1177/1087054713488825

Acknowledgements

This study was founded by a grant from the French Ministry of Health awarded to the last author (*Programme Hospitalier de Recherche Clinique National 2009*). The study is recorded on clinicaltrials.gov under NCT01260792. The authors are grateful to Jacques BOUCHEZ, MD and Professor Franck J. BAYLÉ, MD for their help in translating the ASRS v1.1 into French, to Eric FONTAS, MD for his help in setting up the study procedures, to Mrs Vanina OLIVERI and Mr Kevin DOLLET for their sustained efforts to monitor the study and collect the data, to the *Inspection Académique des Alpes-Maritimes* and to the *Rectorat des Alpes-Maritimes et du Var* for their valuable support, and to the teachers, pupils and parents for participating in this study.

Abstract

Objectives: Recent publications reported that a bifactor model better represented the underlying structure of ADHD than classical models, at least in youth. The Adult ADHD Symptoms Rating Scale (ASRS) has been translated in many languages but a single study compared its structure in adults across DMS-IV and ICD-10 classifications. **Method:** We investigated the factor structure, reliability, and measurement invariance of the ASRS among a community sample of 1,171 adults. **Results:** Results support a bifactor model including one general ADHD factor and three specific Inattention, Hyperactivity and Impulsivity factors corresponding to ICD-10, albeit the Impulsivity specific factor was weakly defined. Results also support the complete measurement invariance of this model across gender and age groups, and that men have higher scores than women on the ADHD G-factor but lower scores on all three S-factors. **Conclusion:** Results suggest that a total ASRS-ADHD score is meaningful, reliable and valid in adults.

Keywords: ADHD, Adult, Rating scale, Psychometrics, Bifactor model.

Attention-Deficit with Hyperactivity (ADHD) is one of the most frequent disorder in children and adolescents with a worldwide estimated prevalence of 5.29% (Polanczyk, Silva de Lima, Lessa Horta, Biederman, & Rohde, 2007). According to DSM-IV, ADHD encompasses a number of pervasive and impairing symptoms including severe problems of inattention and/or hyperactivity and impulsivity (American Psychiatric Association, 1994). To meet diagnostic criteria, at least some of the symptoms must have appeared before age 7 and the main symptoms must be present for more than 6 months, be severe enough to cause significant difficulties, and be documented in at least two settings (e.g. home and school for children, or home and work for adults). The World Health Organization (WHO) (Word Health Organization, 1993) uses the name of Hyperkinetic Disorder (HKD) and lists similar criteria for the disorder, but has a more stringent diagnostic algorithm leading to a lower prevalence of HKD than ADHD (Polanczyk, et al., 2007). The DSM-IV distinguishes three subtypes of ADHD according to whether the predominant symptoms are characterized by Inattention, Hyperactivity-Impulsivity, or both (American Psychiatric Association, 1994). Some studies even suggest that the Predominantly Inattentive subtype may constitute a distinct disorder (Caci, Morin, & Tran, 2013; Capdevila-Brophy et al., 2013; Solanto, 2000).

Although the symptoms related to Hyperactivity and Impulsivity are considered to be part of the same subtype of ADHD, they are defined as two distinct dimensions of HKD. In this regard, the DSM-IV and ICD-10 diagnostic systems also differ regarding the allocation of the symptom “Often talks excessively” (DSM-IV) or “Often talks excessively without appropriate response to social constraints” (ICD-10), to either the hyperactivity dimension in the DSM-IV or to the impulsivity dimension in the ICD-10.

Although it is generally acknowledged by both diagnostic systems that ADHD and HKD symptoms are pervasive and often tend to persist well into adulthood (Faraone, Biederman, & Mick, 2006; Kooij et al., 2010) neither the DSM-IV nor the ICD-10 provides specific criteria for adults (Barkley, Murphy, & Fischer, 2008). Indeed, research shows that ADHD/HKD tends to persist at adulthood in at least two thirds of the cases (Faraone, et al., 2006; Kooij, et al., 2010) and its prevalence is estimated at 3.4% in adults of the general population (Fayyad et al., 2007; Kooij, et al., 2010). This condition is associated with a broad range of negative life outcomes but is also responsive to treatment (Hodgkins et al., 2012; Shaw et al., 2012), pointing the need for efficient screening and diagnosis procedures. Illustrative of this need, the European Network for Adult ADHD recently proposed the DIVA, a semi-structured interview for ADHD, adapted in many languages and available free of charge for clinicians and researchers (Foundation, 2012). However, semi-structured interviews are costly to use in epidemiological research and not well suited to large scale screening procedures.

In 2000, the World Health Organization (WHO) launched the new World Mental Health (WMH) Initiative surveys that included a new questionnaire specifically designed to rigorously assess ADHD in adults. Admittedly, retrospective assessments of childhood ADHD and self-report measures of adult ADHD did exist at that time but were deemed inadequate to assess all 18 symptoms of ADHD criterion A. This new measure, the Adult ADHD Symptoms Self-Report (ASRS), includes 18 items, rated on a 0 (*Never*) to 4 (*Very often*) scale, and provides a comprehensive coverage of ADHD diagnostic criteria. A quota subsample of 154 respondents from the US National Comorbidity Survey Replication (NCS-R) (Kessler et al., 2004) was re-interviewed to assess the performance of the ASRS after thorough weighting of the sample to make it representative of the general population of adults between 18 and 44 years-old (Kessler et al., 2005). This study supported the adequacy of the ASRS as a screening tool for ADHD. In practice, the ASRS items are generally dichotomized: 11 items were scored 1 if the response was *Often* or *Very Often*, and the remaining 7 items were scored 1 if the response was *Sometimes*, *Often* or *Very Often*. However, in research the practice of dichotomizing continuous or ordered-categorical variables including more than two answer categories is known to result in loss of statistical power and measurement accuracy, and to severely change the distributional characteristics of the items (MacCallum, Zhang, Preacher, & Rucker, 2002).

In the last two decades, the ASRS has been successfully used in epidemiological studies yielding computation of mean cross-national prevalence rates in adults that reaches 3.5% (inter-quartile range: 1.3-4.9%) (de Graaf et al., 2008; Fayyad, et al., 2007; Kessler et al., 2006; Kessler et al., 2007). The ASRS has also been used in clinical samples with comorbid conditions (Adler, Guida, Irons, Rotrosen, & O'Donnell, 2009), in clinical trials to monitor changes in adults, and, more recently, in adolescents (Adler & Newcorn, 2011; Adler et al., 2012). Surprisingly, no study to date has attempted to estimate

the underlying measurement model (factor structure) of the ASRS. Clearly, this represents an important limitation of current research on the ASRS as this instrument is usually assumed to closely follow the inherent underlying structure of the DSM-IV and ICD-10 conceptualisations, and utilized as if it did in fact follow this structure.

The fact that the underlying structure of ADHD has recently been questioned reinforces this limitation. More specifically, ADHD is known to represent a relatively stable condition that persists well into adulthood. However, its specific manifestations may change over the course of development (Faraone, et al., 2006). This suggests that there might be a generic (G) component of ADHD that lies at the core of this condition and is stable over time, with remaining specific (S) developmental and symptomatic manifestations that fluctuates over time and contexts (Martel, von Eye, & Nigg, 2010). This distinction is also consistent with the way ADHD is defined in the DSM-IV and ICD-10, with a core G set of ADHD manifestations leading to the main diagnosis, but specificities leading individuals to fit more closely to the inattentive, hyperactive-impulsive, or combined subtypes. Within the Confirmatory Factor Analysis (CFA) framework, a bifactor model (Holzinger & Swineford, 1937) whereby each item is simultaneously allowed to be defined (through factor loadings) by one generic G ADHD factor and one subtype-specific S factor (i.e., inattention or hyperactivity/impulsivity) would be particularly well-suited to investigate this possibility. A bifactor model postulates that an *f*-factor solution exists for a set of *n* items with one G factor and *f*-1 S factors. The items' loadings on the G factor and on one of *f*-1 S factors are estimated; other loadings are constrained to be zero. All factors are set to be orthogonal (i.e., correlations are constrained to be zero). Such a model first analyses the total covariance among the items to extract a global G component underlying all items, and then the residual covariance not explained by the G factor is modelled through the specific S factors.

In psychiatric measurement, an important question has to do with whether a primary dimension (e.g. depression, anxiety, ADHD, etc.) does exist as a unitary disorder including specificities (i.e. as represented by a bifactor model), or whether these specificities rather define a set of distinct facets without a generic common core (i.e. represented by a classical CFA model). Similarly, another issue that deserves investigation in the case of ADHD has to do with the fact that, compared to the DSM-IV, the ICD-10 classification more clearly differentiates hyperactivity symptoms from impulsivity symptoms.

A first study to contrast classical CFA models with bifactor models in studying ADHD symptoms in children and adolescents (Dumenici, McConaughy, & Achenbach, 2004) used the Attention Problems syndrome of the Teacher's Rating Form (Achenbach & Rescorla, 2001). Results supported a bifactor solution including one ADHD G-factor and two specific (Inattention and Hyperactivity-Impulsivity) S-factors among eight separate samples (6-11 years old children, 12-18 years old adolescents, boys, girls, community and clinical samples). This work has since then been replicated by several authors using the DSM-IV symptoms rated by teachers or parents, in children and adolescents with ADHD (Toplak et al., 2009) or in community samples (Martel, et al., 2010). A recent study on teachers' rating of a community sample of children and adolescents, supported a similar bifactor structure, but showed that the Hyperactivity-Impulsivity S-factor was improperly defined calling into question the existence of a Predominantly Hyperactivity-Impulsivity subtype (Caci, et al., 2013). However, few studies assessed a bifactor structure in adults with ADHD (Gibbins, Toplak, Flora, Weiss, & Tannock, 2012), and none did so in community samples. Interestingly, studies show that, over the life span, inattention symptoms tend to have greater persistence (Faraone, et al., 2006), while hyperactivity and impulsivity symptoms present more transient characteristics (Biederman, Mick, & Faraone, 2000; Faraone, et al., 2006; Larsson, Lichenstein, & Larsson, 2006), while still playing a substantial role in diagnoses of ADHD (Katusic et al., 2005). As a corollary, the DSM-IV approach of grouping together Hyperactivity and Impulsivity symptoms has generally been supported in children and adolescents but never systematically investigated in adults. In a sample a 751 ADHD adults rated by a clinician on the ADHD-RS IV, a recent study reported that a bifactor factor structure with one ADHD G-factor and three S-factors (Inattention, Motor-Impulsivity and Verbal-Impulsivity) provided a better fit to the data than alternative models (Gibbins, et al., 2012). Interestingly, the Verbal-Impulsivity factor identified in this study perfectly corresponds to the Impulsivity factor defined by the ICD-10 diagnostic system, while the Motor-Impulsivity factor falls closer to the Hyperactivity factor. Therefore, this study aims at replicating and extending these previous findings using a self-report instrument.

Another important issue that has yet to be systematically investigated has to do with the critical assumption that the various versions of the ASRS measure the same trait in samples from distinct subpopulations among which the instrument will be used (e.g. gender groups, age groups, etc.). This property is known as measurement invariance and represents a pre-requisite to valid comparisons regarding mean levels differences, variability differences, and predictive differences between the targeted subgroups (Meredith, 1993). For instance, in order to verify whether males really do present higher levels of ADHD than females, one must first demonstrate that the measurement instrument assesses the same construct in the same manner in these two subpopulations, and thus that the estimated latent means are comparable. The same question applies to any evaluation of age-related variations in ADHD manifestation. To fully support the argument that ADHD either persists, or tends to fade out, with age, one must first demonstrate that the measurement instrument that is used to support these conclusions present measurement properties that remain unchanged by aging. A lack of invariance could lead to rather serious consequences, including erroneous treatment referrals for members of certain groups because ADHD symptoms are assessed differently in these groups.

We were able locate a single study which reported systematic tests of measurement invariance of a measure of ADHD symptoms (the ADHD-RS IV) based on a clinical sample of adults with ADHD. Interestingly, this study also found evidence supporting a bifactor representation of ADHD including one ADHD G-factor and three S-factors (Inattention, Motor-Impulsivity and Verbal-Impulsivity) (Gibbins, et al., 2012). In this study, the authors also found evidence for the configural invariance of this bifactor model across genders, as well as for its strong measurement invariance (i.e. invariance of the factor loadings and thresholds; Meredith, 1993). However, this study did not investigate the invariance of the measurement model across age-groups, and also did not investigate the invariance of the items' uniquenesses, which is worrying since the non-invariance of items' uniquenesses indicates that the constructs are assessed with different levels of measurement error and precision across subgroups and that the confidence of diagnostic decisions will differ in these groups. The present paper attempts to extend these previous results through the investigation of the full measurement invariance of the ASRS across genders, age groups, and age-by-gender groups.

In summary, this paper aims to investigate the psychometric properties of the ASRS self-rated by adults in order to conduct four specific verifications:

1. How well does the *a priori* two-factor structure of the ASRS (mimicking the DSM-IV and the ICD-10 items' arrangements) fit the data?
2. Will a bifactor model provide a better representation of ADHD self-rated by adults than a classical CFA model, as suggested by some previous studies based on ADHD symptoms?
3. Should specific symptoms of Hyperactivity and Impulsivity be treated as distinct sub-dimensions of ADHD, or as a single sub-dimension, when adults are evaluated? In the second case, a corollary question that will be investigated has to do with whether the "talks excessively" symptom (i.e. item 15 of the ASRS) corresponds more closely with the Hyperactivity or Impulsivity sub-dimension.
4. Is the ASRS a reliable instrument when self-rated by French adults?
5. Is the ASRS measurement model invariant across genders, age groups, and gender-by-age groups?

Methods

This paper is based on data from the Children and Adults with ADHD and Related Disorders (ChiP-ARD) study. This study is the largest general population study of ADHD to have been conducted in France, and one of the largest such studies to have ever been conducted in Europe. This study received the support of the Commissioner of Education and the Department of Education, complied with normative ethical prescriptions for French medical research, and the procedures used to keep paper-based and electronic data secured and anonymous were approved by the *Commission Nationale Informatique et Liberté*.

In this study, schools were randomly drawn from the list of public schools (kindergarten, elementary and secondary) located in the city of Nice and its suburbs and school principals were invited to participate and so on until a sufficient number of schools had accepted to participate to reach approximately 1,000 participants, equally distributed by age group. In these schools, teachers received a presentation of the study objectives and questionnaires before they individually accepted or declined participation. A secured web site was created for the study. Teachers could log in using a password and create a record for their class (level, number of pupils of each gender, etc.) also specifying their

age and gender. To obtain parental consents for the study, the system then randomly returned an alphabetical letter and teachers were asked to contact the parents of the children whose name began with this letter (or the next one if no name matched the first letter). A total of 2 to 4 youths were targeted from each of the selected classes. Parents of these youths had to return a signed consent form that was kept anonymous by teachers who allocated them an 8-digit unique unifier.

Data were collected in spring, during two distinct non-longitudinal waves in 2010 and 2011. A total of 20 kindergarten schools (*pré-élémentaires* or *maternelles*), 30 primary schools (*élémentaires*), 14 secondary-schools (*colleges and lycées*) from Southern France (Nice) participated in the study and teachers provided ratings of a total of 892 students (89.2% of the targeted 1,000 students). Both parents of each participating youths were asked to complete questionnaires about themselves, including the ASRS in relation to their own manifestations of ADHD. Out of 1784 questionnaires thus sent out, a total of 1,171 adults (i.e. parents of the participating youths) returned completed questionnaires and form the main sample used in this study (for a response rate of 65.6%). However, based on a priori knowledge regarding the prevalence of single-parents families, reconstituted families, and families with parents with only a limited mastery of the French language provided by the 2009 Census for France (see http://www.insee.fr/en/default.asp), the protocol predicted that only about 1.5 parent/youth would return the questionnaire (for a response rate of 87.5%).

Twelve (9 women and 3 men) did not report their month and/or year of birth preventing us to compute their age. For women (n=651; 56.45%), age is normally distributed (Shapiro-Wilk test=.997, n.s.; skewness=-.031, n.s.; kurtosis=3.326, n.s.), but not for men (n=507; Shapiro-Wilk normality test=.991; p<.005; skewness=.261, p<.02; kurtosis=3.469, p<.05). Women were younger than men: t(1023.4)=-6.85 (p<.0001, Cohen's d=.412), 40.73 (S.D. = 5.60, median = 40.75, range = 22.00-59.17) vs. 43.16 (S.D. = 6.28, median = 42.67, range = 26.17-64.67), respectively. For tests of measurement invariance based on age, the sample was split into subgroups of younger (20 to 40 years old; n = 425) and middle-age adults (40 to 65 years old; n = 733), following cut-off score commonly used in research on ageing (e.g., Haase, Seider, Shiota, & Levenson, 2012; Kray, & Lindenberger, 2000; Lang, John, Lüdtke, Schupp, & Wagner, 2011), classical theories of human development (e.g., Erikson, 1968), and willingness to keep each gender-by-age subsample reasonably large to conduct tests of measurement invariance.

The ASRS has been adapted to up to ten languages including French (Caci, Bayle, & Bouchez, 2008) but never systematically validated in French. These different linguistic versions can be downloaded at http://www.hcp.med.harvard.edu/ncs/asrs.php. The French adaptation was done following a classical translation-back-translation procedure. The resulting version was approved by the WHO as the official French version of the full ASRS (Caci, et al., 2008). It should be noted here that French is the official language in 32 countries and territories worldwide (Francophonie, 2012), including five European countries (France, Belgium, Switzerland, Monaco, and Luxembourg) and Canada, is one of the European institutions' United Nations' official languages and remains the most often taught second language worldwide, reinforcing the need to investigate the psychometric properties of this version of the ASRS.

Statistical Analyses

The main models were estimated with Mplus 7.0 (Muthén & Muthén, 2012) from polychoric correlation matrices using the robust weight least square estimator (WLSMV). WLSMV estimation has been found to outperform Maximum Likelihood with ordered-categorical items involving 5 or less answers categories such as those used in the present study (Beauducel & Herzberg, 2006; Finney & DiStefano, 2006; Flora & Curran, 2004; Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Muthén, du Toit, & Spisic, 1997).

The fit of seven a priori alternative models of answers to the ASRS was contrasted: a one-factor ADHD model (M1), a model including 2 correlated factors (Inattention and Hyperactivity-Impulsivity: M2), a model including 3 correlated factors (Inattention, Hyperactivity and Impulsivity) defined according to DSM-IV with item 15 allocated to the Hyperactivity factor (M3A), a similar a model including 3 correlated factors (Inattention, Hyperactivity and Impulsivity) defined according to ICD-10 with item 15 allocated to the Impulsivity factor (M3B), a bifactor model including one ADHD G-factor and two specific S-factors (Inattention and Hyperactivity-Impulsivity: M4), a bifactor model including one ADHD G-factor and three specific S-factors (Inattention, Hyperactivity, and Impulsivity), with item 15 allocated to the Hyperactivity factor (M5A), and one similar bifactor model

including one ADHD G-factor and three specific S-factors (Inattention, Hyperactivity, and Impulsivity), with item 15 allocated to the Impulsivity factor (M5B).

We performed measurement invariance tests across gender (male versus females), age groups (young adults versus middle-age adults) and combinations of gender and age groups in a sequential strategy following Meredith (1993) recommendations as adapted for ordered-categorical items by Millsap and Tein (Millsap & Tein, 2004; Morin et al., 2011). The sequence of tests is as follows: (i) configural invariance, (ii) metric/weak invariance (invariance of the factor loadings); (iii) scalar/strong invariance (invariance of the loadings and thresholds); (iv) strict invariance (invariance of the loadings, thresholds and uniquenesses), (v) invariance of the latent variances (invariance of the loadings, thresholds, uniquenesses and variances), and (vi) latent means invariance (invariance of the loadings, thresholds, uniquenesses, variances and latent means). It should be noted that, since bifactor models are specified as orthogonal, tests of the invariance of the latent covariances are precluded.

The fit of all models was evaluated using various indices (Hu & Bentler, 1999; Yu, 2002): the WLSMV chi-squared statistic (χ^2), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA) and the 90% confidence interval of the RMSEA. These fit indices are interpreted the same way as with ML/MLR estimation, with values greater than .90 for CFI and TLI are considered to be indicative of adequate model fit. Values smaller than .08 or .06 for the RMSEA support respectively acceptable and good model fit. In order to test for fit improvement, we used the MPlus DIFFTEST function (MD$\Delta\chi^2$; (Asparouhov & Muthén, 2006; B.O. Muthén, 2004). As the χ^2 itself, MD$\Delta\chi^2$ tends to be oversensitive to sample size and to minor model misspecifications. In this regard, and to take into account the overall number of MD$\Delta\chi^2$ tests used in this study, the significance level to identify non-invariance was fixed at $p < .01$ (Bollen, 1989; Morin, Madore, Morizot, Boudrias, & Tremblay, 2009; Rensvold & Cheung, 1998). It is also generally recommended to use additional indices to complement MD$\Delta\chi^2$ tests when comparing nested models (Chen, 2007; Cheung & Rensvold, 2002): a CFI diminution of .01 or less and a RMSEA augmentation of .015 or less between a model and the preceding model in the invariance hierarchy indicate that the measurement invariance hypothesis should not be rejected.

Results

Confirmatory Factor Analyses and Reliability

The single factor model (M1) showed the worst fit to the data (Table 1), and the two-factor (M2) model provided a suboptimal level of fit to the data according to a TLI under .90 and a RMSEA over .08. However, both three-factor (M3A and M3B) models presented a satisfactory level of fit to the data (CFI and TLI>.90; RMSEA<.08), though the model defined according to ICD-10 specifications (M3B) resulted in an appreciable increase in terms of fit to the data (ΔCFI and ΔTLI >. 10). However, in this model, the estimated correlations between the Inattention, Hyperactivity and Impulsivity factors were high enough (.520 to .641) to suggest the presence of a common core of ADHD symptoms, justifying the investigation of bifactor models.

Accordingly, the fit to the data of the *a priori* bifactor models was systematically higher than the fit of the corresponding correlated factor models (ΔCFI and ΔTLI >. 10), suggesting the superiority of a bifactor representation of ADHD. Once again, the model defined according to ICD-10 specifications (i.e., with item 15 associated with the Impulsivity factor, M5B) proved superior to the alternative specifications in terms of fit to the data and was retained as the final model for this study. Interestingly, the fit of this model was also fully satisfactory (see the lower portion of Table 1) in all possible subgroups of participants based on gender (males versus females), age groups (younger versus middle-age adults) and gender by age groups (female younger or middle-age adults, and male younger or middle-age adults) with CFI and TLI>.90 and RMSEA<.08.

Table 2 presents the parameters estimates for this final model (M5B). The ADHD G-factor is well defined with most loadings between .407 and .658, the only exception being related to the first four items associated with the Inattention S-factor, that present relatively low loadings (λ=.296 to .347) on the ADHD G-factor, and that are surprisingly all included in the 6-item ASRS screener (Kessler, et al., 2007) However, the fact that most loadings on the ADHD G-factor remain moderate in magnitude clearly indicates that substantial specificity is included in the items. In fact, the Inattention S-Factor is itself relatively very well-defined, with 7 out of 9 loadings between .405 and .660, suggesting that only 2 Inattention items mostly serve to define the generic ADHD with low levels of

specificity (item 9 and 11, with λ = .339 and .276 on the Inattention S-factor versus λ = .513 and .496 on the ADHD G-factor, respectively). In other words, items 9 and 11 should not be used in defining the Inattentive subtype of ADHD as these items apparently present no meaningful specificity once the ADHD G-factor is taken into account. Similarly, the Hyperactivity S-factor also appears to be relatively well defined by at least 3 out of 5 items with loadings between .452 and .572. Conversely, the remaining 2 items, items 12 and 14, apparently present very low levels of specificity (λ = .165 and .232 on the Hyperactivity S-factor) and mostly serve to define the ADHD G-factor (λ = .608 and .457). Finally, contrary to what was observed for the other two dimensions, the Impulsivity S-Factor appears to be generally poorly defined, with only item 16 (λ = .556) presenting a substantial level of association with this S-factor. These results thus apparently support the presence of ADHD subtypes Inattentive and Hyperactive subtypes defined based on, respectively, items 1-2-3-4-7-8-10 and 5-6-13. Similarly, these results suggest that Impulsive items should be treated as distinct from the Hyperactivity items but mostly used to define the common core of ADHD symptoms.

Looking at the scale score reliability, Cronbach's α coefficients appear to be quite high for all factors (.654 to .850) due to the specific, and inadequate in this case, manner in which α computes composite reliability (Sijtsma, 2009) (Table 2). McDonald proposed an alternative model-based omega (ω) coefficient providing a more realistic estimate of scale-score reliability, especially when based on complex measurement model (McDonald, 1970). Here, coefficients ω revealed a very high level of reliability of the global ADHD ratings (ω=.877) when these are modelled while also taking into account the presence of S-factors. In accordance with the standardized model results, the scale score reliability estimate of the Inattention S-factor remains fully satisfactory (ω=.771). However, the scale score reliability estimate of the Hyperactivity (ω=.594) and the Impulsivity (ω=.510) S-factors are much lower.

Measurement Invariance

Starting from the bifactor model M5B, systematic tests of measurement invariance were conducted according to gender, age, and gender by age groupings (Table 3). Interestingly, throughout the full sequence of invariance tests, all of the increasingly restrictive models estimated across all possible groupings of students provided a satisfactory level of fit to the data, with CFI and TLI>.90 and RMSEA<.08. The tests of metric/weak, scalar/strong, strict, and latent variance invariance across gender, age, and age-by-gender categories are fully supported. In many cases, the fit indices incorporating a control for model parsimony (i.e. TLI and RMSEA) improved when invariance constraints are added to the model; the more restricted model with strict invariance and invariance of the latent variances even show a substantially higher degree of fit to the data than the baseline model. Furthermore, when equality constraints are placed on the latent means across genders, but not age categories, the MD$\Delta\chi^2$ is significant. Although this conclusion was not supported by changes in fit indices, which remain under the recommended cut-off values, it did suggest potential gender-related latent mean differences. Given the substantive interest of gender-related differences in ADHD levels, we systematically probed these differences. When females' latent means are fixed to 0 for identification purposes, males' latent means (expressed as differences in SD units from females' means) are significantly higher on the ADHD G factor (M=.529; s.e.=.095; p<.01) and significantly lower on the Inattention (M=-.455; s.e.=.093; p<.01), Hyperactivity (M=-.702; s.e.=.130; p<.01), and Impulsivity (M=-.962; s.e.=.146; p<.01) S-factors. This last result should be put into perspective of the nature of the bifactor model as showing that, once overall levels of ADHD are extracted from the ratings, females present higher levels on the residual ratings related to the S-factors. This suggests that, for girls, specific ratings tend to have a greater tendency to be interpreted as something different from a generic ADHD syndrome. A similar pattern of results was also observable when latent means differences were probed across the gender-by-age groups, showing similar latent mean differences across genders, and confirming the absence of differences across age groups.

Discussion

The ASRS, specifically designed to assess adult's symptoms of ADHD, is widely used in many languages around the world and has been extensively validated previously, although studies are quite rare outside the USA. This is worrying as Wu, Li and Zumbo (2007, p.1) emphasized that "the validity of cross-country (or cross-cultural score comparisons is vital to many practices in applied psychology and educational research. […] Unless evidence is demonstrated, construct comparability should never be naively assumed" (Wu, Li, & Zumbo, 2007). Similarly, the real underlying structure of the ADHD

diagnosis, and the evolution of its symptomatic expression over the lifespan, has recently been called into question, making cross-cultural comparisons even riskier. Although we did not specifically verify the equivalence of multiple linguistic versions of the ASRS, something that should be investigated more thoroughly in future studies, this study is the first to thoroughly assess the structure of the ASRS in a large general population sample of French adults. We used CFA and state-of-the-art methodology to compare the fit to the data of alternative representations of ADHD symptoms. Our results provide a clear support to the superiority of a three-factor bifactor model of ADHD symptoms apparently corresponding more to the ICD-10 definition of Hyperkinetic Disorder than to the DSM-IV definition of ADHD, thus replicating the results from a recent American study of a clinical population of adults with ADHD rated by a clinician (Gibbins, et al., 2012). Both classifications use the same 18 behavioural symptoms but differ in what may seem a minor point. Item 15 ("Speaks to much") is allocated to the Hyperactivity factor in the DSM-IV and to the Impulsivity factor in the ICD-10. Criticisms have been raised toward the somewhat weak definition of Impulsivity in the DSM-IV measured by only three items. We do not think that subtypes really need to be relabelled based on the change in location of a single item to avoid creating unnecessary confusion in psychiatric communications. In this case, it simply appears more appropriate to refer to the ICD-10 conceptualization and to note its superiority over the DSM-IV classification system.

More specifically, our results suggest the presence of one general ADHD G-factor and three specific S-factors representing Inattention, Hyperactivity and Impulsivity subtypes. Overall this means that the total score of the ADHD is meaningful and reliable as shown by satisfactory scale score reliability coefficients (α=.850, and ω=.877). This total score thus appears to provide an appropriate measure of global ADHD symptoms in adults from the general population, supporting the use of the ASRS as a potential screening tool for the evaluation of ADHD. But what the bifactor models also show is that the three phenomenological components of ADHD (Inattention, Hyperactivity and Impulsivity) need to be distinguished. Each of these factors still explain enough residual variance after taking into account the variance of the ADHD G-factor to justify them as potential subtypes of ADHD. A bifactor model supports the multiple-pathways conception of ADHD (Nigg, Goldsmith, & Sachek, 2004; Sonuga-Barke, 2002, 2005) and suggests that there are distinct etiological influences that converge on the same core syndrome (Chen, West, & Sousa, 2006) with some remaining specificities. However, future studies should investigate the validity, sensibility and specificity of ADHD assessments based on the self-rated ASRS and formal clinical diagnoses of ADHD made by experiences clinicians using structured interview schedules. Compared to studies conducted among younger participants, which generally support of two-factor bifactor model of ADHD where symptoms of Hyperactivity and Impulsivity are grouped together (Caci, et al., 2013; Martel, et al., 2010; Toplak, et al., 2009), our results suggests that ADHD is a clinical condition that tends to evolve toward greater differentiation in adulthood. This calls into question the direct transferability of children and adolescents diagnoses criteria to adults.

Furthermore, our results suggest that the items mainly associated with the various subtypes may not all present a sufficient level of specificity, at least in adults, to be used to define formal diagnostic subtypes of ADHD, further calling into question the need to refine ADHD diagnosis criteria for adults, or maybe to develop items that would be more suitable for adults (Barkley, et al., 2008). For instance, the Impulsivity S-Factor includes items presenting a very low level of specificity (items 15, 17 and 18) and remains defined mostly by a single item (item 16), calling into question the appropriateness of the Impulsivity S-subscale in adults. Conversely, the Hyperactivity S-factor presents a satisfactory level of both specificity, and generality, with only items 12 and 14 showing a clearly dominant association to the ADHD G-Factor and a low level of specificity. The remaining Hyperactivity items (items 5, 6 and 13) present stronger associations to both the ADHD G-Factor and the Hyperactivity S-Factor. Finally, an altogether opposite pattern is observed for four of the items associated with the Inattentive subtype (items 1 to 4), which rather present a very high level of specificity, so high in fact as to call into question their appropriateness in defining global ADHD symptoms. This observation is worrying as these four items are all part of the 6-item ASRS screener, which is recommended for use as a first screening step in the detection of potential ADHD diagnoses in adults (Kessler, et al., 2007). Clearly, at least some of the impulsivity items included in the ASRS (items 15, 17 and 18), as well as perhaps items 12 and 14 from the Hyperactivity subscale, seems far more suited to the detection of ADHD. These items present lower levels of subtypes-specificity and represent much cleaner indicators of the

global G-construct of ADHD. Clearly, replication of the present results would strongly call into question the current composition of the ADHD screener and suggest the need for further investigation of the best scoring procedures for the use of the ASRS to detect potential ADHD cases in the context of general population surveys or extensive screening programs.

More precisely, our results show that Impulsivity symptoms merge to some extent with Hyperactivity symptoms to define a global, general, condition of ADHD, whereas Inattentive symptoms, as well as some Hyperactivity symptoms, may appear on their own, potentially linked to different causal pathways. For clinicians, this means, that patients can be placed on a continuum with regard of their total ASRS score and that such evaluations should be complemented by specific evaluations of inattention and hyperactivity. In patients with marked levels of inattention, impulsivity or hyperactivity could potentially become a comorbid condition, as suggested in recent deliberations related to the development of a novel "Inattentive (restrictive)" subtype for DSM-V. However, fully validating this proposal would require moving to person-centred profile analyses (Martel, Roberts, Gremillion, von Eye, & Nigg, 2011). Additional studies are needed to examine the changes over time in these ratings, as well as their state and trait components (Normand, Flora, Toplak, & Tannock, 2012). An important extension of the present study would involve using, and contrasting, multiple measures of ADHD developed specifically for adults, such as the ASRS and the Conners Adult ADHD Rating Scale (Christiansen et al., 2011), rather than simply contrast ICD-10 and DSM-IV conceptions as applied to a single instrument as these diagnostic systems have yet to propose diagnostic criteria specific to adults (Barkley, et al., 2008).

A final objective of this study was to extend previous investigations of the ASRS by verifying the measurement invariance of the final bifactor model across genders, age-groups, and gender-by-age groups. We thus verified whether group membership introduced any measurement bias in adults self-reports of ADHD symptoms. Interestingly, our results provide strong support to the total invariance of the factor loadings, thresholds, uniquenesses, and variances across all possible subgroups, only alluding to possible (and expected) mean-level differences across subgroups. We found that levels on the ADHD G-factor were significantly higher for men than women, but that women presented lower levels than men on all three S-factors. This last result concerning the specific factors was unexpected, albeit in line with results from a recent study (Kessler, et al., 2007).This last result should be put into perspective of the nature of the bifactor model as showing that, once overall levels of ADHD are extracted from the ratings, females present higher levels on the residual ratings related to the S-factors. This suggests that, for girls, specific ratings tend to have a greater tendency to be interpreted as something different from a generic ADHD syndrome. At last, latent means comparisons show that levels on the general ADHD factor are higher for males. This is directly in line with a recent meta-analysis reporting the male-female ratio to be about 1.6 (Willcutt, 2012).

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles*. Burlington, VT: Research Centre for Children, Youth and Families.

Adler, L. A., Guida, F., Irons, S., Rotrosen, J., & O'Donnell, K. (2009). Screening and imputed prevalence of ADHD in adult patients with comorbid substance use disorder at a residential treatment facility. *Postgraduate Medicine*, *121*, 7-10.

Adler, L. A., & Newcorn, J. H. (2011). Administering and evaluating the results of the adult ADHD Self-Report Scale (ASRS) in adolescents. *Journal of Clinical Psychiatry*, *72*, e20.

Adler, L.A., Shaw, D.M., Spencer, T., Newcorn, J., Hammerness, P., Sitt, D., et al. (2012). Preliminary examination of the reliability and concurrent validity of the Attention-Deficit/Hyperactivity disorder Self-Report Scale v1.1 to rate symptoms of Attention-Deficit/Hyperactivity Disorder in adolescents. *Journal of Child and Adolescent Psychopharmacology*, 22, 238-244.

American Psychiatric Association. (1994). *Diagnosis and Statistical Manual of Mental Disorders, IV*. Washington, DC: American Psychiatric Press.

Asparouhov, T., & Muthén, B. O. (2006). *Robust chi-square difference testing with mean and variance adjusted test statistics*. Los Angeles, CA: Muthén and Muthén; http://www.statmodel.com/examples/webnote.shtml#web10.

Barkley, R. A., Murphy, K. R., & Fischer, M. (2008). *ADHD in adults: What the science says*. New York: Guilford.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*, 186-203.

Biederman, J., Mick, E., & Faraone, S. V. (2000). Age-dependent decline of symptoms of Attention-Deficit/Hyperactivity Disorder: Impact of remission definition and symptom type. *The American Journal of Psychiatry, 157*, 816-818.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Caci, H., Bayle, F. J., & Bouchez, J. (2008). Adult ADHD: Translation and factor analysis of the ASRS v1.1. *European Psychiatry, 23*, 367-368.

Caci, H., Morin, A. J. S., & Tran, A. (2013, Ahead of print). Teacher ratings of the ADHD-RS IV in a community sample: Results from the ChiP-ARD study. *Journal of Attention Disorders*. DOI: 10.1177/1087054712473834

Capdevila-Brophy, C., Artigas-Pallarés, J., Navarro-Pastor, J. B., Garcia-Nonell, K., Rigau-Ratera, E., & Obiols, J.E. (2013, Ahead of print). ADHD Predominantly Inattentive Subtype with high sluggish cognitive tempo: a new clinical entity? *Journal of Attention Disorders*. DOI: 10.1177/1087054712445483

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464-504.

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*, 189-225.

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

Christiansen, H., Kis, B., Hirsch, O., Philipsen, A., Henneck, M., Panczuk, A., et al. (2011). German validation of the Conners Adult ADHD Rating Scales-self report (CAARS-S) I: Factor structure and normative data. *European Psychiatry, 26*, 100-107.

de Graaf, R., Kessler, R. C., Fayyad, J., ten Have, M., Alonso, J., Angermeyer, M., et al. (2008). The prevalence and effects of adult Attention-Deficit/Hyperactivity Disorder (ADHD) on the performance of workers: Results from the WHO World Mental Health Survey Initiative. *Occupational & Environmental Medicine, 65*, 835-842.

Dumenici, L., McConaughy, S. H., & Achenbach, T. M. (2004). A hierarchical three-factor model of inattention-hyperactivity-impulsivity derived from the attention problems syndrome of the Teacher's Report Form. *School Psychology Review, 33*, 287-301.

Erikson, Erik H. (1968). *Identity, Youth and Crisis*. New York, NY: Norton.

Faraone, S.V., Biederman, J., & Mick, E. (2006). The age-dependent decline of attention-deficit hyperactivity disorder: Meta-analysis of follow-up studies. *Psychological Medicine, 36*, 159-165.

Fayyad, J., De Graaf, R., Kessler, J., Angermeyer, A. M., Demyttenaere, K., De Girolamo, G., et al. (2007). Cross-national prevalence and correlates of adult attention-deficit hyperactivity disorder. *British Journal of Psychiatry, 190*, 402-409.

Finney, S. J., & DiStefano, C. G., CO: IAP. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (pp. 269-314). Greenwich, CT: IAP.

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.

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling, 16*, 625-641.

Foundation, D. (2012). *Diagnostic interview for ADHD in adults*. Retrieved July 21, 2012, from http://www.divacenter.eu/DIVA.aspx

Francophonie (2012). *Organisation Internationale de la Francophonie* [International Organization of French-Speaking Countries]. Retrieved June 14, 2012 from www.francophonie.org.

Gibbins, C., Toplak, M. E., Flora, D. B., Weiss, M. D., & Tannock, R. (2012). Evidence for a general factor model of ADHD in adults. *Journal of Attention Disorders, 16*, 635-644.

Haase, C.M., Seider, B.H., Shiota, M.N., & Levenson, R.W. (2012). Anger and sadness in response to an emotionally neutral film: Evidence for age-specific associations with well-being. *Psychology & Aging, 27*, 305-317.

Hodgkins, P., Arnold, L. E., Shaw, M., Caci, H., Kahle, J., Woods, A. G., et al. (2012). A systematic review of global publication trends regarding of long-term outcomes of ADHD. *Frontiers in Psychiatry*, *2*, 1-17.

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, *2*, 41-54.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1-55.

Katusic, S. K., Barbaresi, W. J., Colligan, R. C., Weaver, A. L., Leibson, C. L., & Jacobsen, S. J. (2005). Case definition in epidemiologic studies of AD/HD. *Annals of Epidemiology*, *15*, 430-437.

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., et al. (2005). The World Health Organization adult ADHD self-report scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine*, *35*, 245-256.

Kessler, R. C., Adler, L., Barkley, R., Biederman, J., Conners, C. K., Demler, O., et al. (2006). The prevalence and correlates of adult ADHD in the United States: Results from the national comorbidity survey replication. *The American Journal Psychiatry*, 163, 716-723.

Kessler, R. C., Adler, L., Gruber, M. J., Sarawate, C. A., Spencer, T., & Van Brunt, D. L. (2007). Validity of the World Health Organization Adult ADHD Self-Report Scale (ASRS) screener in a representative sample of health plan members. *International Journal of Methods in Psychiatry Research*, *16*, 52-65.

Kessler, R. C., Berglund, P., Chiu, W. T., Demler, O., Herringa, S., Hiripi, E., et al. (2004). The US National Comorbidity Survey Replication (NCS-R): Design and field procedures. *International Journal of Methods in Psychiatry Research*, *13*, 69-92.

Kooij, S. J. J., Bejerot, S., Blackwell, A., Caci, H., Casas-Brugué, M., Carpentier, P. J., et al. (2010). European consensus statement on diagnosis and treatment of adult ADHD: The European Network Adult ADHD. *BMC Psychiatry*, *10*, 67-91.

Kray, J., & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology & Aging*, *15*, 126-147.

Lang, F.R., John, D., Lüdtke, O., Schupp, J., & Wagner, G. (2011). Short assessment of the Big-Five: Robust across survey methods except telephone interviewing. *Behavioral Research*, *43*, 548-567.

Larsson, H., Lichenstein, P., & Larsson, J. O. (2006). Genetic contributions to the development of ADHD subtypes from childhood to adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry*, *45*, 973-981.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19-40.

Martel, M. M., Roberts, B., Gremillion, M., von Eye, A., & Nigg, J. T. (2011). External validation of the bifactor model of ADHD: Explaining heterogeneity in psychiatric comorbidity, cognitive control, and personality trait profiles within DSM-IV ADHD. *Journal of Abnormal Child Psychology*, *39*, 1111-1123.

Martel, M. M., von Eye, A., & Nigg, J. T. (2010). Revisiting the latent structure of ADHD: Is there a 'g' factor? *Journal of Child Psychology and Psychiatry*, *51*, 905-914.

McDonald, R.P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical & Statistical Psychology*, *23*, 1-21.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525-543.

Millsap, R. E., & Tein, J. Y. (2004). Assessing Factorial Invariance in Ordered-Categorical Measures. *Multivariate Behavioral Research*, *39*, 479-515.

Morin, A. J. S., Madore, I., Morizot, J., Boudrias, J.-S., & Tremblay, M. (2009). The Workplace Affective Commitment Multidimensional Questionnaire: Factor structure and Measurement Invariance. *International Journal of Psychology Research*, *4*, 307-344.

Morin, A. J. S., Moullec, G., Maïano, C., Layet, L., Just, J.-L., & Ninot, G. (2011). Psychometric properties of the Center for Epidemiologic Studies Depression Scale (CES-D) in French Clinical and Non-Clinical Adults. *Epidemiology and Public Health*, *59*, 327-340.

Muthén, B. O. (2004). Mplus Technical Appendices. Los Angeles, CA: Muthén and Muthén; http://www.statmodel.com/techappen.shtml

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous

outcomes. http://gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf

Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide. Seventh Edition*. Los Angeles, CA: Muthén & Muthén.

Nigg, J. T., Goldsmith, H. H., & Sachek, J. (2004). Temperament and attention deficit hyperactivity disorder: The development of a multiple pathway model. *Journal of child and Adolescent Psychology*, *33*, 42-53.

Normand, S., Flora, D. B., Toplak, M. E., & Tannock, R. (2012). Evidence for a general ADHD factor from a longitudinal general school population study. *Journal of Abnormal Child Psychology*, *40*, 555-567.

Polanczyk, G., Silva de Lima, M., Lessa Horta, B., Biederman, J., & Rohde, L. A. (2007). The worldwide prevalence of ADHD: A systematic review and metaregression analysis. *The American Journal of Psychiatry*, *164*, 942-948.

Rensvold, R. B., & Cheung, G. W. (1998). Testing measurement model for factorial invariance: A systematic approach. *Educational and Psychological Measurement*, *58*, 1017-1034.

Shaw, M., Hodgkins, P., Caci, H., Young, S., Kahle, J., Woods, A. G., et al. (2012). A systematic review and analysis of long-term outcomes in ADHD: Effects of treatment and non-treatment. *BMC Medicine*, *10*, 1-15.

Sijtsma, K. (2009). On the use, misuse, and the very limited usefulness of Cronbach's alpha [Introduction to a special issue]. *Psychometrika*, *74*, 107-120.

Solanto, M. V. (2000). The predominantly inattentive subtype of attention-deficit/hyperactivity disorder. *CNS Spectrums*, *5*, 45-51.

Sonuga-Barke, E. J. S. (2002). Psychological heterogeneity in AD/HD — a dual pathway model of behaviour and cognition. *Behavioural Brain Research*, *130*, 29-36.

Sonuga-Barke, E.J. (2005). Causal models of Attention-Deficity/Hyperactivity Disorder: From simple deficits to multiple developmental pathways. *Biological Psychiatry*, *57*, 1231-1238.

Toplak, M. E., Pitch, A., Flora, D. B., Iwenofu, L., Ghelani, K., Jain, U., et al. (2009). The unity and diversity of inattention and hyperactivity/impulsivity in ADHD: Evidence for a general factor with separable dimensions. *Journal of Abnormal Child Psychology*, *37*, 1137-1150.

Willcutt, E. G. (2012). The prevalence of DSM-IV Attention-Deficit/Hyperactivity Disorder: A meta-analytic review. *Neurotherapeutics*, *9*, 490-499.

Word Health Organization. (1993). *The ICD-10 Classification of Mental and Behavioral Disorders: Diagnostic Criteria for Research*. Geneva, Switzerland: World Health Organization.

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation*, *12*, 1-26.

Yu, C. Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Los Angeles, CA: University of California.

Table 1. Goodness-of-fit statistics for the alternative measurement models.

Model	χ^2-test (df)	CFI	TLI	RMSEA	RMSEA CI 90%
M1: One factor	2545.061 (135)*	.783	.755	.123	[.119-.128]
M2: Two correlated factors	1182.687 (134)*	.906	.892	.082	[.078-.086]
M3A: Three correlated factors DSM-IV	1060.389 (132)*	.917	.903	.077	[.073-.082]
M3B: Three correlated factors ICD-10	906.085 (132)*	.930	.919	.071	[.066-.075]
M4: Two-factor Bifactor	951.029 (117)*	.925	.902	.078	[.073-.083]
M5A: Three-factor Bifactor DSM-IV	756.130 (117)*	.943	.925	.068	[.064-.073]
M5B: Three-factor Bifactor ICD-10	655.508 (117)*	.952	.937	.063	[.058-.067]
Model 5B estimated in the subsamples					
M5B-W: Women (*n* =661)	439.148 (117)*	.946	.929	.065	[.058-.071]
M5B-M: Men (*n=510*)	331.899 (117)*	.960	.948	.060	[.052-.068]
M5B-Y: Younger than 40 (*n=438*)	319.589 (117)*	.953	.938	.063	[.055-.071]
M5B-O: 40+ years old (*n=733*)	469.025 (117)*	.949	.933	.064	[.058-.070]
M5B-YW: Younger women (*n=289)*	238.751 (117)*	.957	.944	.060	[.049-.071]
M5B-OW: Middle-age women (*n=372*)	321.700 (117)*	.935	.915	.069	[.060-.078]
M5B-YM: Younger men (*n=361*)	202.096 (117)*	.944	.927	.070	[.053-.086]
M5B-OM: Middle-age men (*n=149*)	285.352 (117)*	.957	.943	.063	[.054-.072]

Note. *p<0.01; χ^2: chi-square test of model fit and its associated degrees of freedom (*df*); CFI: Comparative Fit Index; TLI: Tucker-Lewis Index; RMSEA: Root Mean Square Error of Approximation and its 90% Confidence Interval (CI); The fact that WLSMV χ^2 values are not exact, but "*estimated*" as the closest integer necessary to obtain a correct *p*-value explains the fact that sometimes the χ^2 and resulting CFI values can be non-monotonic with model complexity.

Table 2. Standardized solution for the final retained model (M5B).

	General ADHD	Inattention	Hyperactivity	Impulsivity	R^2
1. Wrapping up final details of a project	.347	.566			.441
2. Getting things in order	.296	.660			.523
3. Remembering appointments	.302	.494			.336
4. Avoid or delay getting started	.308	.586			.439
5. Squirm with your hands or feet	.556		.452		.513
6. Overly active and compelled to do things	.502		.511		.513
7. Make careless mistakes	.550	.405			.467
8. Difficulty keeping attention	.607	.431			.553
9. Difficulty concentrating	.513	.339			.378
10. Misplacing things	.407	.429			.349
11. Distracted by activity or noise	.496	.276			.311
12. Leaving seat in meetings	.608		.165		.397
13. Feeling restless or fidgety	.658		.572		.761
14. Difficulty unwinding and relaxing	.457		.232		.263
15. Talking too much	.423			.394	.334
16. Finishing sentences of other people	.447			.556	.509
17. Waiting your turn	.469			.276	.296
18. Interrupting busy others	.566			.359	.449
Model-based scale score reliability (ω)	.877	.771	.594	.510	—
Classical scale score reliability (α)	.850	.820	.744	.654	—

Note. All loadings and R2 are significant at $p < 0.01$

Table 3. Tests of measurement invariance for the final model (M5B).

	$\chi^2(df)$	CFI	TLI	RMSEA	RMSEA 90% CI	MD$\Delta\chi^2(\Delta df)$	ΔCFI	ΔTLI	ΔRMSEA
Tests of measurement invariance across genders									
Configural invariance	772.571 (234)*	.952	.938	.063	[.058-.068]	--	--	--	--
Metric/weak invariance	803.212 (266)*	.952	.945	.059	[.054-.063]	84.284 (32)*	.000	.007	-.004
Scalar/strong invariance	871.539 (316)*	.951	.952	.055	[.050-.059]	134.796 (50)*	-.001	.007	-.004
Strict invariance	863.859 (334)*	.953	.957	.052	[.048-.056]	37.608 (18)*	.002	.005	-.003
Latent variance invariance	729.018 (338)*	.965	.969	.044	[.040-.049]	7.091 (4)	.012	.012	-.008
Latent means invariance	756.839 (342)*	.963	.967	.046	[.041-.050]	19.413 (4)*	-.002	-.002	.002
Tests of measurement invariance across age groups									
Configural invariance	763.204 (234)*	.952	.937	.062	[.058-.067]	--	--	--	--
Metric/weak invariance	734.174 (266)*	.957	.951	.055	[.050-.060]	46.333 (32)	.005	.014	-.007
Scalar/strong invariance	705.276 (316)*	.964	.966	.046	[.042-.051]	51.377 (50)	.007	.015	-.009
Strict invariance	713.990 (334)*	.965	.968	.044	[.040-.049]	35.234 (18)*	.001	.002	-.002
Latent variance invariance	624.547 (338)*	.974	.976	.038	[.034-.043]	5.349 (4)	.009	.008	-.006
Latent means invariance	636.082 (342)*	.973	.976	.039	[.034-.043]	12.138 (4)	-.001	.000	.001
*Tests of measurement invariance across age*gender groups*									
Configural invariance	1002.380 (468)*	.951	.937	.063	[.057-.068]	--	--	--	--
Metric/weak invariance	1118.679 (564)*	.950	.945	.058	[.053-.063]	186.858 (96)*	.001	.008	-.005
Scalar/strong invariance	1263.044 (714)*	.950	.957	.052	[.047-.056]	242.504 (150)*	.000	.012	-.006
Strict invariance	1319.190 (768)*	.950	.960	.050	[.045-.054]	102.247 (54)*	.000	.003	-.002
Latent variance invariance	1220.212 (780)*	.960	.969	.044	[.039-.049]	19.610 (12)	.010	.009	-.006
Latent means invariance	1259.435 (792)*	.958	.967	.045	[.040-.050]	33.151 (12)*	-.002	-.002	.001

Notes. *p<0.01; χ^2: chi-square test of model fit and its associated degrees of freedom (*df*); CFI: Comparative Fit Index; TLI: Tucker-Lewis Index; RMSEA: Root Mean Square Error of Approximation and its 90% Confidence Interval (CI); Δ change relative to the previous model in the sequence; MD$\Delta\chi^2$: chi-square difference test calculated with the Mplus DIFFTEST function for the robust weighted least square estimator (WLSMV). The fact that WLSMV χ^2 values are not exact, but "*estimated*" as the closest integer necessary to obtain a correct *p*-value explains the fact that the χ^2 and resulting CFI values can be non-monotonic with model complexity.