RESEARCH ARTICLE

Measurement invariance of maternal ratings of ADHD symptoms across clinic referred children's with and without ADHD

Rapson Gomez*   Alasdair Vance[1]   Vasileios Stavropoulos[2]

Abstract: The study examined the measurement invariance (configural, metric, scalar, and error variances) and factor mean scores equivalencies of a modified version of the Strengths and Weaknesses of ADHD-Symptoms and Normal Behavior Scale (SWAN-M) across ratings provided by mothers of clinic-referred children and adolescents, diagnosed with (N = 666) and without (N = 202) ADHD. Confirmatory factor analysis (CFA) of these ratings provided support for the bi-factor model of ADHD [orthogonal general and specific factors for inattention (IA) and hyperactivity/impulsivity (HI) symptoms]. Multiple-group confirmatory factor analysis (CFA) of the bi-factor model supported full measurement invariance. Findings also showed that for latent mean scores, the ADHD group had higher scores than the non-ADHD group for the ADHD general and IA specific factors. The findings indicate that observed scores (based on maternal ratings of the SWAN-M) are comparable, as they have the same measurement properties. The theoretical, psychometric and clinical implications of the findings are discussed.

Keywords: strengths and weaknesses of ADHD-Symptoms and Normal Behavior Scale, factor structure, measurement invariance, children and adolescents, mother ratings

1   Introduction

Since the publication of the 4th edition of the Diagnostic and Statistical Manual of Mental Disorders,[1] a number of rating scales comprising the ADHD symptoms for completion by parents and teachers have been developed. The symptoms proposed for ADHD in the current DSM-5[2] are highly comparable to those in DSM-IV and its text revised edition DSM-IV TR.[3] Thus DSM-IV/DSM-IV TR based ADHD scales can be used to measure the current DSM-5 ADHD symptoms. DSM-5, DSM-IV TR and DSM-IV list the same eighteen ADHD symptoms under two separate groups, namely inattention (IA) and hyperactivity/impulsivity (HI), with nine symptoms for each group. Concurrent to this, most recently used ADHD rating scales have the 18 ADHD symptoms, word-to-word, as presented in DSM-IV, but with the word often omitted in the description of the symp-

toms.[4]

ADHD rating scales are often used in research and clinical settings. In that line, Burns, Walsh, Servera, Lorenzo-Seva, Cardo and Rodrguez-Fornells (2013) have noted that ADHD rating scales have made extensive contributions to our knowledge and understanding of ADHD. Furthermore, ADHD rating scales have been validated and show specificity/sensitivity in identifying individuals with ADHD.[5, 6] Consequently, ADHD ratings have been used for screening ADHD (*e.g.* identifying cases for more comprehensive evaluation for presence of ADHD), identifying individuals with ADHD,[7, 8] facilitating formal diagnosis (e.g. obtaining teacher ratings to establish cross situational consistency of ADHD manifestations), and monitoring treatment (including medication) effects.[9] Given the extensive and diverse clinical use of ADHD rating scales, it is important that there is appropriate psychometric data supporting the ways they are used.

As already mentioned, among other uses, ADHD rating scales have been used clinically to facilitate the detection of children who could potentially have ADHD, and for separating children into groups of those with and without ADHD.[9] Creditable use for this purpose requires that there is measurement invariance confirmed for the ADHD symptom ratings across these groups. Provided the dearth of such findings, the present study examined the measurement invariance across those with

*   **Correspondence to**: Rapson Gomez, School of Health and Life Sciences, Federation University Australia, Australia;   Email: rapson.gomez@federation.edu.au

[1] Royal Childrens Hospital and The University of Melbourne, Victoria, Australia

[2] Cairnmillar Institute, Australia

and without an ADHD diagnosis, based on mother-ratings of their clinic-referred children's and adolescent's ADHD symptoms. These were examined using a slightly modified version of the Strengths and Weaknesses of ADHD-Symptoms and Normal Behavior Scale.[10] The SWAN is an ADHD rating scale, corresponding to DSM-IV ADHD symptoms, and is accordingly reflective of the relevant DSM-5 criteria.

Interestingly, confirmatory factor analysis (CFA) of the ratings of ADHD symptoms in many different versions of ADHD rating scales of community and clinic-referred children and adults have generally provided support for the theorized bi-factor model, with separate factors for the IA and HI symptoms.[11,12] However, many recent CFA studies of ADHD rating scales have shown more support for a bi-factor model,[13,14] across informants (parent, teacher, self), methods (questionnaires, interviews), participants age groups (preschool, school-aged, adolescents, adults) and participants cultural background.[15] A bi-factor model is an orthogonal first-order factor model with a general factor and specific or group factors for different dimensions in the model (see Figure 1). In such a model, the general factor explains the covariance across all the items, and the specific factors explain the unique covariance of the items within the relevant dimensions, after accounting for the general factor.[16] Thus, the ADHD bi-factor model (see Figure 1) has an ADHD general factor on which all the IA and HI symptoms load, and separate orthogonally related specific factors for the IA and HI symptom groups, after removing the variances captured by the general factor. It is notable that past studies have reported that much of the reliable variance for ADHD is captured by the general factor, with very low variances remaining to be explained by the specific factors.[17]
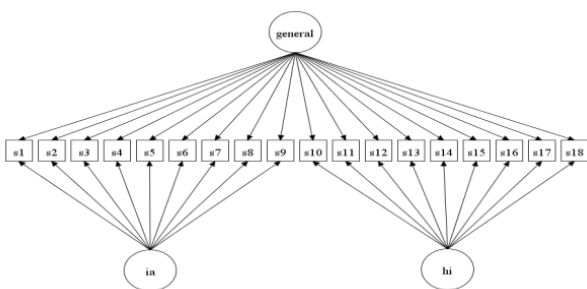

Figure 1. Schematic diagram of bi-factor ADHD models

Regardless of what is the best factor model for the ADHD symptoms, it is critical for the accurate use of the ADHD ratings to distinguish those with and without ADHD diagnosis, that there is measurement invariance for these ratings across these groups. In general, mea-surement invariance across groups deals with whether the observed scores on a measure are the same across the groups when these scores "represent" the same level (intensity, severity) of the underlying latent trait score.[18,19] Lack of support for invariance indicates that the scores obtained by the groups cannot be accurately compared on the measure used, since any difference could be confounded by discrepancies in the scaling properties of the measure for the groups.[20] When applied to ADHD rating scales, measurement invariance across groups of children with and without ADHD refers to observed ADHD scores being the same across these groups, when individuals in the groups have the same level of the underlying ADHD latent trait.[21] Lack of support for measurement invariance means that the ADHD scores obtained by these two groups cannot be accurately compared, as their differences could be explained by variations in the psychometric properties of the measures across the groups. Expressed differently, the same observed scores for the two groups may not reflect the same level of the underlying ADHD construct. Thus, lack of support for invariance across the groups for ratings of ADHD symptoms could seriously question the current practice of deriving groups of children with and without ADHD, using ADHD rating scales.[22] If ADHD symptom ratings for these groups are to be compared, then measurement invariance for them needs to be confirmed in the first instance.[23] Additionally, in relation to research, lack of measurement invariance would raise questions about the validity of existing ADHD findings from studies based on administration of ADHD rating scales across children with and without an ADHD diagnosis.

A powerful method for examining measurement invariance is the multiple-group mean and covariance structures CFA approach. Assuming that the indicator-ratings are treated as continuous scores, this approach can test for configural invariance (same overall factor structure), item factor loadings invariance (same strength of the associations of items with the first-order factors), item intercepts invariance (equivalency in item intercepts values), and error variances or uniqueness invariance (equivalency in the error variances of the items or variances of the items not attributed to the underlying constructs). When there is support for invariance for item factor loadings and intercepts, the groups can be also compared for structural invariance (equivalencies for variances and covariances), and differences in their latent factor mean scores.[24,25]

There are reasons to suspect that ADHD symptoms may lack full measurement invariance across maternal ratings of children with and without ADHD. Specifi-

cally, available studies have shown that ADHD ratings are vulnerable to the "halo" effect.[26–29] The "halo" effect occurs when a person, who is displaying one discrete behavior is rated as exhibiting other behaviors, even when those behaviors are not observed. The "halo" effect could be unidirectional (presence of primary symptom leads to a secondary symptom being falsely endorsed) or bidirectional (presence of a primary symptom inflates ratings of a secondary symptoms and the presence of the secondary symptom also inflates ratings of the primary symptom). Irrespective of these, the outcome of "halo" effect is that it may lead to a "self-fulfilling prophecy process", that could in term contribute to the excessive overestimation (biases and distortions) of one or multiple symptoms. In relation to ADHD rating scales in particular, "halo" effect has been revealed for teacher, college student, and parent ratings of ADHD symptoms. Specifically, the DeVries, *et al.* (2017) study, involving parent ratings, identified the IA symptoms of "Difficulty sustaining attention", and "Doesnt seem to listen", as particularly prone to "halo" effect. In terms of proneness to "halo" effect responses, it could be assumed that ones knowledge and experience of the set of symptoms being rated could influence his/her proneness to engage in responses characterized and/or influenced by "halo" effect. Following that line of thought, an Australian-based study has shown that although the core features (symptoms) of ADHD are well-known in the community, there are misconceptions and discrepancies about many ADHD aspects, especially between individuals who have had contact with ADHD in their own families and those who have had no such exposure.[30] There is also evidence of high co-occurrence of ADHD in parents and children.[31, 32] Thus, it could be speculated that compared to parents of children without ADHD, parents of ADHD children could be more knowledgeable about and experienced with ADHD, and consequently more prone/sensitive to "halo" effect responses, or a distorted estimation of ADHD symptoms in their off-springs. Expressed differently, "halo" effects could result in maternal ratings of the ADHD symptom that may differ from the actual trait level, confounded by their level of experience of the diagnosed behaviors. Viewed in terms of measurement invariance, this could be reflected by varying intercepts for the ADHD symptoms, or lack of scalar invariance.

Given the lack of measurement invariance data for the ADHD symptoms across those with and without the ADHD diagnose, and the possibility that these symptoms could lack measurement invariance across these groups, the major goal of this study was to used multiple-group CFA to examine measurement invariance for the ADHD symptoms across those with and without the ADHD diagnose. The study used successive CFA models to examine measurement invariance (configural, metric, scalar and uniqueness) across mother ratings of clinic-referred children and adolescents (referred henceforth as youth) with ADHD and without ADHD. Ratings were obtained using a modified version of the SWAN,[33] a widely used ADHD rating scale.[34] As there is most support for the ADHD bi-factor model, the study examined measurement invariance across the groups for the ADHD symptoms, based on this model. Given the possible "halo" effects for the ADHD symptoms reported by DeVries, *et al.*, (2017), lack of scalar invariance was expected particularly so for the IA symptoms "Difficulty sustaining attention", and "Doesnt seem to listen". It is to be noted that this is the first study to evaluated measurement invariances for the ADHD symptoms across youths with and without ADHD. Thus the findings from the study would be novel and add importantly to existing measurement invariance data for the ADHD symptoms, and thereby contribute to research and clinical practice in ADHD.

2 Methods

2.1 Participants

The current study used archival data collected at the Academic Child Psychiatry Unit (ACPU) of the Royal Children's Hospital (RCH), Melbourne, Australia. The ACPU is an out-patient psychiatric unit that provides services for children and adolescents with behavioral, emotional, and learning problems. For the present study, records of children and adolescents, aged between 7 and 17 years, referred between 2008 and 2016, who were assessed with the modified SWAN (SWAN-M) were used. In all, there were 868 children and adolescents. These individuals were divided into those with a diagnosis of ADHD (N = 666) and those without a diagnosis of ADHD (N = 202). The ADHD group included those with Combined type (N = 422), Inattentive type (N = 187) and Hyperactive/Impulsivity type (N = 57). The ACPU diagnostic procedure for all disorders is described in the "Procedure" subsection of the present manuscript.

Table 1, 2 presents some background and demographic information of the two groups, including age, gender, mother and father employment status and highest education levels completed, family income, parental relationship status. The results of the comparisons between those with and without ADHD for the background and demographic information together with effect sizes for these comparisons are also presented. We have not included information on race/ethnicity as this data was not

recorded in the archival data collected at the ACPU.

As shown in Table 1, 2, the mean (SD) for ages for the ADHD group and non-ADHD groups were 11.22 (3.34) years and 10.54 (3.11) years, respectively. The non-ADHD group was significantly older, and comprised relatively more females than males. The Cohen's d values for differences considering age and gender composition of the groups were small. The scores' comparisons for mother and father employment and education, family income and parental relationship status showed no significant group differences. Thus, on the whole, the groups were reasonably well matched for all background and demographic variables examined, except for small differences for age and gender.

Table 1, 2 also includes the frequencies and percentages of different groups of disorders in the groups with and without ADHD diagnosis. In the table, the label "any anxiety disorder" includes Separation Anxiety, Social Phobia, Specific Phobia, Panic, Agoraphobia, Generalized Anxiety, Obsessive Compulsive and/or Post-Traumatic Stress Disorders, while the label "depression disorders" includes those with Dysthymic and/or Major Depressive Disorders. In terms of clinical diagnoses, based on DSM-IV TR, 41.1% reached criteria for depression disorders, 74.9% had at least one or more anxiety disorders, and 31.1% presented with either ODD or CD. Although, there were significantly more individuals with ODD/CD in the ADHD group, the Cramer's V value for this difference was small, while there was no significant statistical difference in the number of individuals with depression disorders or anxiety disorders across the two groups. Nevertheless, there was high comorbidity, with 81.00% of the participants being diagnosed with two or more disorders.

2.2 Measures

First, we searched for the keyword "Google Trends" in the "Abstract-Title-Keywords" field for the journal articles. The first two articles using Google Trends were begun in 2009. The search returned 96 publications.

Strengths and Weaknesses of ADHD-Symptoms and Normal Behavior Scale.[35] The SWAN lists the 18 DSM-IV symptoms for ADHD. It is noted that although the instrument is developed with DSM-IV ADHD symptoms in mind, these 18 symptoms are the same in DSM-5. Unlike the way the symptoms are worded in DSM-IV or DSM-5, and also other ADHD rating scales, the SWAN has the ADHD symptoms reworded such that they reflect strengths rather than weaknesses. For example, the DSM-IV symptom "Often avoids, dislikes, or reluctantly engages in tasks requiring sustained mental effort is reworded as "Engage in tasks that require

sustained mental effort." In this study, the SWAN was completed by the mothers of the children. Respondents rated the occurrence of each symptom over the past 6-months on a 5-point scale, ranging from "far below average" (scored 1) to "far above average" (scored 5), relative to other children of the same age. Although the original SWAN has a reference period of one month, a six months reference period was used here for reasons of clinical utility/relevance within the context of ACPU (six months is the ADHD symptoms reference period in DSM-IV/DSM-IV TR). Furthermore, although the original version of the SWAN involved a 7-point scale, initial piloting of the 7-point scale version of the SWAN in the ACPU indicated virtually no endorsement of levels -1 (slightly below average) and +1 (slightly above average). Thus, it was advocated to collapse/merge levels -1 and -2 of the original scale into a single category (-1; below average), and levels +1 and +2 into another single category (+1; above average), thereby resulting to the 5-point scale used in the current study (-1 = far below; -2= below; 3 = average; +1 = above, +2 = far above). Additionally, to ease interpretation of the findings, all symptoms were recoded so that higher scores reflected higher symptoms (-1 = far below average (-1) recoded 5; below average (-2) recoded 4, average (3) recoded as 3; below average (-2) recoded 4, and far below average (-1) recoded 5. Considering the present data, the internal consistency coefficient alpha values were 0.89, 0.89. 0.92 for the IA and HI and combined ADHD (IA plus HI) factors, respectively.

Anxiety Disorders Interview Schedule for Children. The ADISC-IV is a semi-structured interview, based on the DSM-IV-TR diagnostic system. The diagnoses reported earlier (see sample description) were derived from this schedule. Although ADISC-IV has been designed primarily to facilitate the diagnosis of the major childhood internalizing disorders, it can also be used for diagnosing the major childhood externalizing disorders. In that context, there is support for the concurrent validity of the ADISC-IV ADHD module in the ADHD groups [based on parent interviews, and parent and child interviews using the ADISC-IV, ADHD groups (IA, HI) did not differ from one another, and showed greater externalizing and attention problems than a no ADHD group].[35] The ADISC-IV diagnostic guideline instructs that the child be given diagnosis of all disorders meeting the relevant criteria. The scores of ADISC-IV have been shown to present sound psychometric properties,[36] excellent test-retest reliability over a 7 to 14-day interval and Kappa (an index of inter-rater agreement) values for interviews with parents ranging from 0.65 to 1.00. However, it should be highlighted that there are differ-

Table 1. Demographics Information for Participants in the Without ADHD and With ADHD Groups

Variable/Categories	Without ADHD	With ADHD	Test Statistic	Effect size
Number	202	666		
Age – Mean (*SD*)	11.22 (3.34)	10.54 (3.11)	*t(df = 866) = 2.66**	0.22
Gender [N, (percentage)]				
Boy	125 (14.4)	509 (58.6)	χ^2 (*df*= 1) = 16.65	0.14
Girl	77 (8.9)	157 (18.1)		
Mother employment [N, (percentage)]				
Employed	110 (13.0%)	304 (36.0%)	χ^2 (*df*= 6) = 11.17	0.115
Home duties	59 (7.0%)	238 (28.2%)		
Pensioner	9 (1.1%)	39 (4.6%)		
Unemployed	12 (1.4%)	21 (2.5%)		
Student	5 (0.6%)	19 (2.2%)		
Retired	0 (0.0%)	4 (0.5%)		
Other	3 (0.4%)	22 (2.6%)		
Total	198 (23.4%)	647 (76.6%)		
Mother education [N, (percentage)]				
No schooling	0 (0.0%)	1 (0.1%)	χ^2 (*df*= 6) = 5.95	0.084
Some primary school	1 (0.1%)	1 (0.1%)		
Primary school	1 (0.1%)	2 (0.2%)		
Some secondary school	51 (6.1%)	218 (26.0%)		
Secondary school	30 (3.6%)	83 (9.9%)		
Technical, trade	47 (5.6%)	152 (18.1%)		
Tertiary	66 (7.9%)	186 (22.2%)		
Total	196 (23.4%)	643 (76.6%)		
Father employment [N, (percentage)]				
Employed	145 (18.8%)	462 (59.8%)	χ^2 (*df*= 6) = 10.25	0.117
Home duties	7 (0.9%)	8 (1.0%)		
Pensioner	4 (0.5%)	33 (4.3%)		
Unemployed	19 (2.5%)	53 (6.9%)		
Student	2 (0.3%)	2 (0.3%)		
Retired	1 (0.1%)	1 (0.1%)		
Other	7 (0.9%)	28 (3.6%)		
Total	185(24.0%)	587(76.0%)		
Father education [N, (percentage)]				
No schooling	0 (0.0%)	1 (0.1%)	χ^2 (*df*= 6) = 7.02	0.096
Some primary school	2 (0.3%)	2 (0.3%)		
Primary school	1 (0.1%)	13 (1.7%)		
Some secondary school	61 (8.0%)	221 (28.9%)		
Secondary school	21 (2.7%)	67 (8.7%)		
Technical, trade	47 (6.1%)	145 (18.9%)		
Tertiary	53 (6.9%)	132 (17.2%)		
Total	185 (24.2%)	581 (75.8%)		

Note: *p<0.05; ***p<0.001

Table 2. Demographics Information for Participants in the Without ADHD and With ADHD Groups

Variable/Categories	Without ADHD	With ADHD	Test Statistic	Effect size
Number	202	666		
Age – Mean (*SD*)	11.22 (3.34)	10.54 (3.11)	*t(df = 866) = 2.66**	0.22
Gender [N, (percentage)]				
Boy	125 (14.4)	509 (58.6)	χ^2 (*df*= 1) = 16.65	0.14
Girl	77 (8.9)	157 (18.1)		
death of both parents	0 (0.0%)	1 (0.1%)		
other, please describe	4 (0.5%)	19 (2.2%)		
Total	198 (23.3%)	652 (76.7%)		
Family income				
$0	0 (0.0%)	2 (0.2%)	χ^2 (*df*= 4) = 3.03	0.061
$0-$30,000	53 (6.4%)	205 (24.9%)		
$30,000 - $40,000	21 (2.6%)	69 (8.4%)		
$40,000 - $50,000	17 (2.1%)	50 (6.1%)		
$50,000 and over	104 (12.7%)	301 (36.6%)		
Total	195 (23.7%)	627 (76.3%)		
Presence of any anxiety disorder [N,(percentage)]				
Without [218 (25.1%)]	50 (5.8%)	168 (19.4%)	χ^2 (*df*= 1) = 0.02	-0.005
With [650 (74.9%)]	152 (17.5%)	498 (57.4%)		
Presence of any depressive disorder [N, (percentage)]				
Without [511 (58.9%)]	114 (13.1%)	397 (45.7%)	χ^2 (*df*= 1) = 0.65	-0.027
With [357 (41.1%)]	88 (10.1%)	269 (31.0%)		
Presence of Oppositional Defiant Disorder/Conduct Disorder (ODD/CD) [N, (percentage)]				
Without [270 (31.1%)]	102 (11.8%)	168 (19.4%)	χ^2 (*df*= 1) =46.18***	0.231
With [598 (68.9%)]	100 (11.5%)	498 (57.4%)		

Note: *p<0.05;   ***p<0.001

ent ADISC-IV versions for parent interview and for child interview, and clinical diagnosis can be based either on parent or child interview or on both interviews considered together.[37]   All diagnoses reported in this study were based on parent interviews as there is evidence of poor levels of agreement for diagnosis between information across the child and parent versions of the ADISC-IV,[38] alongside with evidence that clinical interviews of children can lead to unreliable diagnosis.[39] Finally, in the present study, ADISC-IV interview data could not be considered for establishing measurement invariance of the ADHD symptoms, due to lack of relevant symptom level information in the ACPU archival data used.

2.3   Procedure

The study was approved by the RCH ethics committee as part of ACPU's comprehensive examination of children and adolescent referred for psychological problems.

Each legal guardian and participant provided informed written consent for any data provided by them to be used in future research studies. This is a standard part of the ACPU assessment procedure.

All participants and their parents participated in separate interviews and testing sessions which were held over two days during the admission of the child. Breaks were provided as necessary. In all cases, parental consent forms were completed prior to the assessment. The parent and child comprehensive data collected covered demographic, medical (primarily neurological and endocrinological), child educational (including standardized measures of IQ and academic achievement tests of reading, arithmetic and language), child psychological (standardized measures of behavioral and emotional self-rating and parent rating scales, diagnostic interviews using the child and parent versions of the ADISC-IV, and neuropsychological measures), family related (standardized measures of family maladjustment, and mari-

tal satisfaction), and maternal mental health (standardized measures of behavioral and emotional symptoms) aspects. Information was also obtained from teachers using various checklists and questionnaires, such as the Teacher Report Form[40] and the Conners 3-Teacher.[41] However, for the current study only the information for the ADISC-IV from parents and parent completed SWAN-M ratings were used.

All psychological data were collected by (specially trained) research assistants, who were advanced masters or doctoral students in clinical psychology, and under the supervision of two registered clinical psychologists. Prior to data collection, the research assistants were provided with extensive supervised training and practice by the two ACPU employed registered psychologists. The training for the ADISC-IV-P included observations of the interview process being administered by the psychologists. The research assistants commenced administering the ADISC-IV only after they attained competence in its administration, as assessed by their supervisors. At this point it should be noted that there was adequate inter-rater reliability for the diagnoses made between the research assistants and the supervisors, and between the research assistants themselves (kappa = 0.88). Standardized procedures were applied for the administration of all measures. Where necessary (due to English literacy reasons), researchers read the items to the participants (approximately 5% of the sample). Approximately 95% of the parent ADISC-IV interviews involved mothers only, and the rest involved fathers only or both fathers and mothers together. Using the categorical data from the parent ADISC-IV, clinical diagnosis was determined by two consultant child and adolescent psychiatrists, who independently reviewed these data. The inter-rater reliability for diagnoses of the two psychiatrists was high (kappa = 0.90). As noted earlier, for the current study, only the records of children and adolescents which involved scores for the SWAN-M, rated by mothers were used.

2.4 Statistical procedures

All CFA models in the study were conducted using the robust maximum likelihood (MLR) estimation in the Mplus, software Version 7.[42] Given that there were five response options, the use of MLR-based extraction is appropriate,[43, 44] and can correct for potential deviations from normality in the data set. At the statistical level, model fit was examined using robust maximum likelihood (MLR) χ^2 values. However, as χ^2 values, including MLR χ^2 values, are inflated by large sample sizes, the fit of the models was also examined using the approximate fit indices of the root mean squared error

of approximation (RMSEA), the comparative fit index (CFI), and the Tucker-Lewis Index (TLI). According to the guidelines suggested by Hu and Bentler (1998), RMSEA values close to 0.06 or below can be considered as good fit, 0.07 to < 0.08 as moderate fit, 0.08 to 0.10 as marginal fit, and < 0.10 as poor fit. For the CFI and TLI, values of close to 0.95 or above are taken as indicating good model-data fit, and values of 0.90 and < 0.95 are taken as acceptable fit. Differences between nested models were computed using the difference in MLRχ^2 values (computed using the scaling correction formula for MLR;[45] An alpha value of 0.01 was used to allow for more stringent Type II error control in the models compared.

Measurement invariance across the ADHD and non-ADHD groups for the bi-factor model was tested using the multiple-group CFA invariance procedure proposed by in the literature. Specifically, this study tested in sequence, configural, metric, scalar and error variances invariance (equality for items uniqueness) across the groups. Metric, scalar and uniqueness invariance are alternatively referred as weak, strong and strict invariance. Due to space limitations, details of the procedure used are not provided. Readers are referred to Brown (2006) for details, including the steps for testing partial invariance. When there is some support for measurement invariance (full or partial), the groups can be compared for latent mean scores, taking into account the non-invariance in the measurement model. For the current study, the non-ADHD group served as the reference group.

3 Results

3.1 Missing Values

There were no missing values in the data set used.

3.2 Fit for the Bi-factor ADHD Models for the ADHD and non-ADHD Groups

Prior to the test for measurement invariance, the fit of the bi-factor ADHD models in the two groups were examined. The findings for the non-ADHD group indicated close to good fit in terms of the RMSEA value, marginally acceptable fit in terms of the CFI value, and poor fit in terms of the TLI value, χ^2 (df = 117) = 218.63, p < 0.001; RMSEA = 0.066 (90% CI = 0.052 - 0.079); CFI = 0.909 and TLI = 0.881. For the ADHD group, the CFI, TLI and RMSEA values indicated good fit, χ^2 (df = 117) = 267.89, p < 0.001; RMSEA = 0.0446 (90% CI = 0.037 - 0.0519); CFI = 0.962 and TLI = 0.950. These findings can be interpreted as indicating reasonable level

Table 3. Invariance for the ADHD Symptoms across Children With and Without ADHD Diagnosis

Models (M)	MLRχ^2	df	RMSEA (90% CI)	CFI	TLI	ΔM	Δdf	ΔMLRχ^2
M1: Configural invariance	486.905	234	0.050 (0.0443 -0.056)	0.95	0.935	-	-	-
M2: Metric invariance	541.27	270	0.048 (0.042 - 0.054)	0.946	0.939	M2 – M1	36	56.52
M3: Thresholds invariance	567.77	285	0.048 (0.042 -0.054)	0.944	0.94	M3 – M2	15	26.08
M4: Error variances invariance	586.4	303	0.047 (0.041 -0 .052)	0.944	0.943	M4– M3	18	22
M5: Invariance for latent factor mean	794.57	306	0.061 (0.056 - 0.066)	0.903	0.903	M5 – M4	3	258.03***
M5.1: Invariance for ADHD latent factor mean	644.72	305	0.051 (0.045 - 0.056)	0.933	0.932	M5.1– M4	1	91.05***
M5.2: Invariance for IA latent factor mean	587.91	304	0.047 (0.041 - 0.052)	0.944	0.943	M5.2– M4	2	1.43

of fit for the the bi-factor ADHD models for both the groups.

3.3 Measurement Invariance for the Bi-factor Model Across the ADHD- and ADHD+ Groups

For the bi-factor model, the fit indices for the baseline or configural invariance model (M1 in Table 3) were χ^2 (df = 234) = 486.90, p < 0.001; RMSEA = 0.050; CFI = 0.950 and TLI = 0.935. Thus, the CFI, and RMSEA values indicated good fit, and the TLI indicted acceptable fit. Overall, there was adequate support for the configural invariance. There was no difference between the configural invariance model (M1 in Table 3) and the metric invariance model (M2 in Table 3); Δdf = 36; Δ MLMχ^2 = 56.52, p < 0.01); the metric invariance model and the scalar invariance model (M3 in Table 3); Δdf = 15; Δ MLMχ^2 = 26.08, ns), and the scalar invariance model and the error variances invariance model (M4 in Table 3); Δdf = 18; Δ MLMχ^2 = 22.00, ns). Thus there was support for full measurement invariance for ratings of the ADHD symptoms across the ADHD and non-ADHD groups.

Given support for invariance for the measurement model, further analysis was conducted for equivalency in latent mean scores. For this analysis, the reference group was the non-ADHD group (thus their latent scores were fixed to 0), and the focus group was the ADHD group (thus their latent scores were freely estimated). As indicated in Table 3, for the criteria used here (p < 0.01), the groups differed for the general ADHD latent factor, and IA specific factor. The standardized mean score (SE) for ADHD and IA factors for the ADHD group were 1.182 (0.091) and 0.872 (0.110), respectively. The positive value suggests that the ADHD group had higher scores for ADHD. The standardized mean score difference (same as unstandardized mean score in Mplus) can be interpreted as akin to Cohen's (1992) d effect sizes. Considering d effect sizes differences, Cohen's recommended magnitudes are as follows: < 0.20 = negligible; $\geqslant$ 0.20 and < 0.50 = small; $\geqslant$ 0.50 and < 0.80 =

medium; $\geqslant$ 0.80 = large. This means the magnitude of the differences between ADHD and non-ADHD for the ADHD general and the IA specific factors were both of large effect size.

4 Discussion

The major aim of the study was to use multiple group CFA to examine measurement invariance across maternal ratings of the SWAN-M (47) for those with and without ADHD diagnosis. Initial analyses indicated reasonable level of support for the bi-factor models across the two groups. Consistent with these findings, existing data also show support for the bi-factor model. The findings for the multiple-group CFA analyses showed acceptable fit for the configural model. Furthermore, there was no difference between the configural model and the metric invariance model, the scalar invariance model and the metric invariance model and the error variances model and the scalar invariance model. Thus the findings for SWAN-M indicated support for the configural model (same pattern of factor structure), and for full measurement invariance for the metric (same factor loadings), scalar (same response categories), and error variances (same unique variances) models, respectively, for ratings from clinic-referred children and adolescents with and without an ADHD diagnosis. Additionally, the findings revealed that the ADHD group had higher scores, with large effect sizes, for the ADHD general factor and the IA specific factor.

To date, no previous study had assessed measurement invariance for the ADHD symptoms across those with and without an ADHD diagnosis. Thus, the findings in the current study can be seen as providing novel measurement invariance data for ADHD symptoms. Thus the findings from the study add importantly to existing measurement invariance data for the ADHD symptoms, and thereby contribute to research and clinical practice in ADHD.

The findings have important conceptual, theoretical and clinical implications. At the conceptual and theoretical level, it was speculated that some of the ADHD

symptoms would lack scalar invariance. This hypothesis was based on existing data showing that due to parents of ADHD children having greater knowledge and experience with ADHD symptoms, their ratings of ADHD behaviors of their ADHD diagnosed children would be more prone to distorting "halo" effects, than the parent ratings referring to children without an ADHD diagnosis. Consequently, these parents were assumed to potentially provide subjectively exacerbated scores when rating their children. However, the support for full scalar invariance revealed across the two groups studied, indicate that the greater knowledge and experience of ADHD symptoms, that likely characterize parents of ADHD diagnosed children, do not necessarily lead to "halo" effects distorting their ADHD SWAN-M scores. In relation to clinical implications, the support for measurement invariance signifies that observed scores of maternal ratings of children with and without ADHD considering ADHD symptoms can be compared directly at least as measured by mother ratings on the modified SWAN for a clinical population. Demonstration of measurement invariance, especially scalar invariance is an essential requirement if maternal ratings on ADHD scales are to be considered for across group comparisons in clinical and research settings to evaluate correlates of ADHD. In this respect, it is worth stressing that, as the bi-factor model was supported, it could be more appropriate to use the total score for the ADHD scale (as it related to the general factor), than the separate IA and HI scores, for group comparisons.

Although the current study has provided useful new information about the measurement invariance of ratings of the ADHD symptom based on the SWAN-M, the findings and their interpretations embrace certain limitations. First, it is possible that factors such as age, gender, ethnicity, comorbidity, and maternal psychopathology could influence ratings of ADHD symptoms.[46] The failure to control for these effects in this study could have confounded the results. Second, because measurement invariance was examined specifically for clinic referred children using maternal ratings of the SWAN-M,[47] the findings here could be unique to clinical referred groups, to the form of the SWAN-M used, and/or to maternal ratings. Third, all the participants in this study were from the same clinic, and therefore, they did not constitute a random sample. Thus, it is likely that this may introduce a bias for the sample examined, limiting the generalizability of the findings and the conclusions made in this study. At a practical level, however, it is difficult and virtually impossible to obtain random samples involving clinical samples. Fourth, as it is possible that as the current sample was heterogeneous in terms of psy-

chopathology, the findings may have been additionally confounded. Finally, the use of archival data is interwoven with the the typical limitations of using archival data.[48] In view of these limitations some may wish to consider the findings and interpretations made in this study as tentative. Therefore, it could be useful if future studies took into consideration the limitations illustrated in the present study.

5 Authors' contributions

RG conceived and designed the study, performed the statistical analyses, and contributed to the writing up of all sections of the paper. AC was in charge of collecting the data, and contributed to the writing up of the paper. VS checked the statistical analyses, and contributed to the writing up of the paper.

References

[1] American Psychiatric Association.Diagnostic and Statistical Manual of Mental Disorders (4th ed.). Washington, 1994.

[2] American Psychiatric Association. Diagnostic and statistical manual of mental disorders DSM-IV TR (Text Revision). Washington, 2000.

[3] American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th ed.). Washington, 2013.
https://doi.org/10.1176/appi.books.9780890425596

[4] Gomez R and Stavropoulos V. Parent Ratings of the Strengths and Difficulties Questionnaire: What Is the Optimum Factor Model? Assessment, 2017.
https://doi.org/1073191117721743.

[5] Daley D. Attention deficit hyperactivity disorder: A review of the essential facts. Child Care, Health and Development, 2006, **32**: 193-204.
https://doi.org/10.1111/j.1365-2214.2006.00572.x

[6] Power TJ, Doherty BJ, Panichelli-Mindel SM, *et al*. The predictive validity of parent and teacher reports of ADHD symptoms. Journal of Psychopathology and Behavioral Assessment, 1998, **20**: 5781.
https://doi.org/10.1023/A:1023035426642

[7] Gaub M and Carlson CL. Behavioral characteristics of DSM-IV ADHD subtypes in a school-based population. Journal of Abnormal Child Psychology, 1997, **25**: 103-111.
https://doi.org/10.1023/A:1025775311259

[8] Teegarden LA and Burns GL. 12-month stability of ADHD, opposition defiant disorder, and conduct disorder symptoms in kindergarten through fifth grade children based on a single information source: Usefulness of teacher ratings for the creation and study of the ADHD subtypes. Child & family behavior therapy, 1999, **21**(3): 53-70.
https://doi.org/10.1300/J019v21n03_04

[9] Aman MG and Pearson DA. Monitoring, measuring drug effects. II. Behavioral, emotional, and cognitive effects. In: J. S. Werry & Aman, M. G. (eds). Practitioner's Guide to Psychoactive Drugs for Children and Adolescents. 1999, 99-164. New York: Plenum Press.
https://doi.org/10.1007/978-1-4899-0086-9_5

[10] Gomez R, Vance A and Stavropoulos V. Test-Retest Measurement Invariance of Clinic Referred Childrens ADHD Symptoms. Journal of Psychopathology and Behavioral Assessment, 2017, 1-12.
https://doi.org/10.1007/s10862-017-9636-4

[11] Swanson J, Schuck S, Mann M, *et al*. Categorical and dimensional definitions and evaluations of symptoms of ADHD: the SNAP and SWAN rating scales. University of California, Irvine, 2006.

[12] Gomez R, Burns GL, Walsh JA, *et al*. Trait, source and error variance in ADHD symptoms: A multitrait-multisource analysis with Australian and Brazilian children. Psychological Assessment, 2003, **15**: 3-16.
https://doi.org/10.1037/1040-3590.15.1.3

[13] Gomez R, Harvey J, Quick C, *et al*. 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, 1999, **40**: 265-274.
https://doi.org/10.1111/1469-7610.00440

[14] Arias VB, Ponce FP and Nunez DE. Bifactor models of Attention-Deficit/Hyperactivity Disorder (ADHD): An evaluation of three necessary but underused psychometric indexes. Assessment, 2016(b).
https://doi.org/10.1177/1073191116679260

[15] Gomez R, Vance A and Gomez RM. Validity of the ADHD bifactor model in general community samples of adolescents and adults, and a clinic-referred sample of children and adolescents. Journal of Attention Disorders, 2013.
https://doi.org/10.1177/1087054713480034

[16] Burns GL, Walsh JA, Servera M, *et al*. Construct validity of ADHD/ODD rating scales: Recommendations for the evaluation of forthcoming DSM-V ADHD/ODD scales. Journal of Abnormal Child Psychology, 2013, **41**: 15-26.
https://doi.org/10.1007/s10802-012-9660-5

[17] Reise SP. The rediscovery of bifactor measurement models. Multivariate behavioral research, 2012, **47**(5): 667-696.
https://doi.org/10.1080/00273171.2012.715555

[18] Arias VB, Ponce FP, Martnez-Molina A, *et al*. General and specific attention-deficit/hyperactivity disorder factors of children 4 to 6 years of age: An exploratory structural equation modeling approach to assessing symptom multidimensionality. Journal of Abnormal Psychology, 2016(a), **125**: 125-137.
https://doi.org/10.1037/abn0000115

[19] Cohen J. Statistical power analysis. Current directions in psychological science, 1992, **1**(3): 98-101.
https://doi.org/10.1111/1467-8721.ep10768783

[20] Drasgow F and Kanfer R. Equivalence of psychological measurement in heterogeneous populations. Journal of Applied Psychology, 1985, **70**: 662-680.
https://doi.org/10.1037/0021-9010.70.4.662

[21] Pontes HM, Stavropoulos V and Griffiths MD. Measurement Invariance of the Internet Gaming Disorder ScaleShort-Form (IGDS9-SF) between the United States of America, India and the United Kingdom. Psychiatry research, 2017, **257**: 472-478.
https://doi.org/10.1016/j.psychres.2017.08.013

[22] Hu LT and Bentler PM. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychological Methods, 1998, **3**: 424-453.
https://doi.org/10.1037/1082-989X.3.4.424

[23] Meredith W. Measurement invariance, factor analysis, and factorial invariance. Psychometrika, 1993, **58**: 525-542.
https://doi.org/10.1007/BF02294825

[24] Reise SP, Bonifay WE and Haviland MG. Scoring and modeling psychological measures in the presence of multidimensionality. Journal of personality assessment, 2013 **95**(2): 129-140.
https://doi.org/10.1080/00223891.2012.725437

[25] DeVries LN, Hartung CM and Golden TL. Negative Halo Effects in Parent Ratings of ADHD and ODD. Journal of Psychopathology and Behavioral Assessment, 2017, **39**(2): 179-188.
https://doi.org/10.1007/s10862-016-9560-z

[26] Hartung CM, Lefler EK, Tempel AB, *et al*. Halo effects in ratings of ADHD and ODD: Identification of susceptible symptoms. Journal of Psychopathology and Behavioral Assessment, 2010, **32**(1): 128-137.
https://doi.org/10.1007/s10862-009-9135-3

[27] Hartun CM, Van Pelt JC, Armendariz ML, *et al*. Biases in ratings of disruptive behavior in children: Effects of sex and negative halos. Journal of Attention Disorders, 2006, **9**(4): 620-630.
https://doi.org/10.1177/1087054705284233

[28] Jackson DA and King AR. Gender differences in the effects of oppositional behavior on teacher ratings of ADHD symptoms. Journal of Abnormal Child Psychology, 2004, **32**(2): 215-224.
https://doi.org/10.1023/B:JACP.0000019772.71251.ff

[29] Gilmore L. Community knowledge and beliefs about ADHD. The Educational and Developmental Psychologist, 2010, **27**(1): 20-30.
https://doi.org/10.1375/aedp.27.1.20

[30] Faraone SV, Monuteaux MC, Biederman J, *et al*. Does parental ADHD bias maternal reports of ADHD symptoms in children? Journal of consulting and clinical psychology, 2003, **71**(1): 168.
https://doi.org/10.1037/0022-006X.71.1.168

[31] Manshadi M, Lippmann S, O'Daniel RG, *et al*. Alcohol abuse and attention deficit disorder. The Journal of clinical psychiatry, 1983, **44**(10): 379-380.

[32] Stavropoulos V, Mastrotheodoros S, Papapetrou S, *et al*. Measurement invariance: The case of measuring romantic attachment in Greek and Cypriot adolescents. European Journal of Developmental Psychology, 2017(a): 1-10.
https://doi.org/10.1080/17405629.2017.1414696

[33] Stavropoulos V, Beard C, Griffiths MD, *et al*. Measurement invariance of the Internet Gaming Disorder ScaleShort-Form (IGDS9-SF) between Australia, the USA, and the UK. International Journal of Mental Health and Addiction, 2017(b): 1-16.
https://doi.org/10.1007/s11469-017-9786-3

[34] Jarrett MA, Wolff JC and Ollendick TH. Concurrent validity and informant agreement of the ADHD module of the Anxiety Disorders Interview Schedule for DSM-IV. Journal of Psychopathology and Behavioral Assessment, 2007, **29**(3): 159-168.
https://doi.org/10.1007/s10862-006-9041-x

[35] Silverman WK, Saavedra LM and Pina AA. Test-retest reliability of anxiety symptoms and diagnoses with anxiety disorders interview schedule for DSM-IV: Child and parent versions. Journal of the American Academy of Child and Adolescent Psychiatry, 2001, **40**(8): 937-944.
https://doi.org/10.1097/00004583-200108000-00016

[36] Silverman WK and Albano AM. The Anxiety Disorders Interview Schedule for DSM-IV: Child Interview Schedule. San Antonio, Texas: Psychological Corporation, 1996.

[37] Grills AE and Ollendick TH. Multiple informant agreement and the anxiety disorders interview schedule for parents and children. Journal of the American Academy of Child & Adolescent Psychiatry, 2003, **42**(1): 30-40.
https://doi.org/10.1097/00004583-200301000-00008

[38] Jensen PS, Rubio-Stipec M, Canino G, *et al*. Parent and child contributions to diagnosis of mental disorder: are both informants always necessary? Journal of the American Academy of Child & Adolescent Psychiatry, 1999, **38**(12): 1569-1579.
https://doi.org/10.1097/00004583-199912000-00019

[39] Achenbach TM and Rescorla L. ASEBA school-age forms & profiles, 2001.

[40] Conners CK, Pitkanen J and Rzepa SR. Conners comprehensive behavior rating scale. In Encyclopedia of clinical neuropsychology. Springer New York, 2011.
https://doi.org/10.1007/978-0-387-79948-3_1536

[41] Muthn LK and Muthn BO. Mplus users guide. Seventh Edition. Los Angeles, CA: Muthn & Muthn, 2012.

[42] DiStefano C. The impact of categorization with confirmatory factor analysis. Structural Equation Modeling, 2002, **9**: 327-346.
https://doi.org/10.1207/S15328007SEM0903_2

[43] Lubke GH and Muthen BO. Applyingmultigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. Structural Equation Modeling, 2004, **11**: 514-534.
https://doi.org/10.1207/s15328007sem1104_2

[44] Satorra A and Bentler PM. A scaled difference chi-square test statistic for moment structure analysis. Psychometrika, 2001, **66**(4): 507-514.
https://doi.org/10.1007/BF02296192

[45] Cantwell DP. Attention deficit disorder: a review of the past 10 years. Journal of the American Academy of Child & Adolescent Psychiatry, 1996, **35**(8): 978-987.
https://doi.org/10.1097/00004583-199608000-00008

[46] Rohde LA, Szobot C, Polanczyk G, *et al*. Attention-deficit/hyperactivity disorder in a diverse culture: do research and clinical findings support the notion of a cultural construct for the disorder? Biological psychiatry, 2005, **57**(11): 1436-1441.
https://doi.org/10.1016/j.biopsych.2005.01.042

[47] Jones C. Archival data: Advantages and disadvantages for research in psychology. Social and Personality Psychology Compass, 2010, **4**(11): 1008-1017.
https://doi.org/10.1111/j.1751-9004.2010.00317.x

[48] Brown TA. Confirmatory factor analysis for applied research. New York: Guilford Press, 2006.