Running head: Teacher ratings of ADHD

Teacher ratings of the ADHD-RS IV in a community sample: Results from the ChiP-ARD study

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

This is the prepublication version of the following manuscript:

Caci, H., Morin, A.J.S., & Tran, A. (2016). Teacher ratings of the ADHD-RS IV in a community sample: Results from the ChiP-ARD study. *Journal of Attention Disorders*, 20 (5), 434-444. DOI: 10.1177/1087054712473834

[*]*Corresponding author*
[1]*Hôpitaux Pédiatriques de Nice CHU Lenval*
57, avenue de la Californie
06200 Nice
FRANCE
Phone: +33 (0)4 92 03 08 49
Fax: +33 (0)4 92 03 05 65
Email: caci.h@pediatrie-chulenval-nice.fr

[2]*Centre for Positive Psychology and Education*
University of Western Sydney
Australia

ABSTRACT

Objectives: Validated instruments to assess ADHD are still unavailable in many languages other than English for teachers, which constitutes a clear obstacle to screening, diagnosis and treatment of ADHD in many European countries. **Method:** Teachers rated 892 youths using the ADHD rating scale (ADHD-RS). We investigated the factor structure, reliability, and measurement invariance based on Confirmatory Factor Analyses. **Results:** Results support a bifactor model including one general ADHD factor and two specific Inattention and Hyperactivity-Impulsivity factors. But the latter is improperly defined calling into question the existence of a Predominantly Hyperactivity-Impulsivity subtype. The measurement invariance is fully supported across gender, age groups and gender x age groups. **Conclusion:** Results supports the multiple pathways hypothesis and suggests that a total ADHD score is meaningful, reliable and valid, as well as specific assessments of Inattention. Some youths – especially older ones – may present a profile of ADHD particularly marked by Inattention symptoms.

Keywords*— ADHD, Bifactor model, Rating scales, Children, Adolescent, Teacher rating*

Biographical statements

Hervé CACI, MD, PhD is a child and adolescent psychiatrist interested in personality, chronobiology and pharmacotherapy. He translated into French several instruments related to ADHD (ADHD-RS, SWAN, ASRS, WURS, etc.) and his currently conducting validation studies in children and adults, both in clinical and community samples.

Alexandre J.S. Morin, PhD defines himself as a lifespan developmental psychologist with broad research interests anchored in the exploration of the social determinants of psychological wellbeing and psychopathologies at various life stages. Most of his research endeavors are anchored into a substantive-methodological synergy framework and thus represent joint ventures in which new methodological developments are applied to substantively important issues.

Antoine TRAN, MD is a pediatrician with strong interests in biostatistics and epidemiology.

Attention-Deficit Hyperactivity Disorder (ADHD) is now recognized as a pervasive neurodevelopmental disorder that tends to persist well into adulthood and to be associated with a broad range of negative life outcomes (Faraone, Biederman, & Mick, 2006; Kooij et al., 2010). However, pointing to the need for efficient screening procedures, ADHD is also responsive to treatment (Hodgkins et al., 2012; Shaw et al., 2012). 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). A metaregression performed in a set of 102 carefully selected international studies estimated the worldwide prevalence of ADHD to be 5.29% [95% CI: 5.01-5.56] (Polanczyk, Silva de Lima, Lessa Horta, Biederman, & Rohde, 2007). According to DSM-IV, three types of ADHD can be distinguished according to whether the predominant symptoms are predominantly characterized by inattention, hyperactivity-impulsivity, or both (American Psychiatric Association, 1994).

Teachers can provide clinicians with important information regarding the child’s behaviour and performance at school, like parents would do at home (Sayal & Goodman, 2009). Although it is common to observe discrepancies between observers when rating ADHD symptoms (e.g. parents and teachers (Rettew et al., 2011)) this information is crucial to proper diagnostic procedures that require behavioural disturbances to be documented in more than one setting. Also, this information is useful to monitor the evolution of children diagnosed with ADHD during treatment. Such inter-professional communications could clearly be facilitated by the reliance on a validated, easy-to-use, behavioural observation rating scale for ADHD symptoms. Unfortunately, no such validated scale exists for French-speaking teachers, or professionals. Knowing that laypersons tend to lack information regarding ADHD, this creates a significant obstacle to research, communication, and practice in French-speaking countries. In fact, French is the official language in 32 countries and territories worldwide (Francophonie), 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.

The ADHD Rating Scale-IV (ADHD-RS) is the most commonly used measure of ADHD symptoms (DuPaul et al., 1997), and has already been successfully validated into many other languages (Döpfner et al., 2006; Magnusson, Smari, Gretarsdottir, & Pradardot, 1999; Szomlaiski et al., 2009; Zhang, Faries, Vowles, & Michelson, 2005). This instrument includes 18 items rated on a 4-point scale (0- *rarely or never* to and 3 *very often*) and parallel versions exist for clinicians, teachers and parents. Even-numbered items represent the 9 Inattention criteria of DSM-IV (e.g. "easily distracted") and odd-numbered items represent the 9 Hyperactivity-Impulsivity criteria (e.g. "leaves seat"). The three symptoms of the DSM-IV specific to impulsivity are numbered 14, 16 and 18 ("blurts out answers", "difficult waiting turn" and "interrupts", respectively).

There have been several publications regarding the ADHD-RS psychometric properties rated by teachers (DuPaul, et al., 1997), parents (DuPaul et al., 1998) or clinicians (Magnusson, et al., 1999; Zhang, et al., 2005). In these studies, Exploratory Factor Analyses (EFA) generally contrasted one- (ADHD), two- (Inattention and Hyperactivity-Impulsivity) or three- (Inattention, Hyperactivity, and Impulsivity) factor solutions (Döpfner, et al., 2006; DuPaul, et al., 1998; DuPaul, et al., 1997). Additional studies rather tried to contrast the fit to the data of *a priori* solutions using Confirmatory Factor Analyses (CFA), and these studies generally supported a two-factor structure (Inattention and Hyperactivity-Impulsivity) for the ADHD-RS in both clinical and community samples, and cross-culturally (Davis, Cheung, Takahashi, Shinoda, & Lindstrom, 2011; Gomez, Harvey, Quick, Scharer, & Harris, 1999; Martel, von Eye, & Nigg, 2010; Ohnishi, Okada, Tani, Nakajima, & Tsujii, 2010; Wolraich et al., 2003). The reported scale score reliability coefficients (i.e. Cronbach's α) of the resulting Inattention (.95) and the Hyperactivity-Impulsivity (.94) factors are generally high when rated by teachers (Gomez, et al., 1999).

In psychiatric measurement, the main question is whether a primary dimension (e.g. depression, anxiety, etc.) does exist as a unitary disorder including specificities (i.e. as represented by a bifactor model), or whether these specificities rather define distinct facets without a common core (i.e. represented by a classical CFA model). Recently, this key conceptual issue has been questioned for ADHD. First, ADHD has been found to represent a relatively stable condition across the lifespan that persists at least well into adulthood, although the specific manifestations of this condition 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) manifestations that fluctuates over time and contexts (Martel, et al., 2010). This distinction is also consistent with the way ADHD is defined in the DSM-IV, with a core G set of ADHD manifestations leading to the main diagnosis, but specificities of individuals leading them to fit more closely to the inattentive, hyperactive-impulsive, or combined subtypes. Within the framework of CFA, a bifactor model (Holzinger & Swineford, 1937) whereby each item is simultaneously defined by one generic G ADHD factor and one subtype-specific S factor (hyperactivity/impulsivity or inattention) would be particularly well-suited to this possibility. More precisely, a bifactor model first analyses the total covariance among the items to extract a global G factor underlying all items, and then models the residual covariance not explained by the G factor through the specific S factors.

The few studies that contrasted classical CFA models with bifactor models in studying ADHD symptoms generally supported a bifactor solution including one ADHD G-factor and two specific (Inattention and Hyperactivity-Impulsivity) S-factors among: (a) a mixed clinical-community population of children rated with the teacher version of the ADHD-RS and parental reports on other instruments (Martel, et al., 2010); (b) among clinical (Toplak et al., 2009) or community (Normand, Flora, Toplak, & Tannock, 2012; Ullebø, Breivik, Gillberg, Lundervold, & Posserud, 2012 (Ahead of print)) samples of children rated with other instruments; (c) among community samples of adults rated with other instruments (Caci, Oliveri, & Dollet, 2011). However, these studies are still few and deserve replication, particularly in large community samples where the screening utility of the ADHD-RS needs to be maximized. In particular, although they all supported bifactor solutions, these studies also report that both of the S factors explained relatively little variance in ADHD ratings and systematically showed that at least one of the subtype-specific S factor was weakly defined, calling into question the appropriateness of some diagnostic subtypes of ADHD. Unfortunately, these studies also disagreed as to whether it was the Inattention (Toplak, et al., 2009), the Hyperactivity-Impulsivity (Toplak, et al., 2009; Ullebø, et al., 2012 (Ahead of print)), or both (Martel, et al., 2010) S-factors that posed problem, reinforcing the need for replication. In particular, two studies showed that the conclusions did not change based on the informant (parent versus children), but rather according the nature of the instrument, so that interview ratings resulted in an undefined Inattention S-factor, whereas questionnaire data resulted in an undefined Hyperactivity-Impulsivity factor (Toplak, et al., 2009).

Another important issue that has yet to be systematically investigated has to do with the critical assumption that the various versions of the ADHD-RS 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). In regards to ADHD measurement based on teacher ratings, this verification is particularly important. Indeed, as we previously noted, the specific manifestations of ADHD are known to differ as a function of age and genders (Barkley, Murphy, & Fischer, 2008; Faraone, et al., 2006; Faraone et al., 2006), while the generic assumption is that the common core of the ADHD construct remains the same. Teachers also tend to be more aware of boys disturbing behaviours in the classroom than of girls who tend to disturb differently. Thus, they may provide less reliable ratings of girls ADHD.

In summary, this paper aims to investigate the psychometric properties of the ADHD-RS rated by teachers in order to conduct four specific verifications:

1. How well does the *a priori* two-factor structure of the ADHD-RS (mimicking the DSM-IV subtypes) fit the ratings provided by French teachers?
2. Will a bifactor model provide a better representation of ADHD-RS ratings by teachers, as suggested by some previous studies based on ADHD symptoms?
3. Is the ADHD-RS reliable when rated by French teachers?
4. Is the ADHD-RS measurement model invariant across genders, age groups, and gender-by-age groups?

METHODS

Participants and Material

This paper uses data from the ChiP-ARD (*Children and Parents with ADHD and Related Disorders*) study, targeting French children and adolescents from the general population aged between 4 to 18 years old. The ChiP-ARD study was conducted in 20 kindergarten schools (*pré-élémentaires* or *maternelles*), 30 primary schools (*élémentaires*), 14 secondary schools (*colleges and lycées*) from

Southern France (Nice). The data was collected in spring 2010 and 2011, during two distinct (non-longitudinal) waves of data collection. Overall, 262 teachers participated in the study (mean age=43.9; S.D.=8.6; range=24-61), forty-seven were males (17.94%). A letter was randomly drawn from the alphabet for each class and the teacher was asked to include 2 to 4 youths whose name began with this letter (or the next one if no name matched the random letter, and starting over at letter ‘A’ if letter ‘Z’ was reached). Parents had to return a signed consent form that was kept anonymous by teachers who allocated them upon reception an 8-digit unique identifier. Teachers thus provided ratings of 132 youths in kindergarten (64 girls, 48.49%), 349 youths in primary schools (174 girls, 49.86%), and 411 youths in secondary schools (220 girls, 53.53%). Overall, the sample comprised 892 youths, including 458 girls (51.35%), with a mean age of 10.59 (S.D.=3.50) for girls and 10.18 (S.D.=3.32) for boys (t(890)=1.829, n.s.). 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*.

The French version of the teacher version of the ADHD-RS was developed through classical translation-back translation procedures by members of the research team and the resulting back-translated English was compared to the original version for final adjustments by the main author of the original ADHD-RS (i.e. DuPaul).

Statistical analyses

The main models were estimated with Mplus 6.12 (L. K. Muthén & Muthén, 2010), 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; B. O. Muthén, du Toit, & Spisic, 1997).

The fit of five *a priori* alternative models of teachers answers to the ADHD-RS instrument 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: M3), a bifactor model including one ADHD G-factor and two specific S-factors (Inattention and Hyperactivity-Impulsivity: M4), and a bifactor model including one ADHD G-factor and three specific S-factors (Inattention, Hyperactivity, and Impulsivity: M5).

Measurement invariance tests across gender (male versus females), age groups (defined as children younger than 12 years old versus adolescents aged over 12 years old), and combinations of gender and age groups were performed in a sequential strategy following Meredith recommendations (Meredith, 1993) as adapted for ordered-categorical items by Millsap & Tein (Millsap & Tein, 2004) (see also (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-square 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 .95 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 .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 (F.F. 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. A supplementary file was prepared to accompany this paper in which annotated input codes used to implement these models in Mplus are provided (for the final bifactor model as well as for the full sequence of tests of invariance across gender groups). This file is available upon requests from the first and second authors.

RESULTS

Confirmatory Factor Analyses and Reliability

The single factor model (M1) showed the worst fit to the data (Table 1). Both the two-factor (M2) and three-factor (M3) models presented a satisfactory level of fit to the data (CFI and TLI>.95; RMSEA<.08), though the improvement in fit related to the addition of an Impulsivity factor remained well below the recommended value for differences in these indices. The estimated M3 correlation between the Hyperactivity and Impulsivity factors was also high enough (.813) to call into question their distinctiveness. In the M2 model, the estimated latent factor correlation between the Inattention and Hyperactivity-Impulsivity factors was more reasonable in size (.560), but still suggested the presence of a common core of ADHD symptoms, justifying the investigation of bifactor models.

Accordingly, the fit to the data of two *a priori* bifactor models was also estimated, one based on two specific S factors and one global G factor (M4) and one based on three S factors and one G factor (M5). The comparison once again supported the more parsimonious solution M4 – showing that it presented a similar, yet slightly decreased (-.001 for CFI and TLI and +.001 for RMSEA), level of fit to the data. Since the bifactor model M4 fitted data better than the more classical model M2, this model 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 (children versus adolescents) and gender by age groups (female children or adolescents, and male children or adolescents) with CFI and TLI>.95 and RMSEA<.06.

Table 2 presents the parameters estimates for this final model (M4) and for the comparison model (M2). Both factors are well defined with items presenting very strong and significant factor loadings (λ =.802 to .942) on their respective factors and a high level of communality (h^2=.643 to

.887), suggesting low level of measurement errors as reflected in items' uniquenesses (δ=1-h^2). These results are also observed for model M4 since both models include the same specific factors. Furthermore, the standardized loadings on the ADHD G factor in model M4 are also moderately strong and significant (λ=.553 to .937), suggesting a well-defined common core of ADHD symptoms. Finally, the standardized loadings are high on the specific Inattention factor (λ=.464 to .726), albeit smaller than in M2 and very weak (items 12, 14, 16, 18; λ=.284 to .406), non-significant (items 4, 6, 8 and 10), or even negative (item 2, λ=-.168) on the specific Hyperactivity-Impulsivity factor. This shows that once the common core of ADHD symptoms is taken into account by the G factor, there remains a substantial level of covariance in the items that is explained by a specific Inattention factor but not by a specific Hyperactivity-Impulsivity factor. Therefore, Hyperactivity-Impulsivity symptoms apparently mostly serve to define the ADHD G factor. In fact, the standardized loadings are so low as to suggest that all of the specificity remaining in these items seems to be linked with unreliability in teachers' ratings. This result calls into question the DSM-IV Hyperactive-Impulsive subtype.

Looking at the scale score reliability, Cronbach's α coefficients appear to be quite high for all factors (.931 to .949), and equivalent in both models M2 and M4 (Table 2).. This is due to the specific, and inadequate in this case, manner in which α computes composite reliability (Sijtsma, 2009). 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 such as used in the present study (McDonald, 1970). Expectedly, coefficients ω converge with coefficients α in model M2. However, when the specificities of the bifactor model M4 are taken into account, coefficients ω revealed a very high level of reliability of the global ADHD ratings (ω=.981) when these are modeled 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=.885). However, the scale score reliability estimate of the Hyperactivity-Impulsivity S-factor is much lower (ω=.454), confirming our previous interpretation that their specificity is mostly due to random noise (i.e. unreliability) in ratings of these symptoms by teachers – not in themselves, but once the common core of ADHD ratings (represented by the G factor) are taken into account.

Measurement Invariance

Starting from the bifactor model M4, 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>.95 and RMSEA<.06. The tests of metric/weak, scalar/strong, strict, and latent variance invariance across gender are fully supported.. In many cases, the fit indices incorporating a control for model parsimony (i.e. TLI and RMSEA) improve when invariance constraints are added to the model; the more restricted model with strict invariance and invariance of the latent variances even shows a substantially higher degree of fit to the data than the baseline model (TLI=.998 versus .987 and RMSEA=.022 versus .053). Furthermore, when equality constraints are placed on the latent means, the MD $\Delta\chi^2$ is significant, the ΔRMSEA (.020) is greater than the recommended cut-off of .015, and the ΔCFI, ΔTLI are larger than in the other models. We thus systematically probed these differences (Table 4). When girls' latent means are fixed to 0 for identification purposes, boys' latent means (expressed as differences in SD units from girls' means) are significantly higher on the ADHD G factor (M=.483; s.e.=.089; p<.01), non-significantly different on the Inattention S factor (M=.132; s.e.=.094; p>.05), and significantly lower on the Hyperactivity-Impulsivity S factor (M=-.334; s.e.=.125; p<.01). 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, girls' present higher levels on the residual ratings related to the specific Hyperactivity-Impulsivity factor that was previously showed to be highly unreliable. This suggests that, for girls, Hyperactivity-Impulsivity ratings tend to have a greater tendency to be interpreted as something different from a generic ADHD syndrome.

Before moving on to tests of measurement invariance according to age-groups, and age by gender groups, the items had to be recoded from their original 4 categories answers scales (0 to 4) into a three category answer scale through collapsing the two highest categories. Indeed, an important assumption of models based on ordered-categorical items is that the same number of answer categories is used in all groups, an assumption that is violated when there are empty cells due to one specific answer categories not being used in a specific group. Empty cells are common situation in analyses of ordered-categorical items that is classically solved by collapsing of adjacent answer categories (Lubke & Muthén, 2004; Morin, et al., 2009; Reise, Morizot, & Hays, 2007). In the present study, empty cells were mostly linked to reduced sample sizes in some of the subgroups, causing some empty cells at the highest level (i.e. answer category 4) of the original answering scale. In order to ensure that no bias results from this procedure, all of the previous models were fully replicated with this new coding scheme and the results proved to be equivalent to those reported here.

The metric/weak, scalar/strong, strict, and latent variance invariance assumptions fully hold across age groups and age by gender groups. Although some of the MD $\Delta\chi^2$ tests come up as significant for the models based on age groups, they remained small in magnitude and not supported by the observed changes in fit indices, suggesting that their significance may simply reflect chi-square's known oversensitivity to minor model misspecification and sample size. Examination of the modification indices associated with these models confirms this interpretation. However, once again the results suggest that it may be appropriate to look at age related differences in the estimated factors (significant and large, in relation to the model degrees' of freedom MD $\Delta\chi^2$ and higher than usual Δ RMSEA of .008, albeit still under the suggested cut-off score of .015). Compared to children's, adolescents' latent means are significantly lower on the ADHD G factor (M=-.357; s.e.=.089; p<.01), non-significantly different on the Hyperactivity-Impulsivity S factor (M=.181; s.e.=.128; p>.05), and significantly higher on the Inattention S factor (M=.724; s.e.=.102; p<.01). While the measurement model underlying teachers responses to the ADHD-RS remains perfectly invariant (unbiased) in children and adolescents, our expectations that ADHD manifestations change with age are confirmed with regard to the generic ADHD and Inattention levels. Finally, when looking at mean-levels differences based on gender by age groups combinations, the results essentially replicate the previous results (Table 4). That is (1) levels on the Inattention S factor tend to increase with age but are equivalent across gender-groups; (2) levels on the Hyperactivity-Impulsivity S factor tend to be lower for male children only, but equivalent across the other groups; (3) levels on the ADHD G factor tend to decrease with age, but also to be higher for males.

DISCUSSION

This paper is the first to thoroughly assess the structure of the ADHD-RS in a large French community sample of youths rated by their teachers. 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 the proposed two-factor bifactor model.

Interestingly, when separate factors (M3) or separate specific S factors (M5) were estimated to differentiate Hyperactivity from Impulsivity symptoms, the resulting models did not provide a better fit to the data and suggest a very high correlation between these two factors. This result is in line with those from previous studies showing consistency across rating scales, settings and culture (Amador-Campos, Forns-Santacana, Martorell-Balanzo, Guardia-Olmos, & Pero-Cebollero, 2005; Burns, Boe, Walsh, Sommers-Flanagan, & Teegarden, 2001; Wolraich, et al., 2003). In fact, only two studies retained the 3-factor structure and both reported a very high factor correlation between these two factors (r=.64 to .80) (Gomez, et al., 1999; Span, Earleywine, & Strybel, 2002).

The bifactor structure that we retained has received substantial support in the last five years (Martel, Roberts, Gremillion, von Eye, & Nigg, 2011; Martel, et al., 2010; Toplak, et al., 2009; Toplak et al., 2011 (Ahead of print); Ullebø, et al., 2012 (Ahead of print)), but is still not widely used. Also in line with the results from some of these preceding studies, we found that the items apparently all contribute to properly define a common core of generic ADHD symptoms, as well as a specific Inattention factor. However, we found that once the covariance between items is taken into account by the ADHD general factor, only the Inattention specific factor remains meaningful and most of the covariance modelled in the Hyperactivity-Impulsivity specific factor may be attributed to unreliability in teacher ratings. This result is in line with previous questionnaires studies of ADHD symptoms (Martel, et al., 2011; Martel, et al., 2010; Normand, et al., 2012; Toplak, et al., 2009; Ullebø, et al., 2012 (Ahead of print)) and calls into question the validity of the Hyperactive-Impulsive subtype.

A bifactor model suggests that there are distinct etiological influences that converge on the same core syndrome (Chen, West, & Sousa, 2006) with some remaining specificities. Thus, the bifactor model retained in the present study is in line with multiple-pathways conceptions of ADHD (Nigg, Goldsmith, & Sachek, 2004; Sonuga-Barke, 2002, 2005), at least regarding the development of a specific subtype of ADHD presenting elevated Inattention levels, but not necessarily elevated Hyperactivity-Impulsivity levels. More precisely, our results also show that Hyperactivity-Impulsivity and Inattentive symptoms merge together to define a global, general, condition of ADHD, whereas Inattentive symptoms may appear on their own accord, potentially linked to different causal pathways. For clinicians, this means, that patients can be placed on a continuum with regard of their total score on the ADHD-RS and that specific *dimensional* evaluations of inattention levels would provide valuable additional information. In these patients with marked Inattentive levels, 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, et al., 2011). Similarly, additional studies are needed to examine the changes over time in these ratings, as well as their state and trait components (Normand, et al., 2012). Finally, and most importantly, additional results are needed to explore the differentiated results that are obtained based on questionnaires, versus interview data, and the reasons for these differences (Toplak, et al., 2009; Toplak, et al., 2011 (Ahead of print)).

Scale score reliability estimates for the ADHD-RS confirm that the global ADHD G factor (ω =.981), as well as the specific Inattention S factor (ω =.885) present satisfactory reliability levels when properly estimated by model based methods taking into account the specificities of the bifactor model. These values are fully in line with previous estimates (Danforth & DuPaul, 1996; DuPaul, et al., 1997). However, the reliability estimate of the Hyperactivity-Impulsivity S-factor is much lower (ω =.454), confirming that apparent specificity in these ratings is mostly due to unreliability once the common core of ADHD ratings are taken into account. The present study is, to our knowledge, the first study based on a bifactor model of ADHD to report proper model-based estimates of reliability.

Measurement invariance of the ADHD-RS

A further objective of this study was to investigate the measurement invariance of this final bifactor model. We thus verified whether group membership (gender, age, and age-by-gender groups) introduced any measurement bias in teachers' ratings 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 expected mean-level differences across subgroups. We found that levels on the specific Inattention factor tended to increase with age in both gender groups. This may reflect the interaction between pupils' abilities and the increasing difficulty with grades. In our clinical practice, we often notice that teachers interpret inattention difficulties as a marker for "immaturity", which is more than rarely the reason invoked to justify a repeating a grade or, when the pupil is old enough, to argue for a orientation toward special needs schools or professional. This is fully in line with previous studies showing that pupils with predominantly inattentive ADHD are generally diagnosed much later than pupils with combined ADHD (Solanto, 2000). A second finding of this study is that male children exhibit lower levels on the specific Hyperactivity-Impulsivity, whereas female adolescents present higher levels. This unexpected result may be related to the lack of reliability observed in these specific S Hyperactivity-Impulsivity ratings made by teachers. Alternatively, it may also suggest that teachers more easily excuse disturbing behaviours as expected from male children, but are more concerned when older female students exhibit such unusual behaviours. At last, latent means comparisons show that levels on the general ADHD factor decrease with age and are higher for males. This is directly in line with epidemiological results in which the boy:girl ratio is commonly reported that ADHD to be around 3. Similarly, the observed age-related trend is in line with the fact that inhibition abilities tend to increase with age making general ADHD symptoms less intense.

CONCLUSION

Based on a large community sample of French children and adolescents, our data showed that French teachers, even knowing that they tend not to be familiar with ADHD, can reliably rate the French version of the ADHD-RS. However, these results also call into question the existence, and reliability, of a subtype of ADHD mostly characterized by Hyperactive-Impulsive characteristics.

Acknowledgements

This study was founded by a grant to the first author from the French Health Ministry and is recorded on clinicaltrials.gov under the reference NCT01260792. The authors are grateful to Dr. Eric FONTAS, Vanina OLIVERI and Kevin DOLLET for their help in the data collection process, to the *Inspection Académique des Alpes-Maritimes* and the *Rectorat des Alpes-Maritimes et du Var* for their support, and to the teachers, pupils and parents for participating in this study.

The authors declare they have no conflict of interest.

REFERENCES

Amador-Campos, J. A., Forns-Santacana, M., Martorell-Balanzo, B., Guardia-Olmos, J., & Pero-Cebollero, M. (2005). Confirmatory factor analysis of parents' and teachers' ratings of DSM-IV symptoms of Attention Deficit Hyperactivity Disorder in a Spanish sample. *Psychological Reports, 97*(3), 847-860.

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: The Guilford Press.

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.

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

Burns, G. L., Boe, B., Walsh, J. A., Sommers-Flanagan, R., & Teegarden, L. A. (2001). A confirmatory factor analysis on the DSM-IV ADHD and ODD symptoms: What is the best model for the organization of these symptoms? *Journal of Abnormal Child Psychology, 29*, 339-349.

Caci, H., Oliveri, V., & Dollet, K. (2011). *Psychometric properties of the ASRS in a general population: findings from the ChiP-ARD study.* Paper presented at the 3rd International Congress on ADHD.

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*(2), 189-225.

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

Danforth, J. S., & DuPaul, G. J. (1996). Interrater reliability of teacher rating scales for children with Attention-Deficit Hyperactivity Disorder. *Journal of Psychopharmacology and Behavioral Assessment, 18*(3), 227-237.

Davis, J. M., Cheung, S. F., Takahashi, T., Shinoda, H., & Lindstrom, W. A. (2011). Cross-national invariance of Attention-Deficit/Hyperactivity Disorder factors in Japanese and U.S. university students. *Research in Developmental Disabilities, 32*, 2972-2980.

Döpfner, M., Steinhausen, H.-C., Coghill, D., Dalsgaard, S., Poole, L., Ralston, S. J., et al. (2006). Cross-cultural reliability and validity of ADHD assessed by the ADHD Rating Scale in a pan-Europen study. *European Child and Adolescent Psychiatry, 15*(Suppl.1), I/46-I/55.

DuPaul, G. J., Anastopoulos, A. D., Power, T. J., Reid, R., Ikeda, M., & McGoey, K. (1998). Parent ratings of attention-deficit/hyperactivity disorder symptoms: factor structure and normative data. *Journal of Psychopathology and Behavioral Assessment, 20*(1), 83-102.

DuPaul, G. J., Power, T. J., Anastopoulos, A. D., Reid, R., McGoey, K., & Ikeda, M. (1997). Teacher ratings of attention-deficit/hyperactivity disorder: factor structure and normative data. *Psychological Assessment, 9*, 436-444.

Faraone, S. V., Biederman, J., & Mick, E. (2006). The age-dependent decline of attention-deficit hyperactivity disorder: a meta-analysis of follow-up studies. *Psychological Medicine, 36*(2), 159-165.

Faraone, S. V., Biederman, J., Spencer, T., Mick, E., Murray, K., Petty, C., et al. (2006). Diagnosing adult attention deficit hyperactivity disorder: are late onset and subthreshold diagnoses valid? *American Journal of Psychiatry, 163*, 1720-1729.

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*(4), 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.

Francophonie, O. I. d. l. Organisation International de la Francophonie [International Organisation of La Francophonie]. Retrieved June 14th, 2012

Gomez, R., Harvey, J., Quick, C., Scharer, I., & Harris, G. (1999). DSM-IV AD/HD: confirmatory factor models, prevalence, and gender and age differences based on parent and teacher ratings of Australian primary school children. *Journal of Child Psychology and Psychiatry, 40*(2), 265-274.

Hodgkins, P., Arnold, L. E., Shaw, M., Caci, H., Kahle, J., Woods, A. G., et al. (2012). A systematic review of long-term outcomes in ADHD: global publication trends. *Frontiers in Psychiatry, 2*(84), 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), 1-55.

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.

Lubke, G. H., & Muthén, B. O. (2004). Applying Multigroup Confirmatory Factor Models for Continuous Outcomes to Likert Scale Data Complicates Meaningful Group Comparisons. . *Structural Equation Modeling, 11*, 514-534.

Magnusson, P., Smari, J., Gretarsdottir, H., & Pradardot, H. (1999). Attention-Deficit/Hyperactivity symptoms in Icelandic schoolchildren: assessment with the Attention Deficit/Hyperactivity Rating Scale-IV. *Scandinavian Journal of Psychology, 40*, 301-306.

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*(8), 905-914.

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

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 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 [Revue d'Épidémiologie et de Santé Publique], 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. Mplus Technical Reports (http://gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf).

Muthén, L. K., & Muthén, B. O. (2010). *Mplus User's Guide. Sixth 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), 555.

Ohnishi, M., Okada, R., Tani, I., Nakajima, S., & Tsujii, M. (2010). Japanese version of school form of the ADHD-RS: an evaluation of its reliability and validity. *Research in Developmental Disabilities, 31*, 1305-1312.

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*(6), 942-948.

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*, 19-31.

Rensvold, R. B., & Cheung, G. W. (1998). Testing measurement model for factorial invariance: A systematic approach. *Educational and Psychological Measurement, 58*(6), 1017-1034.

Rettew, D. C., van Oort, F. V., Verhulst, F. C., Buitelaar, J. K., Ormel, J., Hartman, C. A., et al. (2011). When parent and teacher ratings don't agree: the Tracking Adolescents' Individual Lives Survey (TRAILS). *Journal of Child and Adolescent Psychopharmacology, 21*(5), 389-397.

Sayal, K., & Goodman, R. (2009). Do parental reports of child hyperkinetic disorder symptoms at school predict teacher rating? *European Child and Adolescent Psychiatry, 18*(6), 336-344.

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*(99).

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*(6), 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. S. (2005). Causal models of Attention-Deficity/Hyperactivity Disorder: from common simple deficits to multiple developmental pathways. *Biological Psychiatry, 57*, 1231-1238.

Span, S. A., Earleywine, M., & Strybel, T. Z. (2002). Confirming the factor structure of Attention Deficit Hyperactivity Disorder symptoms in adult, nonclinical samples. *Journal of Psychopharmacology and Behavioral Assessment, 24*, 129-136.

Szomlaiski, N., Dyrborg, J., Rasmussen, H., Schumann, T., Koch, S. V., & Bilenberg, N. (2009). Validity and clinical feasibility of the ADHD rating scale (ADHD-RS): A Danish nationwide multicenter study. *Acta Paediatrica, 98*(2), 397-402.

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.

Toplak, M. E., Sorge, G. B., Flora, D. B., Chen, W., Banaschewski, T., Buitelaar, J., et al. (2011 (Ahead of print)). The hierarchical factor model of ADHD: Invariant across age and national groupings? *Journal of Child Psychology and Psychiatry*.

Ullebø, A. K., Breivik, K., Gillberg, C., Lundervold, A. J., & Posserud, M.-B. (2012 (Ahead of print)). The factor structure of ADHD in a general population of primary school children. *Journal of Child Psychology and Psychiatry*.

Wolraich, M. L., Lambert, E. W., Baumgaertel, A., Garcia-Tornel, S., Feurer, I. D., Bickman, L., et al. (2003). Teacher's screening for Attention Deficit/Hyperactivity Disorder: comparing multinational samples on teacher ratings of ADHD. *Journal of Abnormal Child Psychology, 31*(4), 445-455.

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.

Zhang, S., Faries, D. E., Vowles, M., & Michelson, D. (2005). ADHD Rating Scale IV: psychometric properties from a multinational study as a clinician-administered instrument. *International Journal of Methods in Psychiatric Research, 14*(4), 186-201.

Table 1. Fit indices for the CFA models (WLSMV estimator, N=892).

	$\chi 2(df)$	CFI	TLI	RMSEA	RMSEA 90% CI
Models estimated on the full sample (*n*=892)					
M1: One-factor model	1861.99 (135)*	0.943	0.935	0.120	[0.115 ; 0.125]
M2: 2-factor (oblique)	800.48 (134)*	0.978	0.975	0.075	[0.070 ; 0.080]
M3: 3-factor (oblique)	751.73 (132)*	0.979	0.976	0.073	[0.068 ; 0.078]
M4: 2-factor (bifactor)[a]	421.30 (117)*	0.990	0.987	0.054	[0.048 ; 0.060]
M5: 3-factor (bifactor)[a]	436.22 (117)*	0.989	0.986	0.055	[0.050 ; 0.061]
Model 4 estimated in the subsamples					
M4f: Females (*n*=458)	255.432 (117)*	.989	.985	.051	[0.042 ; 0.059]
M4m: Males (*n*=434)	273.224 (117)*	.990	.987	.055	[0.047 ; 0.064]
M4c: Children (*n*=578)	319.908 (117)*	.990	.987	.055	[0.048 ; 0.062]
M4a: Adolescents (*n*=314)	193.141 (117)*	.994	.992	.046	[0.034 ; 0.057]
M4fc: Female children (*n*=288)	186.339 (117)*	.993	.990	.045	[0.033 ; 0.057]
M4fa: Female adolescents (*n*=170)	150.700 (117)	.994	.992	.041	[0.018 ; 0.059]
M4mc: Male children (*n*=290)	235.692 (117)*	.990	.986	.059	[0.048 ; 0.070]
M4ma: Male adolescents (*n*=144)	134.073 (117)	.998	.997	.032	[0.000 ; 0.054]

Notes. 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. *p<0.01.

[a] Bifactor models based on the same items but including any number of G or S factors will always present the same degrees of freedom. More precisely, for each item, two loadings and one uniqueness are estimated, and no latent covariance is estimated, meaning that the total number of factors has no impact on the model's degrees of freedom (latent variances may be estimated, but the loading of one referent indicator per latent factor then need to be fixed for identification purposes).

Table 2. Standardized Parameters Estimates for the Retained 2-Factor Correlated and Bifactor Models.

	2 correlated factors			Orthogonal Bifactor			
	I	H-I	h^2	G	I	H-I	h^2
1: Close attention	0.819 (0.016)		0.670 (0.026)	0.578 (0.030)	0.601 (0.030)		0.696 (0.026)
2: Fidgets		0.904 (0.013)	0.818 (0.024)	0.924 (0.015)		-0.168 (0.054)	0.882 (0.028)
3: Sustaining attention	0.942 (0.008)		0.887 (0.015)	0.778 (0.020)	0.497 (0.027)		0.852 (0.015)
4: Leaves seat		0.898 (0.013)	0.806 (0.024)	0.907 (0.014)		*-0.076 (0.053)*	0.828 (0.025)
5: Does not listen	0.822 (0.020)		0.676 (0.033)	0.663 (0.032)	0.464 (0.035)		0.655 (0.032)
6: Runs about		0.932 (0.013)	0.869 (0.024)	0.937 (0.013)		*-0.006 (0.053)*	0.878 (0.024)
7: No follow through	0.875 (0.013)		0.766 (0.024)	0.564 (0.035)	0.709 (0.027)		0.821 (0.020)
8: Difficult playing		0.909 (0.012)	0.826 (0.022)	0.913 (0.012)		*0.020 (0.048)*	0.835 (0.023)
9: Difficult organizing	0.892 (0.012)		0.796 (0.021)	0.573 (0.033)	0.726 (0.026)		0.854 (0.018)
10: On the go		0.888 (0.017)	0.788 (0.030)	0.888 (0.018)		*0.060 (0.052)*	0.793 (0.030)
11: Avoids tasks	0.836 (0.017)		0.699 (0.028)	0.553 (0.035)	0.662 (0.029)		0.745 (0.023)
12: Talks excessively		0.802 (0.020)	0.643 (0.031)	0.764 (0.028)		0.335 (0.055)	0.696 (0.030)
13: Loses things	0.861 (0.016)		0.741 (0.027)	0.682 (0.029)	0.509 (0.033)		0.725 (0.026)
14: Blurts out answers		0.844 (0.016)	0.712 (0.027)	0.792 (0.028)		0.406 (0.052)	0.792 (0.026)
15: Easily distracted	0.881 (0.011)		0.776 (0.020)	0.696 (0.025)	0.530 (0.029)		0.765 (0.020)
16: Difficult waiting turn		0.914 (0.012)	0.834 (0.021)	0.865 (0.024)		0.379 (0.053)	0.892 (0.020)
17: Forgetful	0.895 (0.0013)		0.801 (0.023)	0.685 (0.030)	0.575 (0.032)		0.800 (0.022)
18: Interrupts		0.926 (0.011)	0.857 (0.020)	0.895 (0.018)		0.284 (0.049)	0.882 (0.018)
Reliability (α)	0.931	0.937		0.949	0.931	0.937	
Reliability (ω)	0.938	0.941		0.981	0.885	0.454	

Notes. Standard errors are reported in parentheses; I: Standardized loadings on the Inattention Factor; H-I: Standardized loadings on the Hyperactivity-Impulsivity Factor; G: Standardized loadings on the Global ADHD factor; h^2: communality of the items; Italicized parameters estimates are non-signifgicant at p<.05 – all other parameters estimates are significant; α : scale score reliability estimate based on Cronbach coefficient alpha; ω : scale score reliability estimate based on McDonald coefficient omega.

Table 3. Tests of measurement invariance for the final 2-factor bifactor model.

	$\chi^2(df)$	CFI	TLI	RMSEA	RMSEA 90% CI	MD Δ $\chi^2(\Delta df)$	ΔCFI	ΔTLI	Δ RMSEA
Tests of measurement invariance across genders									
Configural invariance	526.58 (234)*	0.990	0.987	0.053	[0.047 ; 0.059]	—	—	—	—
Metric/weak invariance	507.53 (267)*	0.992	0.990	0.045	[0.039 ; 0.051]	41.49 (33)	+0.002	+0.003	-0.008
Scalar/strong invariance	529.08 (300)*	0.992	0.992	0.041	[0.036 ; 0.047]	49.28 (33)	0.000	+0.002	-0.004
Strict invariance	471.18 (318)*	0.995	0.995	0.033	[0.026 ; 0.039]	16.44 (18)	+0.003	+0.003	-0.008
Latent variance invariance	392.76 (321)*	0.997	0.998	0.022	[0.013 ; 0.030]	4.37 (3)	+0.002	+0.003	-0.009
Latent means invariance	583.61 (324)*	0.991	0.991	0.042	[0.037 ; 0.048]	62.30 (3)*	-0.006	-0.007	+0.020
Tests of measurement invariance across age groups									
Configural invariance	423.93 (234)*	0.994	0.992	0.043	[0.036 ; 0.049]	—	—	—	—
Metric/weak invariance	459.85 (267)*	0.994	0.993	0.040	[0.034 ; 0.046]	65.21 (33)*	0.000	+0.001	-0.003
Scalar/strong invariance	480.86 (282)*	0.994	0.993	0.040	[0.034 ; 0.046]	29.02 (15)	0.000	0.000	0.000
Strict invariance	495.96 (300)*	0.994	0.994	0.038	[0.032 ; 0.044]	42.09 (18)*	0.000	0.000	-0.002
Latent variance invariance	384.94 (303)*	0.997	0.997	0.025	[0.016 ; 0.032]	2.59 (3)	+0.003	+0.003	-0.013
Latent means invariance	446.69 (306)*	0.996	0.996	0.032	[0.025 ; 0.038]	23.39 (3)*	-0.001	-0.001	+0.008
*Tests of measurement invariance across age*gender groups*									
Configural invariance	626.97 (468)*	0.995	0.993	0.039	[0.031 ; 0.047]	—	—	—	—
Metric/weak invariance	725.26 (567)*	0.995	0.995	0.035	[0.027 ; 0.043]	132.06 (99)	0.000	0.002	-0.004
Scalar/strong invariance	782.06 (612)*	0.995	0.995	0.035	[0.027 ; 0.042]	69.14 (45)	0.000	0.000	0.000
Strict invariance	819.85 (666)*	0.995	0.996	0.032	[0.024 ; 0.039]	72.26 (54)	0.000	0.001	-0.003
Latent variance invariance	759.04 (675)*	0.997	0.998	0.024	[0.012 ; 0.032]	10.42 (9)	0.002	0.002	-0.008
Latent means invariance	978.99 (684)*	0.991	0.992	0.044	[0.038 ; 0.050]	91.72 (9)*	-0.006	-0.006	+0.020

Notes. 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 Δ χ^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 sometimes the χ^2 and resulting CFI values can be non-monotonic with model complexity; *p<0.01.

Table 4. Latent mean comparisons across groups defined on the basis of gender and age.

Factor	Latent means (s.e.) for female children	Latent means (s.e.) for female adolescents	Latent means (s.e.) for male children	Latent means (s.e.) for male adolescents
ADHD G factor	0	-0.26 (0.13)*	0.52 (0.10)**	0.12 (0.13)
Hyperactivity-Impulsivity S factor	0	0.18 (0.19)	-0.31 (0.16)	-0.13 (0.18)
Inattention S factor	0	0.59(0.13)**	0.07 (0.12)	0.89 (0.14)**
ADHD G factor	0.26 (0.13)*	0	0.77 (0.13)**	0.38 (0.15)*
Hyperactivity-Impulsivity S factor	-0.18 (0.19)	0	-0.49 (0.20)*	-0.32 (0.20)
Inattention S factor	-0.59 (0.13)**	0	-0.52 (0.14)**	0.304 (0.158) 0.054
ADHD G factor	-0.52 (0.10)**	-0.77 (0.13)**	0	-0.40 (0.12)**
Hyperactivity-Impulsivity S factor	0.31 (0.16)	0.49 (0.20)*	0	0.17 (0.17)
Inattention S factor	-0.07 (0.12)	0.52 (0.14)**	0	0.82 (0.14)**
ADHD G factor	-0.12 (0.13)	-0.38 (0.15)*	0.40 (0.12)**	0
Hyperactivity-Impulsivity S factor	0.13 (0.18)	0.32 (0.20)	-0.17 (0.17)	0
Inattention S factor	-0.89 (0.14)**	-0.30 (0.158) 0.054	-0.82 (0.14)**	0

Supplemental materials for:
Teacher ratings of the ADHD-RS IV in a community sample: Results from the ChiP-ARD study

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

[*]*Corresponding author*
[1]*Hôpitaux Pédiatriques de Nice CHU Lenval*
57, avenue de la Californie
06200 Nice
FRANCE
Phone: +33 (0)4 92 03 08 49
Fax: +33 (0)4 92 03 05 65
Email: caci.h@pediatrie-chulenval-nice.fr

[2]*Centre for Positive Psychology and Education*
University of Western Sydney
Australia

(1) Mplus syntax for the estimation of the final bifactor model on the total group.

```
!In Mplus code, text placed after an exclamation mark are ignored by the program.
! Next line used to provide a title to the input.
TITLE:  Bifactor (2 S 1 G) CFA of ADHD-RS Teacher Version / All dataset
! Next section used to identify the data file to use in the analyses.
DATA:
FILE IS "ADHDRS_T All.dat";
! Next section to identify the name of all variables in the data set in their order of appearance
!(following NAMES ARE), the variables used in the analyses (following USEVARIABLES
ARE),
! and the variables that are ordered categorical (following CATEGORICAL ARE).
VARIABLE:
NAMES ARE CODE SEXE AGE I1-I23;
USEVARIABLES ARE I1-I18;
CATEGORICAL ARE I1-I18;
! The specifications required for the analyses, including the WLSMV estimator, the THETA
! parameterisation (allowing for the estimation of items' uniquenesses) and an increase in the
number
! of iterations (often useful with WLSMV).
ANALYSIS:
ESTIMATOR IS WLSMV;
PARAMETERIZATION=THETA;
ITERATIONS = 10000;
! to compute the WLSMV chi-square difference test, indicate here the data saved from the
model
! under which this one is nested (see last section of input).
DIFFTEST=BIF3.dat;
! The model itself is defined in the MODEL section. Asociations between items and latent
factors
! (loadings) and marked with BY. Two S factors are defined here (Inatt and Hyper) and one G
factor
! (g). All loadings are estimated (to avoid picking a referent indicator) and thus all latent
variances
! fixed to 1. However, a referent indicator can be picked as long as the one used to identify
the G
! factor is not also used to identify one of the S factors. The factors in a bifactor model are
specified as
! orthogonal and thus, their correlations are fixed to 0 (to ensure that G absorbs the
covariance of the
! items not modelled by the S factors and Vice versa).
  MODEL:
  Inatt BY I1* I3 I5 I7 I9 I11 I13 I15 I17;
  Hyper BY I18* I4 I6 I8 I10 I12 I14 I16 I2;
  g BY I1 * I3 I5 I7 I9 I11 I13 I15 I17
  I18 I4 I6 I8 I10 I12 I14 I16 I2;
  Inatt@1;
  Hyper@1;
  g@1  ;
  g with Inatt-Hyper@0;
```

```
  Inatt with Hyper@0;
```
! Requests for specific sections of outputs
```
OUTPUT:
SAMPSTAT STANDARDIZED RESIDUAL CINTERVAL MODINDICES (3.0);
TECH2 TECH3 TECH4 SVALUES;
```
! To request a file that can be used to compute chi square difference tests associated with this model.
```
SAVEDATA:
DIFFTEST=BIF2.dat;
```

(2) Mplus syntax for the estimation of the configural invariance model across gender.

! Throughout these supplements, we only comment on sections which differ from previous inputs.

```
TITLE:  Configural model/Gender
DATA: FILE IS "ADHDRS_T All.dat";
VARIABLE:
NAMES ARE CODE SEXE AGE I1-I23;
USEVARIABLES ARE I1-I18;
CATEGORICAL ARE I1-I18;
```

! Grouping is used to define the groups, here base on a variables namesd "SEXE" and coded 1 and 2.
! Here, we provide labels to the 1 and 2 values to define the groups as including girls and boys.

```
GROUPING IS SEXE (1=girls 2=boys);
ANALYSIS:
ESTIMATOR IS WLSMV;
PARAMETERIZATION=THETA;
ITERATIONS = 10000;
```

! The first section of the "MODEL" section defines the generic model used in all groups. The next
! section will include statement specific to the second group. Thus, the generic group statement refers
! to the first group also.

```
MODEL:
Inatt BY I1* I3 I5 I7 I9 I11 I13 I15 I17;
Hyper BY I18* I4 I6 I8 I10 I12 I14 I16 I2;
g BY I6* I1 I3 I5 I7 I9 I11 I13 I15 I17
I18 I4  I8 I10 I12 I14 I16 I2;
Inatt@1; Hyper@1; g@1  ;
g with Inatt-Hyper@0;
Inatt with Hyper@0;
```

! Statements in brackets refer to the thresholds of the categorical indicators. There is one less
! threshold than answer category as these reflect the point where the answer changes from one
! category to the other. By default, thresholds are invariant (equal) across group. We need to relax
! this assumption by including these requests for their free estimation. For identification purposes, the
! first threshold of all items is constrained to be invariant and thus not listed here. Similarly, the
! second threshold for one referent indicator per factor is constrained to invariance and thus
! exclamated out here. The referent indicator needs to be different for all factors.

```
![I1$2];
 [I1$3];  [I3$2];  [I3$3];
 [I5$2];  [I5$3]; [I7$2];  [I7$3];
 [I9$2];  [I9$3]; [I11$2];  [I11$3];
 [I13$2];  [I13$3]; [I15$2];  [I15$3];
 [I17$2];  [I17$3];
 ![I18$2];
 [I18$3];  [I4$2];  [I4$3];
```

```
 ![I6$2];
 [I6$3]; [I8$2];  [I8$3];
 [I10$2];  [I10$3]; [I12$2];  [I12$3];
 [I14$2];  [I14$3]; [I16$2];  [I16$3];
 [I2$2];  [I2$3];
! For the second group, we request the free estimation of all parameters that are not fixed in the
! generic section for identification purposes.
! By default, uniquenesses are freely estimated in the second group and fixed to 1 in the first group
! and thus do not need to be specified here.
! By default, the factors means will be freed in the second group.
MODEL boys:
Inatt BY  I1* I3 I5 I7 I9 I11 I13 I15 I17;
Hyper BY I18* I4 I6 I8 I10 I12 I14 I16 I2;
g BY I6* I1 I3 I5 I7 I9 I11 I13 I15 I17
I18 I4  I8 I10 I12 I14 I16 I2;
![I1$2];
 [I1$3];  [I3$2];  [I3$3];
 [I5$2];  [I5$3]; [I7$2];  [I7$3];
 [I9$2];  [I9$3]; [I11$2];  [I11$3];
 [I13$2];  [I13$3]; [I15$2];  [I15$3];
 [I17$2];  [I17$3];
 ![I18$2];
 [I18$3];  [I4$2];  [I4$3];
 ![I6$2];
 [I6$3]; [I8$2];  [I8$3];
 [I10$2];  [I10$3]; [I12$2];  [I12$3];
 [I14$2];  [I14$3]; [I16$2];  [I16$3];
 [I2$2];  [I2$3];
OUTPUT:
SAMPSTAT STANDARDIZED RESIDUAL CINTERVAL MODINDICES (3.0);
TECH2 TECH3 TECH4 SVALUES;
SAVEDATA:
DIFFTEST=M0sex.dat;
```

(3) Mplus syntax for the estimation of the weak (loadings) invariance model across gender.

! We skip TITLE, VARIABLES and OUTPUT sections.

```
ANALYSIS:
ESTIMATOR IS WLSMV;
PARAMETERIZATION=THETA;
ITERATIONS = 10000;
DIFFTEST = M0sex.dat;
MODEL:
Inatt BY I1* I3 I5 I7 I9 I11 I13 I15 I17;
Hyper BY I18* I4 I6 I8 I10 I12 I14 I16 I2;
g BY I6* I1 I3 I5 I7 I9 I11 I13 I15 I17
I18 I4  I8 I10 I12 I14 I16 I2;
Inatt@1; Hyper@1; g@1  ;
g with Inatt-Hyper@0;
Inatt with Hyper@0;
![I1$2];
 [I1$3];  [I3$2];  [I3$3];
 [I5$2];  [I5$3]; [I7$2];  [I7$3];
 [I9$2];  [I9$3]; [I11$2];  [I11$3];
 [I13$2];  [I13$3]; [I15$2];  [I15$3];
 [I17$2];  [I17$3];
 ![I18$2];
 [I18$3];  [I4$2];  [I4$3];
 ![I6$2];
 [I6$3]; [I8$2];  [I8$3];
 [I10$2];  [I10$3]; [I12$2];  [I12$3];
 [I14$2];  [I14$3]; [I16$2];  [I16$3];
 [I2$2];  [I2$3];
```

! By default, the loadigns are specified as invariant cross groups. So to fix them as invariant they only
! need to be taken out of the second group section. This allows for the free estimation of the factor
! variances in the second group.

```
MODEL boys:
! Inatt BY  I1* I3 I5 I7 I9 I11 I13 I15 I17;
! Hyper BY I18* I4 I6 I8 I10 I12 I14 I16 I2;
! g BY I6* I1 I3 I5 I7 I9 I11 I13 I15 I17
! I18 I4  I8 I10 I12 I14 I16 I2;
Inatt*; Hyper*; g*  ;
![I1$2];
 [I1$3];  [I3$2];  [I3$3];
 [I5$2];  [I5$3]; [I7$2];  [I7$3];
 [I9$2];  [I9$3]; [I11$2];  [I11$3];
 [I13$2];  [I13$3]; [I15$2];  [I15$3];
 [I17$2];  [I17$3];
 ![I18$2];
 [I18$3];  [I4$2];  [I4$3];
 ![I6$2];
 [I6$3]; [I8$2];  [I8$3];
```

```
 [I10$2];  [I10$3]; [I12$2];  [I12$3];
 [I14$2];  [I14$3]; [I16$2];  [I16$3];
 [I2$2];  [I2$3];
SAVEDATA:
DIFFTEST=M1sex.dat;
```

(4) Mplus syntax for the estimation of the strong (loadings+ thresholds) invariance model across gender.

```
! We skip TITLE, VARIABLES and OUTPUT sections.
ANALYSIS:
ESTIMATOR IS WLSMV;
PARAMETERIZATION=THETA;
ITERATIONS = 10000;
DIFFTEST = M1sex.dat;
MODEL:
Inatt BY I1* I3 I5 I7 I9 I11 I13 I15 I17;
Hyper BY I18* I4 I6 I8 I10 I12 I14 I16 I2;
g BY I6* I1 I3 I5 I7 I9 I11 I13 I15 I17
I18 I4  I8 I10 I12 I14 I16 I2;
Inatt@1; Hyper@1; g@1  ;
g with Inatt-Hyper@0;
Inatt with Hyper@0;
! As all thresholds are invariant by default, they don't need to be specified at this step.
MODEL boys:
Inatt*;
Hyper*;
g*  ;
SAVEDATA:
DIFFTEST=M2sex.dat;
```

(5) Mplus syntax for the estimation of the strict (loadings+ thresholds + uniquenesses) invariance model across gender.

```
! We skip TITLE, VARIABLES and OUTPUT sections.
ANALYSIS:
ESTIMATOR IS WLSMV;
PARAMETERIZATION=THETA;
ITERATIONS = 10000;
DIFFTEST = M2sex.dat;
MODEL:
Inatt BY I1* I3 I5 I7 I9 I11 I13 I15 I17;
Hyper BY I18* I4 I6 I8 I10 I12 I14 I16 I2;
g BY I6* I1 I3 I5 I7 I9 I11 I13 I15 I17
I18 I4  I8 I10 I12 I14 I16 I2;
Inatt@1; Hyper@1; g@1  ;
g with Inatt-Hyper@0;
Inatt with Hyper@0;
MODEL boys:
Inatt*;
Hyper*;
g*  ;
! With WLSMV, the uniquenesses always need to be fixed to 1 in the first group so that
constraining
! them to invariance involve fixing them to be 1 in both groups.
I1-I18@1;
SAVEDATA:
DIFFTEST=M3sex.dat;
```

(6) Mplus syntax for the estimation of the invariance of the factor variances across gender.

```
! We skip TITLE, VARIABLES and OUTPUT sections.
ANALYSIS:
ESTIMATOR IS WLSMV;
PARAMETERIZATION=THETA;
ITERATIONS = 10000;
DIFFTEST = M3sex.dat;
MODEL:
Inatt BY I1* I3 I5 I7 I9 I11 I13 I15 I17;
Hyper BY I18* I4 I6 I8 I10 I12 I14 I16 I2;
g BY I6* I1 I3 I5 I7 I9 I11 I13 I15 I17
I18 I4  I8 I10 I12 I14 I16 I2;
Inatt@1; Hyper@1; g@1  ;
g with Inatt-Hyper@0;
Inatt with Hyper@0;
MODEL boys:
! Given that all loadings are freely identified (albeit invariant across group) the factors are
! identified by constraining the latent variance to 1 in the first group. Thus, constraining them
! to be equal across group simply involves taking out the request for them being freely
estimated in the
! second group.
! Inatt*;
! Hyper*;
! g*  ;
I1-I18@1;
SAVEDATA:
DIFFTEST=M4sex.dat;
```

(7) Mplus syntax for the estimation of the invariance of the factor means across gender.

```
! We skip TITLE, VARIABLES and OUTPUT sections.
ANALYSIS:
ESTIMATOR IS WLSMV;
PARAMETERIZATION=THETA;
ITERATIONS = 10000;
DIFFTEST = M4sex.dat;
MODEL:
Inatt BY I1* I3 I5 I7 I9 I11 I13 I15 I17;
Hyper BY I18* I4 I6 I8 I10 I12 I14 I16 I2;
g BY I6* I1 I3 I5 I7 I9 I11 I13 I15 I17
I18 I4  I8 I10 I12 I14 I16 I2;
Inatt@1; Hyper@1; g@1  ;
g with Inatt-Hyper@0;
Inatt with Hyper@0;
! By default, the means are constrained to 0 in group 1 and freely estimated in the
other
!  groups. To constrain them to be invariant, they only have to be constrained to zero
in the
! generic model section (applied to all groups).
[Inatt@0];
```

```
[Hyper@0];
[g@0] ;
MODEL boys:
I1-I18@1;
SAVEDATA:
DIFFTEST=M5sex.dat;
```