**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. MORIN¹*, Antoine TRAN², Hervé CACI²

¹Corresponding author

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

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

This is the prepublication version of the following manuscript:


© 2016. This paper is not the copy of record and may not exactly replicate the authoritative document published in the *Journal of Attention Disorders*.

**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, Lichtenstein, & 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...
 Factor Validity of the Adult Symptom Rating Scale

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: r(1023.4)=.685 (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 ($\chi^2$), 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 $\chi^2$ 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 ($\Delta$CFI and $\Delta$TLI >.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 ($\Delta$CFI and $\Delta$TLI >.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 ($r=.296$ 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 $\lambda = .339$ and .276 on the Inattentive S-factor versus $\lambda = .513$ 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 ($\lambda = .165$ and .232 on the Hyperactivity S-factor) and mostly serve to define the ADHD G-factor ($\lambda = .608$ 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 ($\lambda = .556$) 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 $\alpha$ coefficients appear to be quite high for all factors (.654 to .850) due to the specific, and inadequate in this case, manner in which $\alpha$ computes composite reliability (Sijtsma, 2009) (Table 2). McDonald proposed an alternative model-based omega ($\omega$) coefficient providing a more realistic estimate of scale-score reliability, especially when based on complex measurement model (McDonald, 1970). Here, coefficients $\omega$ revealed a very high level of reliability of the global ADHD ratings ($\omega=.877$) 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 Inattentive S-factor remains fully satisfactory ($\omega=.771$). However, the scale score reliability estimate of the Hyperactivity ($\omega=.594$) and the Impulsivity ($\omega=.510$) 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$\lambda^2$ 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 suggest 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
Factor Validity of the Adult Symptom Rating Scale

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


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

<table>
<thead>
<tr>
<th>Model</th>
<th>χ²-test (df)</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
<th>RMSEA CI 90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1: One factor</td>
<td>2545.061 (135)*</td>
<td>.783</td>
<td>.755</td>
<td>.123</td>
<td>[.119-.128]</td>
</tr>
<tr>
<td>M2: Two correlated factors</td>
<td>1182.687 (134)*</td>
<td>.906</td>
<td>.892</td>
<td>.082</td>
<td>[.078-.086]</td>
</tr>
<tr>
<td>M3A: Three correlated factors DSM-IV</td>
<td>1060.389 (132)*</td>
<td>.917</td>
<td>.903</td>
<td>.077</td>
<td>[.073-.082]</td>
</tr>
<tr>
<td>M3B: Three correlated factors ICD-10</td>
<td>906.085 (132)*</td>
<td>.930</td>
<td>.919</td>
<td>.071</td>
<td>[.066-.075]</td>
</tr>
<tr>
<td>M4: Two-factor Bifactor</td>
<td>951.029 (117)*</td>
<td>.925</td>
<td>.902</td>
<td>.078</td>
<td>[.073-.083]</td>
</tr>
<tr>
<td>M5A: Three-factor Bifactor DSM-IV</td>
<td>756.130 (117)*</td>
<td>.943</td>
<td>.925</td>
<td>.068</td>
<td>[.064-.073]</td>
</tr>
<tr>
<td>M5B: Three-factor Bifactor ICD-10</td>
<td>655.508 (117)*</td>
<td>.952</td>
<td>.937</td>
<td>.063</td>
<td>[.058-.067]</td>
</tr>
</tbody>
</table>

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; χ²: 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 χ² values are not exact, but "estimated" as the closest integer necessary to obtain a correct p-value explains the fact that sometimes the χ² and resulting CFI values can be non-monotonic with model complexity.
Table 2. Standardized solution for the final retained model (M5B).

<table>
<thead>
<tr>
<th></th>
<th>General ADHD</th>
<th>Inattention</th>
<th>Hyperactivity</th>
<th>Impulsivity</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Wrapping up final details of a project</td>
<td>.347</td>
<td>.566</td>
<td></td>
<td></td>
<td>.441</td>
</tr>
<tr>
<td>2. Getting things in order</td>
<td>.296</td>
<td>.660</td>
<td></td>
<td></td>
<td>.523</td>
</tr>
<tr>
<td>3. Remembering appointments</td>
<td>.302</td>
<td>.494</td>
<td></td>
<td></td>
<td>.336</td>
</tr>
<tr>
<td>4. Avoid or delay getting started</td>
<td>.308</td>
<td>.586</td>
<td></td>
<td></td>
<td>.439</td>
</tr>
<tr>
<td>5. Squirm with your hands or feet</td>
<td>.556</td>
<td></td>
<td>.452</td>
<td></td>
<td>.513</td>
</tr>
<tr>
<td>6. Overly active and compelled to do things</td>
<td>.502</td>
<td></td>
<td>.511</td>
<td></td>
<td>.513</td>
</tr>
<tr>
<td>7. Make careless mistakes</td>
<td>.550</td>
<td>.405</td>
<td></td>
<td></td>
<td>.467</td>
</tr>
<tr>
<td>8. Difficulty keeping attention</td>
<td>.607</td>
<td>.431</td>
<td></td>
<td></td>
<td>.553</td>
</tr>
<tr>
<td>10. Misplacing things</td>
<td>.407</td>
<td>.429</td>
<td></td>
<td></td>
<td>.349</td>
</tr>
<tr>
<td>11. Distracted by activity or noise</td>
<td>.496</td>
<td>.276</td>
<td></td>
<td></td>
<td>.311</td>
</tr>
<tr>
<td>12. Leaving seat in meetings</td>
<td>.608</td>
<td></td>
<td>.165</td>
<td></td>
<td>.397</td>
</tr>
<tr>
<td>13. Feeling restless or fidgety</td>
<td>.658</td>
<td>.572</td>
<td></td>
<td></td>
<td>.761</td>
</tr>
<tr>
<td>14. Difficulty unwinding and relaxing</td>
<td>.457</td>
<td>.232</td>
<td></td>
<td></td>
<td>.263</td>
</tr>
<tr>
<td>15. Talking too much</td>
<td>.423</td>
<td></td>
<td></td>
<td>.394</td>
<td>.334</td>
</tr>
<tr>
<td>16. Finishing sentences of other people</td>
<td>.447</td>
<td></td>
<td>.556</td>
<td>.509</td>
<td></td>
</tr>
<tr>
<td>17. Waiting your turn</td>
<td>.469</td>
<td></td>
<td>.276</td>
<td>.296</td>
<td></td>
</tr>
<tr>
<td>18. Interrupting busy others</td>
<td>.566</td>
<td></td>
<td></td>
<td>.359</td>
<td>.449</td>
</tr>
</tbody>
</table>

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).

<table>
<thead>
<tr>
<th>Tests of measurement invariance across genders</th>
<th>χ² (df)</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
<th>RMSEA 90% CI</th>
<th>MDΔχ²(Δdf)</th>
<th>ΔCFI</th>
<th>ΔTLI</th>
<th>ΔRMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configural invariance</td>
<td>772.571 (234)*</td>
<td>.952</td>
<td>.938</td>
<td>.063</td>
<td>[.058-.068]</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Metric/weak invariance</td>
<td>803.212 (266)*</td>
<td>.952</td>
<td>.945</td>
<td>.059</td>
<td>[.054-.063]</td>
<td>84.284 (32)*</td>
<td>.000</td>
<td>.007</td>
<td>-.004</td>
</tr>
<tr>
<td>Scalar/strong invariance</td>
<td>871.539 (316)*</td>
<td>.951</td>
<td>.952</td>
<td>.055</td>
<td>[.050-.059]</td>
<td>134.796 (50)*</td>
<td>-.001</td>
<td>.007</td>
<td>-.004</td>
</tr>
<tr>
<td>Strict invariance</td>
<td>863.859 (334)*</td>
<td>.953</td>
<td>.957</td>
<td>.052</td>
<td>[.048-.056]</td>
<td>37.608 (18)*</td>
<td>.002</td>
<td>.005</td>
<td>-.003</td>
</tr>
<tr>
<td>Latent variance invariance</td>
<td>729.018 (338)*</td>
<td>.965</td>
<td>.969</td>
<td>.044</td>
<td>[.040-.049]</td>
<td>7.091 (4)</td>
<td>.012</td>
<td>.012</td>
<td>-.008</td>
</tr>
<tr>
<td>Latent means invariance</td>
<td>756.839 (342)*</td>
<td>.963</td>
<td>.967</td>
<td>.046</td>
<td>[.041-.050]</td>
<td>19.413 (4)*</td>
<td>-.002</td>
<td>-.002</td>
<td>.002</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tests of measurement invariance across age groups</th>
<th>χ² (df)</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
<th>RMSEA 90% CI</th>
<th>MDΔχ²(Δdf)</th>
<th>ΔCFI</th>
<th>ΔTLI</th>
<th>ΔRMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configural invariance</td>
<td>763.204 (234)*</td>
<td>.952</td>
<td>.937</td>
<td>.062</td>
<td>[.058-.067]</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Metric/weak invariance</td>
<td>734.174 (266)*</td>
<td>.957</td>
<td>.951</td>
<td>.055</td>
<td>[.050-.060]</td>
<td>46.333 (32)</td>
<td>.005</td>
<td>.014</td>
<td>-.007</td>
</tr>
<tr>
<td>Scalar/strong invariance</td>
<td>705.276 (316)*</td>
<td>.964</td>
<td>.966</td>
<td>.046</td>
<td>[.042-.051]</td>
<td>51.377 (50)</td>
<td>.007</td>
<td>.015</td>
<td>-.009</td>
</tr>
<tr>
<td>Strict invariance</td>
<td>713.990 (334)*</td>
<td>.965</td>
<td>.968</td>
<td>.044</td>
<td>[.040-.049]</td>
<td>35.234 (18)*</td>
<td>.001</td>
<td>.002</td>
<td>-.002</td>
</tr>
<tr>
<td>Latent variance invariance</td>
<td>624.547 (338)*</td>
<td>.974</td>
<td>.976</td>
<td>.038</td>
<td>[.034-.043]</td>
<td>5.349 (4)</td>
<td>.009</td>
<td>.008</td>
<td>-.006</td>
</tr>
<tr>
<td>Latent means invariance</td>
<td>636.082 (342)*</td>
<td>.973</td>
<td>.976</td>
<td>.039</td>
<td>[.034-.043]</td>
<td>12.138 (4)</td>
<td>-.001</td>
<td>-.001</td>
<td>.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tests of measurement invariance across age*gender groups</th>
<th>χ² (df)</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
<th>RMSEA 90% CI</th>
<th>MDΔχ²(Δdf)</th>
<th>ΔCFI</th>
<th>ΔTLI</th>
<th>ΔRMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configural invariance</td>
<td>1002.380 (468)*</td>
<td>.951</td>
<td>.937</td>
<td>.063</td>
<td>[.057-.068]</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Metric/weak invariance</td>
<td>1118.679 (564)*</td>
<td>.950</td>
<td>.945</td>
<td>.058</td>
<td>[.053-.063]</td>
<td>186.858 (96)*</td>
<td>.001</td>
<td>.008</td>
<td>-.005</td>
</tr>
<tr>
<td>Scalar/strong invariance</td>
<td>1263.044 (714)*</td>
<td>.950</td>
<td>.957</td>
<td>.052</td>
<td>[.047-.056]</td>
<td>242.504 (150)*</td>
<td>.000</td>
<td>.012</td>
<td>-.006</td>
</tr>
<tr>
<td>Strict invariance</td>
<td>1319.190 (768)*</td>
<td>.950</td>
<td>.960</td>
<td>.050</td>
<td>[.045-.054]</td>
<td>102.247 (54)*</td>
<td>.000</td>
<td>.003</td>
<td>-.002</td>
</tr>
<tr>
<td>Latent variance invariance</td>
<td>1220.212 (780)*</td>
<td>.960</td>
<td>.969</td>
<td>.044</td>
<td>[.039-.049]</td>
<td>19.610 (12)</td>
<td>.010</td>
<td>.009</td>
<td>-.006</td>
</tr>
<tr>
<td>Latent means invariance</td>
<td>1259.435 (792)*</td>
<td>.958</td>
<td>.967</td>
<td>.045</td>
<td>[.040-.050]</td>
<td>33.151 (12)*</td>
<td>-.002</td>
<td>-.002</td>
<td>.001</td>
</tr>
</tbody>
</table>

**Notes.** *p<0.01; χ²: 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; MDΔχ²: chi-square difference test calculated with the Mplus DIFFTEST function for the robust weighted least square estimator (WLSMV). The fact that WLSMV χ² values are not exact, but "estimated" as the closest integer necessary to obtain a correct p-value explains the fact that the χ² and resulting CFI values can be non-monotonic with model complexity.