International Journal of Instruction *October 2025 ● Vol.18, No.4*
e-ISSN: 1308-1470 ● www.e-iji.net *p-ISSN: 1694-609X*
pp. 57-74

Article submission code: 20250128071730
Received: 28/01/2025
Revision: 07/04/2025
Accepted: 16/04/2025
OnlineFirst: 01/07/2025

Factorial Validity and Measurement Invariance of the Arabic Version of the Childhood Autism Rating Scale

Mohammed Sameeh Ashour
Asst. Prof., Special Education Department, College of Education, University of Hail, Saudi Arabia, *m.ashour@uoh.edu.sa*

Bander M. Almohayya
Assoc. Prof., Special Education Department, College of Education, University of Hail, Saudi Arabia, *b.alotaibi@uoh.edu.sa*

The aim of this study is to evaluate the psychometric properties and the measurement invariance of the Arabic version of the Childhood Autism Rating Scale- Second Edition (CARS-2). Two hundred Arabic-speaking children, aged 2 to 12 years, who had received an autism spectrum disorder (ASD) diagnosis were recruited to the study. Confirmatory factor analysis (CFA) and Multi-group confirmatory factor analysis (MGCFA) used to evaluate the structure of the scale. McDonald Omega and Cronbach's alpha total coefficient values were both .88. A three-factor solution was best suited to the data and the third CFA iteration yielded significant improvement, with the RMSEA value indicating a good fit and the CFI and TLI values indicating a suitable fit. For the measurement invariance evaluation, the configural model produced an acceptable match for the age variable only. For the gender variable, we found evidence of configural, metric, scalar, and strict, and measurement invariance was stablished for males and females. CARS-2 is currently accessible for clinical and research use in the Arabic language and has strong psychometric properties.

Keywords: autism, childhood autism, special education, childhood autism rating scale, CARS-2, Arabic version

INTRODUCTION

Autism spectrum disorder (ASD) is a neurological condition that affects people all over the world and has various biological and/or genetic origins (Elsabbagh et al., 2012). Numerous characteristics, such as difficulties with social communication and repetitive, restricted interests and activities, are indicative of ASD (American Psychiatric Association, 2013). Several screening tools have been created to identify ASD in young children. Although the fundamental diagnostic traits associated with ASD have been established, its diagnosis remains particularly challenging because individuals with the disorder present with a wide range of symptoms and symptom intensities (Volker et al., 2016). Diagnosis is further complicated by the need to differentiate people with ASD

Citation: Ashour, M. S., & Almohayya, B. M. (2025). Factorial validity and measurement invariance of the Arabic version of the childhood autism rating scale. *International Journal of Instruction*, *18*(4), 57-74.

from those with other developmental conditions who may exhibit comparable traits (e.g., intellectual disability, and language difficulty).

Only a few studies have been conducted on individuals from low- and middle-income brackets to investigate the cause, diagnostic type, clinical picture, and treatments for ASD (Ashwood et al., 2015; Elsabbagh et al., 2012; de Leeuw et al., 2020; Wallace et al., 2012). Most of the currently available information on ASD is from research performed in countries with abundant resources. The overrepresentation of industrialized nations and the underrepresentation of developing and underdeveloped countries in the data on ASD has prompted concerns about the current global imbalance in the ASD literature (Stevanovic et al., 2021). Significant gaps in clinical practice and research have been caused by this asymmetry. The degree to which ASD is conceptualized and measured using psychometric instruments that are variably accessible in different societies is one issue, while another difficulty is the extent to which therapeutic strategies for ASD are applicable in various cultures and settings. (Ashwood et al., 2015; Stevanovic et al., 2021).

Consequently, it is critically important to develop screening instruments that will enable a reliable diagnosis, considering the diverse symptoms of ASD, the various domains of social communication impairment and repetitive, restricted interests and behaviors impacted by ASD, and the challenges associated with differential diagnosis. People with ASD need to be diagnosed as soon as possible, as this will facilitate intensive interventions that have been linked to optimal long-term outcomes (Volker et al., 2016). In this regard, a variety of factors have influenced the classification of the measures used in ASD evaluations, such as the level of training needed to administer the measurement tool (i.e., trained raters versus untrained raters) and the purpose of the measure (i.e., screening versus diagnosis; Lecavalier, 2005; Pandolfi et al., 2010).

Broad diagnostic guidelines for ASD diagnosis have been established by the World Health Organization (WHO, 2018). These guidelines rely primarily on clinical assessments and employ various diagnostic and evaluation instruments (Falkmer et al., 2013; Randall et al., 2018). Nonetheless, a growing body of research suggests that cultural and geographical factors may impact how ASD is evaluated and how this varies between cultures (de Leeuw et al., 2020; Norbury & Sparks, 2013; Stevanovic et al., 2021).

Research findings show that there are variations in the severity of ASD symptoms (Matson et al., 2017; Zachor et al., 2011) and clinical presentations (Amr et al., 2012; Hussein et al., 2011; Magaña et al., 2012; Stevanovic et al., 2021) across various national or cultural groups. For instance, research by Magaña and Vanegas (2017) reports low levels of restricted and repetitive behaviors in Latino children with ASD, and research by Fombonne et al. (2012) reports larger social deficits in Mexican children with ASD than in children from the USA or Germany who have ASD. Furthermore, certain socioeconomic and cultural characteristics have been linked to delayed ASD diagnosis and detection (Ratto et al., 2016; Windham et al., 2014; Stevanovic, 2021). Several studies have shown that cultural variety is essential for the diagnosis and treatment of ASD (Nichols et al., 2020; Stevanovic, 2021). However, because there is a dearth of adequate cross-cultural validity data on ASD assessment instruments, it is debatable whether the cultural variations in ASD is a natural characteristic of the condition and whether cultural and geopolitical influences are relevant to ASD evaluations, or whether there are validity issues with the measurement tools (Stevanovic, et al., 2021).

Almost all diagnostic and assessment instruments currently in use have been created in Western cultures and validated among only certain regional or racial groups (Falkmer et al., 2013; Randall et al., 2018). Although several of these tools are being adapted into other languages, it is occasionally difficult to tell how these translations and cross-cultural adaptations are being done and omits crucial factors that need to be taken into account when utilizing these tools internationally (Cascio, 2015; DuBay & Watson, 2021). Furthermore, considering the high cost of the extensive training needed to use the instruments, the difficulty of administering the tools, and the lack of adequate validation processes, most diagnostic tools are unfeasible for use globally, especially in socioeconomically challenged developing and underdeveloped global regions (Abubakar et al., 2016; Durkin et al., 2015). In order to address the gap in the world and the large gaps in the clinical and scientific knowledge of ASD, robust and reliable assessments and diagnostic instruments need to be made available for use in various global contexts (Durkin et al., 2015; Stevanovic et al., 2021). Based on psychometric theory, the aforementioned solution is achievable only if evaluation methods function consistently and the underlying constructs share a theoretical framework that applies to two or more regional or cultural groups (Dimitrov, 2010; Stevanovic et al., 2021). This indicates cross-cultural validity and is referred to as cross-cultural or regional measurement invariance. Increasingly, to verify the cross-cultural equivalency of an assessment tool and its utility for cross-cultural and intercountry comparisons, researchers are reproducing the conceptual elements of the tool and adapting it to a different language or culture (Byrne & Watkins, 2003; Stevanovic et al., 2021). Nonetheless, the structure of the ASD assessment instrument must be invariant—i.e., there must be cross-cultural measurement invariance— to analyze and compare assessment instrument estimates among various geographic or cultural groups (Byrne & Watkins, 2003; He & van de Vijver, 2012; Stevanovic et al., 2021).

Because of its simplicity, acceptability, conceptual relevance, cost effectiveness, strong psychometric properties, utility across various populations, and strong agreement with the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) autism diagnosis, the Childhood Autism Rating Scale (CARS) is a promising diagnostic tool (Schopler et al., 1988). Hence, this observational rating scale, the CARS (Schopler et al., 2010), is used to assess ASD symptoms. It addresses elements social/communication, stereotyped behavior, abnormalities of the senses, and emotional control (Park & Kim, 2016)—that fit into the two DSM-5 conceptualization categories. A growing body of research shows that the CARS generates valid and trustworthy evaluations of ASD symptomatology in clinical and research contexts across diverse global economic regions (Breidbord & Croudace, 2013; Magyar & Pandolfi, 2007; Mayes et al., 2014; Moulton et al., 2019; Park & Kim, 2016; Russell et al., 2010; Stevanovic et al., 2021).

According to two recent evaluations (Moulton et al., 2019; Randall et al., 2018), the CARS-2 has acceptable to good psychometric properties and modest levels of specificity when used alone as a diagnostic tool for ASD (Stevanovic et al., 2021). The CARS could potentially be a suitable and feasible evaluation and diagnostic instrument for addressing socioeconomic and geographical differences encountered in identifying ASD, especially in developing countries (Samms-Vaughan et al., 2017). Its good psychometric properties, ease of use in a variety of settings, low cost, and short training requirements further contribute to its potential. However, because there is a dearth of information on the cross-cultural measurement invariance of ASD, the extent to which ASD is classified and, consequently, assessed consistently across cultures and nations using this particular measurement tool remains unclear. This is particularly important because proving cross-cultural validity might provide information on cross-cultural diversity among individuals with ASD.

The CARS-2 has recently been translated into Arabic twice: by Akoury-Dirani et al. (2013) and Alotaibi and Alotaibi (2021). The CARS's English version was translated both forward and backward in creating the Arabic version, thereby validating the CARS-2 for Arabic culture. The standard version of the second edition of the CARS, the CARS2, was translated into Arabic for a study in which a sample of young people aged 4 to 19 years was recruited to assess the psychometric properties of the test. Based on the internal consistency value of (.95), The CARS-2 in Arabic was shown to have a high level of reliability (Akoury-Dirani et al., 2013). To verify the model fit and structure of the CARS, the validity of the constructs in the scale's Arabic version was evaluated through the use of factorial analyses (Alotaibi & Alotaibi, 2021). However, to the best of our knowledge, the measurement invariance of the Arabic translation of the CARS-2 has not been examined in previous studies.

In this study, validity and reliability of the Arabic version of the CARS-2 are examined using a clinically recruited population of young children diagnosed with ASD. Notably, this study's evaluation of multifactor models led to the elucidation of the general constructs (i.e., solutions) and specific constructions (i.e., potential multifactor solutions) that the CARS possess. We also conducted a confirmatory factor analysis (CFA) of the CARS-2 to extend the findings of previous studies on the factor structure of the tool to children aged two to twelve. This study also evaluates the measurement invariance of the Arabic version in terms of age and gender. who had been identified as having autism spectrum disorder (ASD), and social communication disorder (SCD) made up the study sample. This study was conducted at a center for diagnosis and evaluation in Hail City, in Saudi Arabia. The significance of this study is that the Arabic version of the CAR-2 does not yet have measurement invariance established. Thus, this is the first study to assess the psychometric properties of the CARS-2 scale's structure and the first one to gather a sufficient sample of autistic children in order to assess both; the CFA and measurement invariance of the Arabic version of the CAR-2.

METHOD

This cross-sectional study employs a quantitative method and was approved by the University of Hail Scientific Research Ethics Committee.

Participants

Two hundred Arabic-speaking children aged 2 to 12 years who had received an ASD diagnosis were recruited to the study. An assessment and diagnosis center in Hail City served as the study site. Sample was chosen purposefully of those children who were aged 2 to 12 years and had received an ASD or SCD diagnosis based on evaluation using the CARS-2. This is explained in greater details in the procedure.

Measures

Medical professionals use individual interviews with parents and caregivers, as well as direct observation of children, to assess the 15 items on the CARS-2. Every child receives a total score that ranges from 15 to 60, with each item scored on a scale of 1 to 4, where 1 = normal at the relevant age and 4 = severely abnormal at the corresponding age (Ji et al., 2023). Notwithstanding significant variations between populations and age groups, the CARS-2 has generally strong psychometric properties. Its internal consistency (0.896) and inter-rater reliability (0.796) are both high (Breidbord & Croudace, 2013). According to Chlebowski et al. (2010), the CARS-2 has strong reliability consistency, having a 0.93 reported coefficient of reliability for children younger than six years old (Moulton et al., 2019). According to Schopler et al. (1988), children who score between 30.5 and 37 on the CARS-2 are considered somewhat-to-moderately autistic, while those who score between 37.5 and 60 are considered severely autistic.

Procedure

All the children who participated in this study had been referred to the diagnosis and evaluation center with a provisional ASD diagnosis, and they were all referred utilizing the database of all unit procedures. These provisional diagnoses were made by clinicians during the first visit of each child to the center. Every clinic visit each child made was recorded, and if a child was suspected of having ASD at any time during that clinical course, they were considered eligible for our study.

The interdisciplinary unit team at the center collected all available information on ASD and SCD diagnoses at the time of data collection for the study. After the team’s psychiatrist made a clinical diagnosis of ASD, the CARS-2 was administered. However, it was administered independently by a speech pathologist and a clinical psychologist. The raters’ behavioral observations of the children served as the basis for their CARS-2 ratings, which were also confirmed against parental data. The assessment team (i.e., the raters) was not informed of the clinical evaluation given by the psychologist to minimize rater bias. The child’s psychologist independently obtained the DSM-5 clinical diagnosis details. Data from children aged 2 to 12 years with traditional ASD or SCD diagnoses were utilized in this validation study.

Data were gathered on those children who were aged 2 to 12 years and had received an ASD or SCD diagnosis based on evaluation using the CARS-2. The DSM-5's most important symptom domains and the CARS-2 are similar. which is why this criterion was chosen. (Moulton et al., 2019). The two types of ASD; which are AD and SCD were based on the last revision the DSM-5 categories and classification of autism disorders. ASD refers to autism disorder as traditional disorder identified in the DSM-5, and SCD in the DSM-5 includes communication difficulties, is included as SCD. Notably, however, individuals with attention deficit hyperactivity disorder, which is linked to ASD and stereotyped movements, were not included in this study due to their unclear nosological status—in line with the updated DSM-5 symptom categories (Russell et al., 2010). To ensure that all requirements were satisfied, each eligible participant was examined independently. The diagnose and evaluation of all autism cases in this study was prior to the edition publication of the American Diagnostic and Statistical Manual of Mental Disorders, Text Revision (DSM-5-TR) which was in 2022.

The following clinical and psychological information was gathered to be used in this procedure. The data on all the study participants, including their CARS-2 scores, were obtained from the final psychologists' reports, and these data served as the basis for validating the CARS-2. Based on the CARS-2 being an in-person observation cognitive test, a licensed clinical psychologist and a speech pathologist administered the psychological evaluations. Information on the test outcomes was gathered from the aforementioned sources independently by a speech therapist and a licensed psychologist. To protect the privacy of the study participants, the collected data were encrypted using a reversible method.

Data Analysis

The mean and standard deviations of the data gathered from the full sample ($n = 200$), as well as the frequencies and percentages of all variables, were calculated using descriptive statistical analysis. The reliability of the CARS-2 was evaluated using Cronbach's alpha and McDonald's omega to assess the internal consistency of the assessed domains; McDonald's omega and Cronbach's alpha values of 0.70 or above were deemed as suitable. The determined factor structure of the sample ($n = 200$) was examined using CFA analysis. CFA was used to confirm structure solution revealed by the Exploratory Factor Analysis (Ersoy et al., 2023). CFA also used to incorporate theoretical or prior model-based hypotheses about the number of constructs and how well those models fit the data (Zeynivandnezhad et al., 2019). We also utilized the normed chi-square ($\chi 2/df$)—where values < 0.06 indicate a good fit, and values ≤ 0.08 may indicate an adequate fit, especially if the value falls outside of the 90% confidence interval—and the index of comparative fit (CFI)—where values ≥ 0.95 indicate a good fit.

Based on previously published work on model measurement invariance, we examined the measurement invariance of the CARS-2 scale items across age and gender (Muthén & Muthén, 2012). To assess measurement parameter invariance, we used hierarchical tests, and to determine whether the variations among the population matrixes were similar, we used a hypothesis test—an RMSEA score less than the lowest threshold would be indicative of the overall stability of the instrument. We then considered the configuration invariance model. This condition must be met to test for invariance by comparing the configuration invariance model to alternative invariance models using fit indices.

Next, we investigated metric invariance, which confirms that evaluations made using the metrics are on a comparable scale across genders. We then examined the metric invariance model. Under this paradigm, gender invariance applies to item intercepts and factor loadings, allowing for cross-gender comparisons of the underlying factor. Next, the strict invariance model—which necessitates the invariance of residual variances, intercepts, and factor loadings—was looked at.

Based on recommendations taken from the literature, the demonstration of invariance between the less restrictive models (configural invariance model) and the more restricted models (weak measurement invariance models) was established (Cheung & Rensvold, 2002; Wang & Wang, 2012). If the CFI's change in value (ΔCFI) is less than or equal to 0.01, we do not reject the invariance hypothesis. The values of 0.015 and 0.01 are critical for a change in the Tucker–Lewis index (ΔTLI) and the ΔRMSEA, respectively. To compare each model with the previous one, a chi-square difference test was conducted. Factor loadings of 0.40 and higher were employed, along with significant p-values, covariance residuals, and modification indices (Wang & Wang, 2012), to detect any problematic items that might contribute to a mismatch with the data.

We considered it important to include the covariances between the items that were found. Therefore, based on the modification index (MI) values, residual covariances between the items were added as a parameter. We used Amos (version 25) and the Statistical Package for the Social Sciences (version 27) to analyze our data in all the statistical analyses performed in this study.

FINDINGS

A total of 200 children were recruited to this study, with 60.5% of the study sample (N = 121) comprising males and 39.5% comprising females (N = 79). Furthermore, 48% of the participants were between the ages of 2 and 6 years, and 52% were between the ages of 6 and 12 years. In addition, it was determined that 42% of the participants (N = 84) had SCD and 58% (N = 116) had ASD (Table 1). With a 74.4% approximation rate to the maximum score, we employed a CARS-2 threshold of 26 for our sample of 200 children with ASD or SCD, aged two to twelve; the range of scores was 15 to 49, with a diagnosis of mild autism defined as a score of 26.

Table 1
Characteristics of study participants

Characteristics	(n = 200)	
Descriptive Statistics	Frequency	(%)
Age in Years		
2–6	96	(48%)
7–12	104	(52%)
Total	200	(100%)
Gender		
Male	121	(60.5%)
Female	79	(39.5%)
Severity		
Mild autism	78	(39%)
Moderate autism	90	(45%)
Severe autism	32	(16%)
Type of Disorder		
Autism spectrum disorder (ASD)	116	(58%)
Social communication disorder (SCD)	84	(42%)

Confirmatory Factor Analysis (CFA)

We found no univariate or multivariate outliers when the multivariate normality and linearity assumptions were examined using Amos 25.0. A series of CFAs were run to investigate the model fit. We decided to first explore the solutions of the different models by initially entering all 15 items and then testing the model fit of the three factors using the deletion technique. Subsequently, we evaluated the MIs and then added residual covariances between the items as a parameter to improve the model fit, as needed. The first model included all 15 items and produced a poor fit, with $\chi2(df = 87) = 244.444$, CFI = .86, TLI = .83, and RMSEA = .09. The MIs indicated the need to delete the *Listening Response and General Impression* items. The model was then run with 13 items, and the second run also produced a poor fit, with a slight increase in the TLI (.87) and CFI (.89), while the RMSEA remained at .09, indicating a poor fit. Essentially, with the following values: $\chi2(df = 62) = 165.919$, CFI = .89, TLI = .87, and RMSEA = .09, the model fit was poor. Consequently, the residual covariances between the *Imitation and Verbal Communication* items as a parameter were introduced; this was done to enhance the model. This third run showed improvement, producing the following results: $\chi2(df = 78) = 148.017$, CFI = .91, TLI = .88, and RMSEA = .08. In addition, the residual covariances were again added between the *Verbal Communication item and the Relation to People* item. It can be said that verbal communication is a tool with which to initiate relationships with other people. Thus, residual covariances were added between those two items (*Verbal Commination, and Relation to People*). Another noticeable improvement was observed in the fourth run, which yielded the following values: $\chi2(df = 60) = 133.193$, CFI = .90, TLI = .92, and RMSEA = .07. The RMSEA value indicates a good fit and the CFI and TLI values indicate an adequate fit; thus, the fit of the three-factor model is acceptable. Comparing all past attempts to the three-factor model, the Akaike's information criterion and the consistent Akaike's information criterion kept decreasing, Relative improvements in the fit were indicated by the NFI and GFI continuing to rise. (Table 2).

Table 2
Indices of fit for every CFA-tested model (N = 200)

Attempts for Each Model	χ^2	df	χ^2/df	CFI	TLI	RMSEA	NFI	GFI	AIC	ECVI
First run-15 items	244.444	87	2.80	.86	.83	.09	.80	.95	6720.835	1.70
Second run-13 items	165.919	62	2.67	.89	.87	.09	.84	.97	5801.635	1.25
Residual covariance (RC)	148.017	78	2.42	.91	.88	.08	.86	.97	5785.734	1.17
Second run of RC	133.193	60	2.21	.92	.90	.07	.87	.97	5772.909	1.10

The reliability statistics for the CARS-2 was computed using 13 items (N = 13), and it was found to have strong internal consistency. The calculation of the Cronbach's alpha and McDonald's omega values was used to assess the internal consistency of the CARS. Based on Cronbach's alpha and McDonald's omega coefficients, the CARS-2 had good reliability for the total scores (.88) for both tests (Table 2).

Table 3
Cronbach's Alpha values for the total correlation of each factor item when an item is deleted and corrected

Construct	Item	Omega if Item Deleted	Cronbach if Item Deleted	Corrected Item-Total Correlation
Construct one	2	.869	.865	.697
	11	.874	.868	.647
	12	.881	.876	.502
	14	.876	.871	.606
	1	.870	.865	.700
	7	.877	.872	.594
Construct two	3	.867	.862	.749
	10	.886	.880	.427
	6	.891	.889	.212
Construct three	4	.878	.871	.591
	5	.878	.872	.579
	9	.882	.877	.471
	13	.880	.874	.549

Convergent Validity

The Pearson correlation coefficient test was used to test the correlation of each item with the full sample of 200 ($df = N-2 = 200-2 = 198$); the critical value comparison approach was also employed. The results reveal that the total coefficient value of each item was greater than the critical value. Item 1 obtained a value of .56, which is greater than the critical value of .11 and is thus highly significant. In addition, the remainder of the values obtained for all the remaining items were less than 0.11 (Table 4). This indicates that convergent validity was achieved (Table 4).

Table 4
Pearson correlation for each item

Item #	Value obtained	>.13 (critical value)	Significant
1	.47**	Yes	At .05
2	.49**	Yes	At .05
3	.51**	Yes	At .05
4	.37**	Yes	At .05
5	.35**	Yes	At .05
6	.14**	Yes	At .05
7	.34**	Yes	At .05
9	.31**	Yes	At .05
10	.24**	Yes	At .05
11	.50**	Yes	At .05
12	.44**	Yes	At .05
13	.40**	Yes	At .05

Measurement Invariance

We evaluated the gender and age measurement invariance of the Arabic version of the CARS-2 using a configural invariance model (Table 5), which achieved a satisfactory fit with the gender data. The information presented in Table 5 demonstrates that the configural invariance model satisfactorily fits the gender variables in the data. We then contrasted the configural model with a more restrictive measurement invariance model (i.e., metric measurement invariance). In Table 5, it can be seen that the metric invariance model fits the data quite well.

Table 5
Evidence of invariance between all types of measurement invariance for gender and age variables for the Arabic version of CARS-2

Measurement Invariance	χ^2 (df)	CFI	TLI	RMSEA (90% CI)	SRMR	$\Delta\chi^2(\Delta df)$	ΔCFI	ΔTLI	ΔRMSEA	ΔSRMR
Gender										
Configural	188.051 (118)	.929	.906	.077 (.056, .097)	0.061	-	-	-	-	-
Metric	193.168 (128)	.934	.919	.071 (.050, .091)	0.067	5.117 (10)	0.005	0.013	−0.006	0.006
Scalar	197.905 (138)	.939	.931	.066 (.044, .086)	0.068	4.737 (10)	0.005	0.012	−0.005	0.001
Strict	209.965 (154)	.943	.942	.060 (.038, .080)	0.071	12.06 (16)	0.004	0.011	−0.006	0.003
Age										
Configural	275.068 (118)	.860	.814	.115 (.098, .133)	0.072	-	-	-	-	-
Metric	302.781 (128)	.844	.809	.117 (.100, .134)	0.100	27.713 (10)	−0.016	−0.005	0.002	0.028

A comparison between the configural invariance model and the metric invariance model revealed that the factor score metric was gender invariant. The changes in TLI, RMSEA, and CFI were all acceptable (ΔTLI = 0.013, ΔCFI = 0.005, and ΔRMSEA = −0.006). implying that the meaning of the items used to approximate the factor loadings was the same for male and female. The data in Table 5 show how well the scalar invariance model, which was the subsequent restrictive model, fits the data. A scalar invariance model was constructed by restricting the factor loadings and item intercept in the second, stricter model, which exhibited a strong invariance (ΔCFI = −0.005, ΔTLI = 0.012, ΔRMSEA = −0.005). This implies that there is no gender variation in the item intercept and factor loadings.

We next evaluated the last, stricter model, strict or invariant uniqueness model, by lowering the factor loadings, item intercept, and residual variances. The suggested values and revised indices of fit (ΔCFI = −0.004, ΔTLI = 0.011, ΔRMSEA = −0.006) were all in the same range of the changes and was within the recommended values. This indicates that it is valid to compare the average item scores of males with those of females. We then performed the same analysis on the age variable, starting with the configural invariance model that fits the data (Table 5). The results show a poor model fit, with the CFI, TLI, and RMSEA values for the metric run being .860, .814, and .115, respectively.

We then compared this configural model with a more constrained measurement invariance (i.e., metric measurement invariance) model, and the CFI, TLI, and RMSEA values for the metric run were .84, .809, and .117, respectively. Using the matrix measurement invariance worsens the structure, and we found that the model fit for the group comprising the ≤6 year olds was somewhat worse than that of the groups comprising the ≥6 year olds. Notwithstanding knowing that the metric invariance model and the configural invariance model had acceptable differences in CFI, TLI, and RMSEA (ΔCFI = 0.002, ΔTLI = −0.016, ΔRMSEA = 0.002), we chose to characterize the metric invariance model as having a poorer match for the age variable than the configural invariance model.

This result indicates that there was non-invariance in the factor score metric between the age groups. In essence, the items used to evaluate the factor loadings for the groups comprised of <6-year-olds are different from those used for the group comprising children 6 years old and older.

DISCUSSION

Several statistical methods were employed to validate the psychometric parameters of the Confirmatory factor investigations and multigroup confirmatory factor analyses were used to confirm the factorial structure of the scale. The results indicated the need for a three-factor solution. McDonald's omega and Cronbach's alpha were used to evaluate the proposed domains' internal consistency, and it was found to be good. In addition, a multigroup confirmatory component analysis was performed to determine whether the scale is age and gender invariant. The validation studies on the Arabic version of the CARS in cultural contexts that we found in the literature do not evaluate measurement invariance (Akoury-Dirani et al., 2013; Alotaibi & Alotaibi, 2021).

The McDonald's omega (.886) and Cronbach's alpha coefficient (.88) values of the reliability statistics results for the CARS-2 indicate that it has strong internal consistency, which aligns with the findings of previous studies. Our analysis of studies that utilized the CARS-2 between 1980 and 2021 of good internal consistency. For instance, our analysis of the internal consistency of the CARS-2 shows that the *Cronbach's α coefficient* value reported in earlier research studies varies between 0.62 and 0.92 (Kurita et al., 1989; Russell et al., 2010; Moulton et al., 2019). Cronbach's α coefficient of internal consistency obtained in our study was 0.88, which is congruent with the findings of CARS-2validation research conducted in different cultures, as well as in Arabic cultures, (Akoury-Dirani et al. 2013; Alotaibi & Alotaibi 2021). This demonstrates that, when applied in the Arabic context, the CARS-2 factors are consistent. Overall, the literature supports the reliability of the CARS-2. Most studies report an acceptable internal consistency, with alpha coefficients often at or exceeding .90 (Magyar & Pandolf, 2007). Our study found the same result with other previous cultural studies in Arabic context, and this might suggest that the CARS-2 is a reliable and accurate tool to diagnose children with autism. The consistency of those results indicated that there is a potential impact on clinical practice in Arabic-speaking regions to use this scale and benefit from its accuracy in clinical and psychometric evaluation of autism.

In terms of the number of factors kept, this study found that a three-factor solution operated better. The solution employing three factors fits the data most closely, according to the findings of studies by DiLalla and Rogers (1994), Moulton et al. (2019), and Alotaibi and Alotaibi (2021), whose findings most closely align with ours. Three factors were identified in our study—although the specific item loadings varied slightly—and these factors have several similarities with factors identified in research conducted by Russell et al. (2010) and Magyar and Pandolfi (2007), who have reviewed more than three components. Our factor analysis yielded different results from Akoury-Dirani et al. (2013)'s earlier validation research of the Arabic version of the CARS-2, whose results yield a two-factor solution.

To further confirm the CARS-2 validation obtained in this study, measurement invariance was assessed, as it can only be performed after such equivalency is proven. The factor solution was the same for both males and females. However, although the matrix failed to yield a reasonable fit for the age variable, the configural model did. In addition, we discovered evidence of scalar, structural, uniqueness, and metric invariance, indicating that the relationships between the components are the same for both males and females. This finding differs from (Stevanovic et al. 2021), a recent study that found that the CARS-2 lacks cross-cultural validity. For example, Stevanovic et al. (2021) studied measurement invariance across six countries and found that the CARS-2 structure is unstable. As such, across geographically and linguistically diverse groups, raters employing the CARS-2 questions may report and assess observed ASD symptoms in different ways. Stevanovic et al. (2021) discussed that this might be the result of factors like how the child's family rates the scale, the evaluation's setting, and the assessment itself—all of which could have no direct connection to the region, language, or social group of the child being assessed. However, as noted by de Leeuw et al. (2020), there might be real variations between cultures and countries in how children with ASD present symptoms at different severity levels, which would be reflected in the CARS-2 ratings (Stevanovic et al., 2021).

Our results build substantially on a small body of research on cross-cultural measurement invariance in ASD instruments. To the best of our knowledge, this is the second study to investigate the measurement invariance of the CARS-2, with the cross-cultural study by Stevanovic et al. (2021) being the first. In addition, we provide a solution to the models with poor fit tested in our initial CFA runs. For instance, we chose to include covariance between items from the same factor to modify the structure of the Arabic version of the CARS-2. These changes to the proposed model were made after we established that there was adequate theoretical support for the measurement invariance, and the changes were based on the MI values. Items from the same factor were implicated in two residual covariances. We provided residual covariances addition. Given that the two items are derived from the same factor, the covariance between the residuals for both items was reasonable. The theoretical justification was based on the two items being derived from Factor 1, and the covariance between the residuals for both items was reasonable. Furthermore, because imitation is subsumed under verbal communication and is a type of social communication for children, it is possible to connect these two items based on their common social backgrounds. We used this modification to the model to derive a better-fitting model, and significant meaningfulness was considered before incorporating these residual covariances into the models. These factors can make significant substantive sense, particularly in social psychology research; therefore, they ought to be incorporated into the model (Cole et al., 2007; Jöreskog & Sörbom, 1993; Kueh et al., 2018).

We are aware that this study has some limitations. The correlation between the CARS-2 and other gold standard or level one measures were not evaluated for the convergent and discriminant validity, like the Autism Diagnostic Observation Schedule, second edition (ADOS-2). Criterion validity was also not used or included in other measures to compare convergent and discriminant validity. For the results to be generalizable, it may be more appropriate to test the diagnostic accuracy of the CARS-2 against a gold standard diagnostic instrument that is stronger than the CARS (e.g., Autism Diagnostic Interview–Revised, Autism Diagnostic Observation Schedule–Second Edition). Furthermore, we could not use diagnostic accuracy to investigate the cutoff score, sensitivity, or specificity of the CARS-2.

REFERENCES

Abubakar, A., Ssewanyana, D., & Newton, C. R. (2016). A Systematic Review of Research on Autism Spectrum Disorders in Sub-Saharan Africa. *Behavioural Neurology*, *2016*, 1–14. https://doi.org/10.1155/2016/3501910

Akoury-Dirani, L., Alameddine, M., & Salamoun, M. (2013). Validation of the Lebanese Childhood Autism Rating Scale-Second Edition-Standard Version. *Research in Autism Spectrum Disorders*, *7*(9), 1097–1103. https://doi.org/10.1016/j.rasd.2013.05.004

Alotaibi, B., & Alotaibi, A. (2021). Exploratory and confirmatory factor analyses of the Arabic version of the Childhood Autism Rating Scale. *Research in Autism Spectrum Disorders*, *86*, 101827. https://doi.org/10.1016/j.rasd.2021.101827

Amr, M., Bu Ali, W., Hablas, H., Raddad, D., El-Mehesh, F., El-Gilany, A.-H., & Al-Shamy, H. (2012). Sociodemographic factors in Arab children with Autism Spectrum Disorders. *The Pan African Medical Journal*, *13*, 65.

American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5-TR*. American Psychiatric Association.

Ashwood, K. L., Buitelaar, J., Murphy, D., Spooren, W., & Charman, T. (2015). European clinical network: autism spectrum disorder assessments and patient characterisation. *European Child & Adolescent Psychiatry*, *24*(8), 985–995. https://doi.org/10.1007/s00787-014-0648-2

Breidbord, J., & Croudace, T. J. (2013). Reliability Generalization for Childhood Autism Rating Scale. *Journal of Autism and Developmental Disorders*, *43*(12), 2855–2865. https://doi.org/10.1007/s10803-013-1832-9

Byrne, B. M., & Watkins, D. (2003). The Issue Of Measurement Invariance Revisited. *Journal of Cross-Cultural Psychology*, *34*(2), 155–175. https://doi.org/10.1177/0022022102250225

Cascio, M. A. (2015). Cross-Cultural Autism Studies, Neurodiversity, and Conceptualizations of Autism. *Culture, Medicine, and Psychiatry*, *39*(2), 207–212. https://doi.org/10.1007/s11013-015-9450-y

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Chlebowski, C., Green, J. A., Barton, M. L., & Fein, D. (2010). Using the Childhood Autism Rating Scale to Diagnose Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *40*(7), 787–799. https://doi.org/10.1007/s10803-009-0926-x

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *12*(4), 381–398. https://doi.org/10.1037/1082-989X.12.4.381

de Leeuw, A., Happé, F., & Hoekstra, R. A. (2020). A Conceptual Framework for Understanding the Cultural and Contextual Factors on Autism Across the Globe. *Autism Research*, *13*(7), 1029–1050. https://doi.org/10.1002/aur.2276

DiLalla, D. L., & Rogers, S. J. (1994). Domains of the childhood autism rating scale: Relevance for diagnosis and treatment. *Journal of Autism and Developmental Disorders*, *24*(2), 115–128. https://doi.org/10.1007/BF02172092

DuBay, M., Watson, L. R., Baranek, G. T., Lee, H., Rojevic, C., Brinson, W., Smith, D., & Sideris, J. (2021). Rigorous Translation and Cultural Adaptation of an Autism Screening Tool: First Years Inventory as a Case Study. *Journal of Autism and*

Developmental Disorders, *51*(11), 3917–3928. https://doi.org/10.1007/s10803-020-04837-1

Durkin, M. S., Elsabbagh, M., Barbaro, J., Gladstone, M., Happe, F., Hoekstra, R. A., Lee, L., Rattazzi, A., Stapel-Wax, J., Stone, W. L., Tager-Flusberg, H., Thurm, A., Tomlinson, M., & Shih, A. (2015). Autism screening and diagnosis in low resource settings: Challenges and opportunities to enhance research and services worldwide. *Autism Research*, *8*(5), 473–476. https://doi.org/10.1002/aur.1575

Elsabbagh, M., Divan, G., Koh, Y., Kim, Y. S., Kauchali, S., Marcín, C., Montiel-Nava, C., Patel, V., Paula, C. S., Wang, C., Yasamy, M. T., & Fombonne, E. (2012). Global Prevalence of Autism and Other Pervasive Developmental Disorders. *Autism Research*, *5*(3), 160–179. https://doi.org/10.1002/aur.239

Falkmer, T., Anderson, K., Falkmer, M., & Horlin, C. (2013). Diagnostic procedures in autism spectrum disorders: a systematic literature review. *European Child & Adolescent Psychiatry*, *22*(6), 329–340. https://doi.org/10.1007/s00787-013-0375-0

Ersoy, M., Eren, E., Avcı, Z. Y., & Kandemir, C. M. (2023). Development of the perception Scale for flipped Learning model. *Anatolian Journal of Education*, *8*(1), 63–78. https://doi.org/10.29333/aje.2023.815a

Fombonne, E., Marcin, C., Bruno, R., Tinoco, C. M., & Marquez, C. D. (2012). *Screening for Autism in Mexico. Autism Research*, *5*(3), 180–189.

García-López, C., & Narbona, J. (2014). Inventario del espectro autista y childhood autism rating scale: correspondencia con criterios DSM-IV-TR en pacientes con trastornos generalizados del desarrollo. *Anales de Pediatría*, *80*(2), 71–76. https://doi.org/10.1016/j.anpedi.2013.05.012

He, J., & van de Vijver, F. (2012). Bias and Equivalence in Cross-Cultural Research. *Online Readings in Psychology and Culture*, *2*(2). https://doi.org/10.9707/2307-0919.1111

Hussein, H., Taha, G. R., & Almanasef, A. (2011). Characteristics of autism spectrum disorders in a sample of egyptian and saudi patients: transcultural cross sectional study. *Child and Adolescent Psychiatry and Mental Health*, *5*(1), 34. https://doi.org/10.1186/1753-2000-5-34

Ji, S.-I., Park, H., Yoon, S. A., & Hong, S.-B. (2023). A Validation Study of the CARS-2 Compared With the ADOS-2 in the Diagnosis of Autism Spectrum Disorder: A Suggestion for Cutoff Scores. *Journal of the Korean Academy of Child and Adolescent Psychiatry*, *34*(1), 45–50. https://doi.org/10.5765/jkacap.220027

Jöreskog, K. G., & Sörbom, D. (1993). *Structural equation modeling with the SIMPLIS command language. In LISREL 8: Structural equation modeling with the SIMPLIS command language*.

Kueh, Y. C., Abdullah, N., Kuan, G., Morris, T., & Naing, N. N. (2018). Testing Measurement and Factor Structure Invariance of the Physical Activity and Leisure

Motivation Scale for Youth Across Gender. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01096

Kurita, H., Miyake, Y., & Katsuno, K. (1989). Reliability and validity of the Childhood Autism Rating Scale-Tokyo version (CARS-TV). *Journal of Autism and Developmental Disorders*, *19*(3), 389–396. https://doi.org/10.1007/BF02212937/METRICS

Lecavalier, L. (2005). An Evaluation of the Gilliam Autism Rating Scale. *Journal of Autism and Developmental Disorders*, *35*(6), 795–805. https://doi.org/10.1007/s10803-005-0025-6

Magaña, S., Parish, S. L., Rose, R. A., Timberlake, M., & Swaine, J. G. (2012). Racial and Ethnic Disparities in Quality of Health Care Among Children with Autism and Other Developmental Disabilities. *Intellectual and Developmental Disabilities*, *50*(4), 287–299. https://doi.org/10.1352/1934-9556-50.4.287

Magaña, S., & Vanegas, S. B. (2017). Diagnostic Utility of the ADI-R and DSM-5 in the Assessment of Latino Children and Adolescents. *Journal of Autism and Developmental Disorders*, *47*(5), 1278–1287. https://doi.org/10.1007/s10803-017-3043-2

Magyar, C. I., & Pandolfi, V. (2007). Factor Structure Evaluation of the Childhood Autism Rating Scale. *Journal of Autism and Developmental Disorders*, *37*(9), 1787–1794. https://doi.org/10.1007/s10803-006-0313-9

Matson, J. L., Matheis, M., Burns, C. O., Esposito, G., Venuti, P., Pisula, E., Misiak, A., Kalyva, E., Tsakiris, V., Kamio, Y., Ishitobi, M., & Goldin, R. L. (2017). Examining cross-cultural differences in autism spectrum disorder: A multinational comparison from Greece, Italy, Japan, Poland, and the United States. *European Psychiatry*, *42*, 70–76. https://doi.org/10.1016/j.eurpsy.2016.10.007

Mayes, S. D., Calhoun, S. L., Murray, M. J., Pearl, A., Black, A., & Tierney, C. D. (2014). Final DSM-5 under-identifies mild Autism Spectrum Disorder: Agreement between the DSM-5, CARS, CASD, and clinical diagnoses. *Research in Autism Spectrum Disorders*, *8*(2), 68–73. https://doi.org/10.1016/j.rasd.2013.11.002

Moulton, E., Bradbury, K., Barton, M., & Fein, D. (2019). Factor Analysis of the Childhood Autism Rating Scale in a Sample of Two Year Olds with an Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *49*(7), 2733–2746. https://doi.org/10.1007/s10803-016-2936-9

Muthén, L. K., & Muthén, B. O. n. (2012). *Mplus User's Guide* (7th ed.). CA: Muthén & Muthén.

Nichols, H. M., Dababnah, S., Troen, B., Vezzoli, J., Mahajan, R., & Mazefsky, C. A. (2020). Racial Disparities in a Sample of Inpatient Youth with ASD. *Autism Research*, *13*(4), 532–538. https://doi.org/10.1002/aur.2262

Norbury, C. F., & Sparks, A. (2013). Difference or disorder? Cultural issues in understanding neurodevelopmental disorders. *Developmental Psychology*, *49*(1), 45–58. https://doi.org/10.1037/a0027446

Pandolfi, V., Magyar, C. I., & Dill, C. A. (2010). Constructs Assessed by the GARS-2: Factor Analysis of Data from the Standardization Sample. *Journal of Autism and Developmental Disorders*, *40*(9), 1118–1130. https://doi.org/10.1007/s10803-010-0967-1

Park, E., & Kim, J. (2016). Factor structure of the Childhood Autism Rating Scale as per DSM-5. *Pediatrics International*, *58*(2), 139–145. https://doi.org/10.1111/ped.12770

Randall, M., Egberts, K. J., Samtani, A., Scholten, R. J., Hooft, L., Livingstone, N., Sterling-Levis, K., Woolfenden, S., & Williams, K. (2018). Diagnostic tests for autism spectrum disorder (ASD) in preschool children. *Cochrane Database of Systematic Reviews*, *2018*(7). https://doi.org/10.1002/14651858.CD009044.pub2

Ratto, A. B., Reznick, J. S., & Turner-Brown, L. (2016). Cultural Effects on the Diagnosis of Autism Spectrum Disorder Among Latinos. *Focus on Autism and Other Developmental Disabilities*, *31*(4), 275–283. https://doi.org/10.1177/1088357615587501

Russell, P. S., Daniel, A., Russell, S., Mammen, P., Abel, J. S., Raj, L. E., Shankar, S. R., & Thomas, N. (2010). Diagnostic accuracy, reliability and validity of Childhood Autism Rating Scale in India. *World Journal of Pediatrics*, *6*(2), 141–147. https://doi.org/10.1007/s12519-010-0029-y

Samms-Vaughan, M., Rahbar, M. H., Dickerson, A. S., Loveland, K. A., Hessabi, M., Pearson, D. A., Bressler, J., Shakespeare-Pellington, S., Grove, M. L., Coore-Desai, C., Reece, J., & Boerwinkle, E. (2017). The diagnosis of autism and autism spectrum disorder in low- and middle-income countries: Experience from Jamaica. *Autism*, *21*(5), 564–572. https://doi.org/10.1177/1362361317698938.

Schopler, E., Reichler, R. J., & Renner, B. R. (1988). *The Childhood Autism Rating Scale.* Los Angeles, CA: Western Psychological Services.

Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). The childhood autism rating scale (2nd ed.). Western Psychological Services.

Stevanovic, D., Costanzo, F., Fucà, E., Valeri, G., Vicari, S., Robins, D. L., Samms-Vaughan, M., Ozek Erkuran, H., Yaylaci, F., Deshpande, S. N., Deshmukh, V., Arora, N. K., Albores-Gallo, L., García-López, C., Gatica-Bahamonde, G., Gabunia, M., Zirakashvili, M., Machado, F. P., Radan, M., . . . Knez, R. (2021). Measurement invariance of the Childhood Autism Rating Scale (CARS) across six countries. *Autism Research*, *14*(12), 2544–2554. https://doi.org/10.1002/aur.2586.

Vaughan, C. A. (2011). Test Review: E. Schopler, M. E. Van Bourgondien, G. J. Wellman, & S. R. Love Childhood Autism Rating Scale (2nd ed.). Los Angeles, CA: Western Psychological Services, 2010. *Journal of Psychoeducational Assessment*, *29*(5), 489–493. https://doi.org/10.1177/0734282911400873

Volker, M. A., Dua, E. H., Lopata, C., Thomeer, M. L., Toomey, J. A., Smerbeck, A. M., Rodgers, J. D., Popkin, J. R., Nelson, A. T., & Lee, G. K. (2016). Factor Structure, Internal Consistency, and Screening Sensitivity of the GARS-2 in a Developmental Disabilities Sample. *Autism Research and Treatment*, *2016*, 1–12. https://doi.org/10.1155/2016/8243079

Wallace, S., Fein, D., Rosanoff, M., Dawson, G., Hossain, S., Brennan, L., Como, A., & Shih, A. (2012). A Global Public Health Strategy for Autism Spectrum Disorders. *Autism Research*, *5*(3), 211–217. https://doi.org/10.1002/aur.1236.

Wang, J., & Wang, X. (2012). *Structural Equation Modeling*. Wiley. https://doi.org/10.1002/9781118356258

Windham, G. C., Smith, K. S., Rosen, N., Anderson, M. C., Grether, J. K., Coolman, R. B., & Harris, S. (2014). Autism and Developmental Screening in a Public, Primary Care Setting Primarily Serving Hispanics: Challenges and Results. *Journal of Autism and Developmental Disorders*, *44*(7), 1621–1632. https://doi.org/10.1007/s10803-014-2032-y

Zachor, D., Yang, J., Itzchak, E. Ben, Furniss, F., Pegg, E., Matson, J. L., Horovitz, M., Sipes, M., Chung, K.-M., & Jung, W. (2011). Cross-cultural differences in comorbid symptoms of children with autism spectrum disorders: An international examination between Israel, South Korea, the United Kingdom and the United States of America. *Developmental Neurorehabilitation*, *14*(4), 215–220. https://doi.org/10.3109/17518423.2011.568468.

Zeynivandnezhad, F., Rashed, F., & Kanooni, A. (2019). Exploratory Factor Analysis for TPACK among Mathematics Teachers: Why, What and How. *Anatolian Journal of Education*, *4*(1). https://doi.org/10.29333/aje.2019.416a