



Universiteit
Leiden
The Netherlands

Measurement invariance of the Childhood Autism Rating Scale (CARS) across six countries

Stevanovic, D.; Costanzo, F.; Fucà, E.; Valeri, G.; Vicari, S.; Robins, D.L.; ... ; Knez, R.

Citation

Stevanovic, D., Costanzo, F., Fucà, E., Valeri, G., Vicari, S., Robins, D. L., ... Knez, R. (2021). Measurement invariance of the Childhood Autism Rating Scale (CARS) across six countries. *Autism Research*, 14(12), 2544-2554. doi:10.1002/aur.2586

Version: Publisher's Version










License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/3279344>

Note: To cite this publication please use the final published version (if applicable).

RESEARCH ARTICLE

Measurement invariance of the Childhood Autism Rating Scale (CARS) across six countries

Dejan Stevanovic¹  | Floriana Costanzo² | Elisa Fucà² | Giovanni Valeri² | Stefano Vicari² | Diana L. Robins³ | Maureen Samms-Vaughan⁴ | Handan Ozek Erkurun⁵ | Ferhat Yaylaci⁶ | Smita N. Deshpande⁷ | Vaishali Deshmukh⁸ | Narendra K. Arora⁸ | Lilia Albores-Gallo⁹  | Cristina García-López¹⁰  | Gabriel Gatica-Bahamonde¹¹  | Maia Gabunia¹²  | Medea Zirakashvili¹³ | Fernanda Prada Machado¹⁴ | Miruna Radan¹⁵ | Sayyed Ali Samadi¹⁶  | Teck-Hock TOH¹⁷ | Windham Gayle¹⁸  | Laura Brennan¹⁹ | Tatjana Zorcec²⁰ | Alejandra Auza²¹  | Maretha de Jonge²² | Noordeen Shoqirat²³ | Anita Marini²⁴ | Rajna Knez^{25,26} 

¹Department of Psychiatry, Clinic for Neurology and Psychiatry for Children and Youth, Belgrade, Serbia

²Child and Adolescent Neuropsychiatry Unit, Bambino Gesù Children's Hospital, Rome, Italy

³A.J. Drexel Autism Institute, Drexel University, Philadelphia, Pennsylvania, USA

⁴University of the West Indies, Kingston, Jamaica

⁵Child and Adolescent Psychiatry Unit, Dr Behcet Uz Children's Research and Training Hospital, Izmir, Turkey

⁶Child Psychiatry Unit, Bursa Dörtcelik Children's Hospital, Bursa, Turkey

⁷Department of Psychiatry, Centre of Excellence in Mental health, Atal Bihari Vajpayee Institute of Medical Sciences; Dr. Ram Manohar Lohia Hospital, New Delhi, India

⁸INCLIN Trust, New Delhi, India

⁹Research Division, Hospital Psiquiátrico Infantil "Dr. Juan N. Navarro," Secretaría de Salud, Mexico City, Mexico

¹⁰Hospital Sant Joan de Déu, Barcelona, Spain

¹¹Hospital de Villarrica, Department of Mental Health, Universidad de La Frontera, Temuco, Chile

¹²Mental Health Center, Tbilisi, Georgia

¹³Iliia State University, Tbilisi, Georgia

¹⁴Pontificia Universidade Católica de São Paulo, São Paulo, Brazil

¹⁵National Institute for Maternal and Child Health, Bucharest, Romania

¹⁶Institute of Nursing and Health Research, University of Ulster, Coleraine, UK

¹⁷Clinical Research Centre & Department of Pediatrics, Sibü Hospital, Ministry of Health Malaysia, Sibü, Malaysia

¹⁸Environmental Health Investigations Branch, California Department of Public Health, Richmond, California, USA

¹⁹University of Connecticut, Storrs, Connecticut, USA

²⁰Developmental Department, University Children's Hospital, Skopje, Macedonia

²¹Language and Cognition Laboratory, Hospital General Dr. Manuel Gea González, Mexico City, Mexico

²²Faculty of Social Science, Education and Child Studies, Clinical Neuroscience and Developmental Disorders, Leiden University, Leiden, The Netherlands

²³Faculty of Nursing, Mu'tah University, Mu'tah, Jordan

²⁴World Health Organization, Geneva, Switzerland

²⁵Department of Women's and Children's Health, Skaraborgs Hospital, Skövde, Sweden

²⁶Sahlgrenska Academy, Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden

Correspondence

Stevanovic Dejan, Clinic for Neurology and Psychiatry for Children and Youth, Dr Subotica 6a, 11000 Belgrade, Serbia.
Email: stevanovic.dejan79@gmail.com

Abstract

The Childhood Autism Rating Scale (CARS) is a simple and inexpensive tool for Autism spectrum disorder (ASD) assessments, with evidenced psychometric data from different countries. However, it is still unclear whether ASD symptoms are measured the same way across different societies and world regions with this tool, since data on its cross-cultural validity are lacking. This study evaluated the cross-cultural measurement invariance of the CARS among children with ASD from six countries, for whom data were aggregated from previous studies in India ($n = 101$), Jamaica ($n = 139$), Mexico ($n = 72$), Spain ($n = 99$), Turkey ($n = 150$), and the United States of America ($n = 186$). We analyzed the approximate measurement invariance based on Bayesian structural equation modeling. The model did not fit the data and its measurement invariance did not hold, with all items found non-invariant across the countries. Items related to social communication and interaction (i.e., relating to people, imitation, emotional response, and verbal and nonverbal communication) displayed lower levels of cross-country non-invariance compared to items about stereotyped behaviors/sensory sensitivity (i.e., body and object use, adaptation to change, or taste, smell, and touch response). This study found that the CARS may not provide cross-culturally valid ASD assessments. Thus, cross-cultural comparisons with the CARS should consider first which items operate differently across samples of interest, since its cross-cultural measurement non-invariance could be a source of cross-cultural variability in ASD presentations. Additional studies are needed before drawing valid recommendations in relation to the cultural sensitivity of particular items.

KEYWORDS

assessment, children, cross-cultural validity, differential item functioning

INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental condition originating due to different biological and/or genetic causes and it exists worldwide (Elsabbagh et al., 2012). Much of what we know about ASD comes from high-resource areas of the world and there are scarce data about the etiology, diagnostic process, clinical picture, and treatment interventions from low- and middle-income regions (e.g., Ashwood et al., 2015; de Leeuw et al., 2020; Elsabbagh et al., 2012; Wallace et al., 2012). The overrepresentation of developed societies and the underrepresentation of low- and middle-income societies in delivering data related to ASD have created significant research and clinical practice gaps raising concerns in the current global disparity of ASD. One concern is the extent to which ASD is conceptualized and measured with available tools across different societies and the other the degree to which treatment interventions are applicable across different settings and societies if the ASD concept is perceived differently (Ashwood et al., 2015; Zaroff & Uhm, 2012).

The diagnosis of ASD is based on the same diagnostic principles across the globe (American Psychiatric Association – APA, 2013; Doernberg & Hollander, 2016; Kim et al., 2016; World Health Organization – WHO, 2018), mainly relying on clinical assessments and using different diagnostic/evaluation tools (e.g., Falkmer et al., 2013; Randall et al., 2018). However, accumulated evidence

indicates possible cultural/regional effects concerning ASD evaluations and its cross-cultural variability (e.g., de Leeuw et al., 2020; Harrison et al., 2017; Norbury & Sparks, 2013). Studies pointed on differences in ASD clinical presentation (e.g., Amr et al., 2012; Hussein et al., 2011; Magaña et al., 2012) and symptom severity (e.g., Matson et al., 2011; Matson et al., 2017; Zachor et al., 2011) across different national/cultural groups. For example, Magaña and Vanegas (2017) showed that Latino children with ASD had low levels of restrictive and repetitive behaviors, while Fombonne et al. (2012) showed that children of Mexican origin had higher social deficits compared to these from the United States of America (USA) or Germany. In addition, socioeconomic and cultural factors may affect the early detection and diagnosis of ASD (Ratto et al., 2016; Samms-Vaughan, 2014; Windham et al., 2014), with cultural diversity indicated as vital to ASD identification and treatment (Nichols et al., 2020). Nevertheless, it is questionable whether cultural/regional effects in relation to ASD evaluations are meaningful and the cross-cultural variability of ASD is an inherent characteristic of the disorder, or there are effects of the measurement itself, as a validity flaw, since we do not have enough cross-cultural validity data for tools that measure ASD.

Almost all available diagnostic/evaluation tools were developed in Western societies within selected regional/ethnic groups (e.g., Falkmer et al., 2013; Randall et al., 2018). Although many of these tools are being adapted into different languages, the translation and cultural

adaptation process itself is not always clearly outlined and often fails to include the recommended guidelines when these tools are used cross-nationally (Cascio, 2015; DuBay & Watson, 2019). Moreover, most of the diagnostic tools are impractical for worldwide use, especially in low and middle-income regions, given the high costs associated with the intensive training required, complicated implementation, and insufficient validation processes (e.g., Abubakar et al., 2016; Durkin et al., 2015). Thus, to overcome the global disparity and the significant gaps in ASD research and clinical practice (Durkin et al., 2015), sound and sustainable evaluation/diagnostic tools must be available for use in different settings across various world regions. From a psychometric point of view, the latter is only possible if evaluation methods operate the same way and underlying constructs have the same theoretical structure across two or more regional/cultural groups (Dimitrov, 2010), what is cross-cultural/regional measurement invariance as an indicator of cross-cultural validity. The trend for researchers is to replicate the theoretical construct of an assessment tool developed in one language/culture and applied it to another would guarantee its cross-cultural equivalence and suitability for cross-cultural/country comparisons (Byrne & Watkins, 2003). However, to compare and evaluate estimates from one ASD assessment tool across various regional/cultural groups, its structure should be invariant, namely that cross-cultural measurement invariance exists (Byrne & Watkins, 2003; Dimitrov, 2010; He & van de Vijver, 2012).

Studies testing aspects of cross-cultural validity of ASD tools for children has appeared recently. No single study that simultaneously tested the construct validity of any ASD tool across two or more language/cultural groups prior to 2016 (Stevanovic et al., 2017). Across previous studies, cross-cultural variability in ASD assessments and measurement invariance of various degrees are evident. For example, one study testing the predictive validity of the Autism Spectrum Quotient (AQ; Auyeung et al., 2008) in children across samples from India, Japan and the United Kingdom (UK; Carruthers et al., 2018) showed that 28 out of 50 items have acceptable discrimination properties across all three countries. Another study evaluating the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000) found item level bias according to race and/or ethnicity for three out of 10 evaluated items (i.e., unusual eye contact, stereotyped/idiosyncratic use of words or phrases, and immediate echolalia; Harrison et al., 2017). In an investigation of the cross-cultural validity of the Diagnostic and Statistical Manual of Mental Disorders 5th edition (DSM 5; APA, 2013) for the ASD diagnosis, it was documented that the DSM model was fitted equally well in a Finnish and UK sample (Mandy et al., 2014), even at the level of particular symptoms based on the developmental, diagnostic and dimensional interview (3Di; Skuse et al., 2004). In addition, there

is also very limited evidence related to cross-cultural/regional validity of ASD in adults across different cultural samples (e.g., Freeth et al., 2013; Sappok et al., 2017).

The Childhood Autism Rating Scale (CARS; Schopler et al., 1980; Schopler et al., 2010), is an observational rating scale designed to evaluate symptoms of ASD. It covers aspects of social/communication, social interaction, stereotypic behaviors, sensory abnormalities, and emotional regulation, which likely fall under the two categories of the DSM-5 conceptualization (Park & Kim, 2016). Accumulated evidence indicates that the CARS produces reliable and valid assessments of ASD symptomatology in research and clinical settings across regions with different income levels worldwide (e.g., Breidbord & Croudace, 2013; García-López & Narbona, 2014; Magyar & Pandolfi, 2007; Mayes et al., 2014; Moulton et al., 2019; Park & Kim, 2016; Russell et al., 2010; Samms-Vaughan et al., 2017; Santos et al., 2012). Two recent reviews confirmed sound psychometric properties of the CARS, identifying moderate levels of specificity as a standalone, ASD diagnostic tool (Moon et al., 2019; Randall et al., 2018). Good psychometric properties, simple implementation in various settings, inexpensive use, and brief training required have made the CARS potentially a suitable and sustainable evaluation/diagnostic tool to overcome socio-economic and geographic gaps in ASD identification, but especially in low-resource regions (e.g., Samms-Vaughan et al., 2017). Nevertheless, it is still unclear to which extent ASD is defined and consequently measured in the same way across different societies and world regions with this tool, since data on its cross-cultural measurement invariance are lacking. This is of particular relevance because demonstrating levels of cross-cultural validity could inform on levels of cross-cultural variability in ASD possibly originating due to the measurements with the CARS.

The Autism Spectrum Disorder International Consortium (ASDIC; Stevanovic, 2018) organized this study with the aim to evaluate whether children with ASD from different language/culture groups across the globe were rated similarly with the CARS. Thus, the study tested cross-cultural measurement invariance by examining the CARS among children diagnosed with ASD aggregated from studies conducted in six countries; India, Jamaica, Mexico, Spain, Turkey, and the USA. Considering the results of the mentioned studies with other tools and data for cultural/regional variations in ASD evaluations, minor differences in rating children with the CARS were assumed to exist between the countries, thus testing approximate measurement invariance was considered (Muthén & Asparouhov, 2012, 2013). Approximate measurement invariance may provide more encouraging results for the usefulness of one scale for cross-cultural research with more groups, and it is also a recommended approach when small differences between groups are expected (Ciecuch et al., 2014; Muthén & Asparouhov, 2012).

METHODS

Participants

This is a retrospective study and data were aggregated for children who participated in previous studies and from gray literature (i.e., unpublished/clinical data by the authors). Overall, CARS data were available for children diagnosed with ASD from the following countries: India ($n = 101$ from Chakraborty et al., 2015 and from Juneja et al., 2014), Jamaica ($n = 139$ from Samms-Vaughan et al., 2017), Mexico ($n = 72$ from Flores-Rodriguez & Albores-Gallo, 2016), Spain ($n = 99$ from García-López & Narbona, 2014), Turkey ($n = 150$ from Yaylaci & Miral, 2017), and the USA ($n = 186$ from Chlebowski et al., 2013 and unpublished data from Diana L. Robins; Table 1). The income-economy level of the included countries was low-middle for India, upper-middle for Jamaica, Mexico, and Turkey, and high for Spain and the USA. Across these studies, expert clinicians based on the DSM-IV TR, DSM-5, or ICD-10 criteria assign the final ASD diagnosis. The CARS was implemented as a part of the diagnostic process or as a part of the research instrumentation either as a primary or secondary outcome measure. Many studies were missing data related to socio-economic status, race/ethnic background, neuropsychiatric comorbidities, and/or levels of intellectual and adaptive functioning. The reader is referred to the original studies for details on participants' selection and evaluation.

Childhood Autism Rating Scale

The CARS (Schopler et al., 1980) was developed as an observational rating scale for evaluating different ASD symptoms by professionals with experience in ASD through a parent/caregiver interview, the individual's observation, a case history review, and/or a combination of these sources. Its successor, the CARS-2, was developed to include aspects relevant to individuals with high-functioning levels of ASD as a separate version, while for those under aged 6 years and those above this age with low intellectual functioning and/or notably impaired communication, the original CARS version remained

(Schopler et al., 2010). The CARS measures the degree of abnormality considering the individual's chronological age, but also the peculiarity, frequency, intensity, and duration of behaviors in 14 items as the domains generally affected in ASD, plus one general category of impression of severity of ASD symptoms (Schopler et al., 1980; Schopler et al., 2010). The 15 items are: relating to people; imitation; emotional response; body use; object use; adaptation to change; visual response; listening response; taste, smell, and touch response and use; fear or nervousness; verbal communication; nonverbal communication; activity level; level and consistency of intellectual response; and general impressions. The items are not specifically predefined to be asked literally by an interviewer. Still, each has its own definition and considerations, which serve as cues for obtaining as much as possible information to score that item during a semi-structured interview. For example, item Relating to People is defined as "a rating of how the child behaves in a variety of situations involving interaction with other people," with structured and unstructured situations considered to determine how that child interacts with, reacts to, and responds to others. Each item is assigned a score of 1–4, where 1 indicates behavior appropriate for age level, while 4 indicates severe deviance with respect to normal behavior for the age level. The midpoints between them (1.5, 2.5, and 3.5) are used when the behavior appears to fall between two categories. The total raw CARS score is a sum of all rated items and ranges from 15 to 60, where higher scores indicate more severe levels of abnormal ASD behaviors. For all details related to its use, scoring, and interpretations see Schopler et al. (2010). Accumulated evidence indicates strong psychometric properties of the CARS for various aspects of reliability and validity (for details see Moon et al., 2019; Randall et al., 2018).

Across the studies from which the data were used for the present study, the CARS was administered by trained/licensed clinicians/researchers or experienced clinicians in ASD and trained for the CARS use as an interview with parents/primary caregivers (Flores-Rodriguez & Albores-Gallo, 2016; García-López & Narbona, 2014) or combining interviews with parents/primary caregivers and

TABLE 1 Main characteristics of the samples included

| Country ^a , <i>n</i> | Boys, <i>n</i> (%) | Age in years, mean (SD); range | ASD diagnosis | CARS language | CARS score, M (SD); range |
|---------------------------------|--------------------|--------------------------------|---------------------|-------------------------------|---------------------------|
| India, 101 | 84 (83.2) | 6.12 (2.84); 2–16 | ICD-10 or DSM-IV TR | Hindi | 42.02 (6.53); 25–56 |
| Jamaica, 139 | 119 (85.6) | 3.79 (1.41); 2–8 | DSM-IV TR | English | 39.01 (3.98); 30–52 |
| Mexico, 72 | 65 (90.3) | 6.23 (3.03); 2–14 | DSM-IV TR | Spanish | 39.79 (6.01); 24–52 |
| Spain, 99 | 86 (86.9) | 9.03 (3.47); 3–18 | DSM-IV TR | Spanish | 36.23 (8.93); 23–60 |
| Turkey, 150 | 116 (77.3) | 8.22 (3.73); 3–15 | DSM-5 | Turkish | 40.53 (7.89); 23–56 |
| USA, 186 | 147 (79.0) | 1.86 (0.55); 1–5 | DSM-IV TR | English, Spanish ^b | 34.73 (5.60); 21–51 |

Note: DSM-IV TR (APA, 2000); DSM-5 (APA, 2013); ICD-10 (WHO, 1992).

^aIncome-economy level: India – low-middle; Jamaica, Mexico, and Turkey – upper-middle; Spain and USA – high.

^bEnglish in 180 and English/Spanish combined in 6 cases.

direct observations of children (Chakraborty et al., 2015; Chlebowska et al., 2013; Juneja et al., 2014; Samms-Vaughan et al., 2017; Yaylaci & Miral, 2017). The original CARS (Schopler et al., 1980) was used in all studies and it was completed in the following languages: English or English/Spanish combined in the USA (in 180 and six cases, respectively; Chlebowska et al., 2013; Robins unpublished data), English in Jamaica (Samms-Vaughan et al., 2017), Spanish in Spain (García-López & Narbona, 2014) and Mexico (Flores-Rodríguez & Albores-Gallo, 2016), Turkish in Turkey (Yaylaci & Miral, 2017), and Hindi in India (Chakraborty et al., 2015; Juneja et al., 2014). Cronbach's alpha for the internal consistency of the CARS used in the respective studies was 0.89 for the data from India, 0.60 for Jamaica, 0.82 for Mexico, 0.93 for Spain, 0.89 for Turkey, and 0.79 for the USA. In the original studies, no analyses were considered on specific aspects of the reliability and validity of the CARS.

Statistical analysis

The main approach to the data analysis was based on testing for approximate measurement invariance following Bayesian structural equation modeling (Muthén & Asparouhov, 2012, 2013). Nevertheless, it was first tested the exact measurement invariance of the model represented by one underlying factor and 15 items as observable variables using multi-group confirmatory factor analysis (MG-CFA), where factor loadings and/or intercepts are gradually constrained to be exactly equal across the groups. The testing begun with the least restrictive model (i.e., configural invariance), followed by sequentially introduced cross-group equality constraints on the factor loading (i.e., metric invariance) and intercept (i.e., scalar invariance). Absolute model fit to the data was evaluated using the Tucker–Lewis index (TLI), comparative fit index (CFI), and root mean square error of approximation (RMSEA) with the following cut-off points: TLI and CFI ≥ 0.90 , RMSEA ≤ 0.08 as adequate; TLI and CFI ≥ 0.95 , RMSEA ≤ 0.06 as good fit (Brown, 2006).

Afterwards, as suggested by Muthén and Asparouhov (2013), approximate measurement invariance was implemented by estimating the model assuming approximate measurement invariance for all parameters of interest (i.e., factor loading and intercept). This method has been recently introduced to assess testing measurement invariance of an instrument when there are many groups to compare like those in cross-cultural studies (Muthén & Asparouhov, 2013). In the exact methods of measurement invariance testing, such as MG-CFA, factor loadings and intercepts are constrained to be exactly equal across groups, namely the differences in the parameters across the groups should be exactly equal to zero, which is a very restrictive assumption. On the contrary, in approximate measurement invariance, less restrictive assumptions regarding parameter differences are considered

in approximate measurement invariance, where differences are assumed to be approximately zero, but not exactly. Due to the fact that this small amount of variability is rather random, the normal distribution with mean of zero and small variance is considered for the differences in factor loadings and intercepts, which is known as the prior distribution in a Bayesian framework. Bayesian inference has three parts: the prior distribution, the data, and the likelihood. Together, they create the posterior distribution. The prior distribution conveys the information distribution based on the researcher's assumptions about a parameter, namely the assumptions on the possible values for the estimated parameter, while the posterior mean conveys mean, mode, or median, which can serve as a point estimate, and the posterior standard deviation, which serve as an indication of precision. The results of simulation studies showed that small variances such as 0.01 or 0.05 ensure that the differences are ignorable and the construct of interest remains approximately comparable across the groups (Muthén & Asparouhov, 2013; van de Schoot et al., 2013). Following the outline of Asparouhov et al. (2015), we began with a very small variance (0.001); if the model did not fit well, we slightly increased the variance that would lead to an eventually acceptable model fit. In this study, we considered values 0.001, 0.01, 0.05, and 0.1 for the variance of the prior distribution for the differences among loadings and intercepts of across the six countries. The fit of the Bayesian model was assessed based on the posterior predictive probability (PPP) values and the confidence interval (CI) between the observed and replicated chi-square values. When the PPP is higher than zero and CI contains zero, the Bayesian model fits well. If the model does not fit well, the non-invariant items could be detected, determined as the difference of a particular parameter (i.e., factor loading or/and intercept) at a specific country from the average of estimates for that particular parameter across all six countries. If a difference of zero was outside of the 95% CI of the posterior distribution of differences, the difference was assumed to be significant and the item could be considered to be non-invariant (Seddig & Leitgöb, 2018). The Mplus was used to conduct Bayesian approximate measurement invariance (Muthén & Muthén, 1998-2012).

RESULTS

In the MG-CFA (i.e., the exact measurement invariance), the model did not fit sufficiently the data across the six countries, since the values for the fit indices were below acceptable thresholds: configural (TLI = 0.85; CFI = 0.87, and RMSEA = 0.08), metric (TLI = 0.82; CFI = 0.83, and RMSEA = 0.09), and scalar invariance (TLI = 0.62; CFI = 0.59, and RMSEA = 0.13).

Table 2 shows the value of fit indices for approximate measurement invariance for the prior variance of 0.1, which had the greatest value of the variance reported. All values of the variance of the prior distribution and the

PPP values were zero, while the CIs did not include zero. Thus, the models with the smallest prior variance (i.e., 0.001) to the largest (i.e., 0.1) did not fit to the data and approximate scalar measurement invariance did not hold, too. Although in our analyses, the values of 0.2, 0.25, and 0.5 were also considered, the results were not discussed here because increasing the values of prior variance did not lead to an improvement of the model fit and the results were comparable to those reported. The full results of approximate measurement invariance are provided in a Supporting Information file Online for four values of prior variance mentioned in Table 2. For the smaller values of the prior variance, similar results were obtained and they are available on request.

Table 3 presents the deviations of the factor loadings and/or intercepts of the items from the defined priors (mean = 0 variance = 0.1). In at least one country, the estimated posterior parameters of interest (i.e., factor loading and/or intercept) deviated substantially from the average posterior estimates across all countries. The deviations of the factor loadings only were evident for

items relating to people and listening response, the deviations of the intercepts only for items emotional response, body use, fear or nervousness, and activity level, while the deviations of both parameters for items imitation, object, adaptation to change, visual response, taste, smell, and touch response, verbal communication, non-verbal communication, level and consistency of intellectual response, and general impressions. Items relating to people, body use, and listening response deviated for only one of the six countries, items verbal and nonverbal communication, activity level, and general impression for two, and the others for three to five countries. Within each country, the number of items deviating from the estimates varied from six to nine.

DISCUSSION

The results of our study showed that the CARS was lacking cross-cultural validity since its measurement invariance across six countries was not demonstrated. Considering the main assumption of approximate measurement invariance that small differences between groups are expected, even increasing the variance in the factor loadings and intercepts did not lead to an acceptable model fit to the data in our study. This implies that the CARS structure was not stable (i.e., non-invariant) across these regionally and linguistically diverse samples drawn from six countries. However, the source of the non-invariance was not limited to any particular item(s), since the deviations of the factor loadings and/or intercepts of all items varied across the countries in the models

TABLE 2 Fit indices of Bayesian approximate measurement invariance

| | Prior/PPP | 95% CI |
|-------------------|----------------|-------------------|
| Scalar invariance | N (0, 0.001)/0 | (946.04, 1127.03) |
| Scalar invariance | N (0, 0.01)/0 | (454.48, 644.72) |
| Scalar invariance | N (0, 0.05)/0 | (336.98, 522.93) |
| Scalar invariance | N (0, 0.1)/0 | (333.05, 518.67) |

Note: PPP – posterior predictive probability value; CI – confidence interval.

TABLE 3 Deviation of the factor loadings and intercepts of the CARS items from the prior defined parameters

| CARS item | India | | Turkey | | Mexico | | Spain | | Jamaica | | USA | |
|--|-------|----|--------|----|--------|----|-------|----|---------|----|-----|----|
| | FL | IN | FL | IN | FL | IN | FL | IN | FL | IN | FL | IN |
| 1. Relating to People | | | | | ▲ | | | | | | | |
| 2. Imitation | ▲ | ▲ | | ▲ | | | | | | | ▲ | ▲ |
| 3. Emotional Response | | | | ▲ | | | | ▲ | | | | ▲ |
| 4. Body Use | | | | | | | | ▲ | | | | |
| 5. Object Use | ▲ | ▲ | | ▲ | | ▲ | | | | ▲ | | |
| 6. Adaptation to Change | | | ▲ | ▲ | | ▲ | | ▲ | | ▲ | | |
| 7. Visual Response | | | | ▲ | | ▲ | | | | ▲ | | ▲ |
| 8. Listening Response | | | ▲ | | | | | | | | | |
| 9. Taste, Smell, and Touch Response and Use | ▲ | | | | ▲ | | | | | ▲ | | ▲ |
| 10. Fear or Nervousness | | ▲ | | ▲ | | | | ▲ | | ▲ | | ▲ |
| 11. Verbal Communication | | | | | ▲ | | | ▲ | | | | |
| 12. Nonverbal Communication | | | | ▲ | | | | | ▲ | ▲ | | |
| 13. Activity Level | | | | | | ▲ | | | | | | ▲ |
| 14. Level and Consistency of Intellectual Response | ▲ | | ▲ | ▲ | | | ▲ | | ▲ | | ▲ | |
| 15. General Impressions | | ▲ | | | ▲ | | | | | | | |

Note: FL – factor loading; IN – intercept; The sign ▲ indicates on the deviation of a given parameter in a given country from the defined priors (mean = 0, variance = 0.1); a non-invariant item.

tested. For four items, there were evident deviations of the intercepts only (i.e., emotional response, body use, fear or nervousness, and activity level), which may suggest that there were some responding patterns involved not related to the latent construct. For the rest of the items, with evident deviations of the factor loadings, it is suggested the unrelatedness of these items to the presumed underlying factor in one or more countries (Fischer & Karl, 2019). Collectively, the least non-invariant items were those whose measurement parameters differed for one country only, like item relating to people for Mexico, body use for Spain, or listening response for Turkey. On the other pole are items object use, adaptation to change, taste, smell, and touch response/use, fear or nervousness, and level and consistency of intellectual response, which were found the most non-invariant since their measurement parameters differed across four or more countries. In addition, comparing Mexico and Spain due to the Spanish version use of the CARS, there were only three invariant items, or Jamaica and the USA due to the English version used, there were five invariant items. Finally, the total number of non-invariant items per country also varied from six to nine, indicating that even within one country not all items exhibit similar levels of invariance. Taken together, these findings imply that CARS was not operating equally across these six countries, but its items displayed different levels of cross-cultural validity in measuring ASD symptoms.

There are several possible explanations for our findings. On the one hand considering the deviations of the intercepts (Fischer & Karl, 2019), raters using CARS items may differently report and evaluate intended ASD symptoms across regionally and linguistically diverse samples, due to factors related to the child rated, his/her family, and the context and evaluation itself, which all may not necessarily be related to one's regional/linguistic/cultural group. On the other hand, there may be genuine differences in expressing ASD symptoms at various levels among children across cultures/nations (de Leeuw et al., 2020), reflected in rating with the CARS. There are studies across different national/cultural groups pointing to significant variations in overall ASD clinical presentation (e.g., Amr et al., 2012; Hussein et al., 2011; Magaña & Vanegas, 2017) and symptom severity (e.g., Matson et al., 2011; Matson et al., 2017; Zachor et al., 2011). In addition, considering the deviations of the factor loadings, there might be true variability in assessing ASD symptoms with the CARS among professionals across nations, reflecting differences in the ASD conceptualization, which may be attributable to factors such as cultural rules, practices, language, and other systems within the cultural formulation of ASD (APA, 2013; de Leeuw et al., 2020). Similar to other tools of psychopathology including ASD (e.g., Stevanović, 2021; Stevanović et al., 2015), the CARS items may be differently sensitive to one culture than another in measuring ASD constructs

or that they are easily confounded by the culture-specific attributes related to the construct. Thus, the items might not represent specific symptoms the same way and at the same severity level or there might be some items deemed less clinically meaningful in the context of culture-specific and reference norms (e.g., Heine et al., 2002).

Our findings add significantly to the scarce literature available on the cross-cultural measurement invariance of tools for ASD. Harrison et al. (2017) using the ADOS with children and adolescents found seven non-biased items according to race and/or ethnicity in the USA, namely facial expression directed to others, quality of social overtures, quality of social response, overall quality of rapport, unusual sensory interest in play material/person, hand and finger and other complex mannerisms, and self-injurious behavior. Three items were found biased (i.e., unusual eye contact, stereotyped/idiosyncratic use of words or phrases, and immediate echolalia), of which the two later measures restricted and repetitive behaviors. Carruthers et al. (2018) found five items of the AQ (Auyeung et al., 2008) measuring social situations and communication to have almost identical discrimination power across the Indian, Japanese, and the UK sample tested, which could be an indirect measure on the cross-cultural invariance. Matson et al. (2017) used the Baby Infant Screen for Children with aUtism Traits (BISCUIT; Matson et al., 2007) in toddlers diagnosed with ASD from Greece, Italy, Japan, Poland, and the United States; they showed significant differences in overall ASD symptom severity and endorsement between these national groups. However, they observed that patterns of endorsement were fairly consistent across those items related to socialization and communication, with more variation among those related to restricted and repetitive behaviors. Nevertheless, Mandy et al. (2014) showed that ASD symptoms based on the DSM-5 (i.e., non-verbal interaction, peer relationships, sharing, socio-emotional reciprocity, non-verbal communication, conversational abilities, unusual preoccupations, routines and rituals, stereotyped and repetitive behavior, preoccupation with parts of objects, sensory abnormalities, and stereotyped and repetitive language) have sound cross-cultural measurement invariance across Finish and UK children with ASD. In relation to these, our findings indicate that some CARS items conceptually related to ASD symptoms' social communication and interaction (Moulton et al., 2019; Park & Kim, 2016) are less non-invariant (e.g., relating to people; listening response; verbal communication), which may not be the case with items related to other ASD behaviors (e.g., object use, adaptation to change, taste, smell, and touch response/use). Thus, taking together all previous studies and ours, it seems that items measuring ASD behaviors related to socialization, interaction, and communication may be universally present, less likely variable, and less likely culturally sensitive, while restricted, repetitive behaviors and sensory processing are less universally present and probably

more variable among children with ASD when assessed across regionally and linguistically diverse samples. This is supported by the previous observation (Matson et al., 2017) that cultural/ethnic background, which could include the language too, may have a larger influence on how raters perceive sensory, restricted and repetitive behaviors than on the symptoms related to communication and interaction in ASD.

The present study has several potential limitations that should be taken into consideration when interpreting the results. First, we used already available data and not data simultaneously collected for country samples. In this regard, we could not consider and control for major sociodemographic and clinical variables of the participants and their caregivers/parents, like the children's adaptive abilities, intellectual functioning, and ASD severity. In addition, it is possible that some age specific ASD characteristics were present across all samples affecting the rater's variability. The characteristics of participating centers in the original studies and the type and time of recruitment were not considered. Not simultaneously collected data, especially in different times, could also be affected by new knowledge originating. Second, we could not evaluate the appropriateness and quality of the conducting assessments with the CARS in other than English, since this was not available from the studies, although requested from the authors providing the data. Third, although the CARS was administered by trained/licensed clinicians/researchers or experienced clinicians in ASD and trained for the CARS in the primary studies, its inter-rater reliability was not possible to assess. This is an important aspect since finding measurement non-invariance could be also due to a measurement error. Fourth, the study did not consider CARS scores derived from the items weighted differently or different groups of similar items summated together, yielding different results for its measurement invariance. Finally, the measurement invariance of the same language from different regions or the different languages from the same region was not assessed due to a small number of participants. Thus, aside from the possibility that the observed findings in some CARS items could be due to inherent cultural differences and that the concepts of these items are *not equally relevant*, the difference in items' interpretation could also result from the language in which the instrument is used, the educational levels of the raters, due to the genuine differences in the mental and social abilities and functioning of children evaluated, among others that may affect and/or underly the responding to the items.

In summary, this study found that the CARS may not provide cross-culturally valid assessments of ASD, since it was not supported its measurement invariance across the data sampled from six countries. However, its items displayed different levels of cross-cultural invariance and those related to ASD features of social communication and interaction (i.e., relating to people, imitation, emotional response, and verbal and nonverbal communication) might be more cross-culturally valid compared to

items related to stereotyped behaviors and sensory sensitivity (i.e., body and object use, adaptation to change, or taste, smell, and touch response). Even though this is the only study, it is indicative for now that every comparison across two or more regional and/or linguistically diverse groups with the CARS should consider first its cross-cultural measurement invariance, since this could affect ASD assessments and might be a source of cross-cultural variability in ASD assessments. Future studies should consider testing the CARS cross-cultural measurement properties prospectively, simultaneously across regionally and linguistically diverse samples, and pairing regions/countries with same/similar language. Addressing the above limitations of this study are next logical steps, since a prerequisite for establishing measurement invariance of an assessment scale is that the theoretical construct is measured in each culture in the same way and tested simultaneously across cultural groups (He & van de Vijver, 2012). In addition, future cross-cultural research with the CARS should consider the cross-cultural reflection on the nosology of ASD, especially testing how the items relate to symptom domain structure as described in current diagnostic manuals for ASD (de Leeuw et al., 2020). On one hand, it would be essential to evaluate how and to which extent ASD is conceptualized with the CARS taking different approaches like exploratory interviews with professionals and people with autism. On the other, testing different CARS scores in measurement invariance studies derived from the items weighted differently or different groups of items based on specific conceptualizations of ASD, especially items related to social communication and repetitive/restrictive behaviors.

ACKNOWLEDGMENT

We would like to thank Dr. Tony Charman, who provided data related to the M-CHAT to our project. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTERESTS

Lilia Albores-Gallo receives remuneration for training in the assessment of autism with the CRIDI interview. Dr. Diana Robins is a co-owner of M-CHAT, LLC. M-CHAT, LLC licenses use of our intellectual property, the Modified Checklist for Autism in Toddlers (M-CHAT), for use in commercial products. We collect royalties from licensees. Dr. Robins has a 50% share in the LLC. She also is on the advisory board for Quadrant Biosciences, Inc. All other authors have no conflicts of interest.

ETHICS STATEMENT

Community members were not involved in this study. The study was approved by the Ethics committee of the Clinic for Neurology and Psychiatry for children and Youths Belgrade, Serbia.

ORCID

Dejan Stevanovic  <https://orcid.org/0000-0001-8236-5246>

Lilia Albores-Gallo  <https://orcid.org/0000-0001-5862-4404>

Cristina García-López  <https://orcid.org/0000-0002-5472-8158>

Gabriel Gatica-Bahamonde  <https://orcid.org/0000-0001-6906-6540>

Maia Gabunia  <https://orcid.org/0000-0002-4019-1302>

Sayed Ali Samadi  <https://orcid.org/0000-0002-7965-9857>

Windham Gayle  <https://orcid.org/0000-0002-3882-8343>

Alejandra Auza  <https://orcid.org/0000-0002-7592-4625>

Rajna Knez  <https://orcid.org/0000-0003-1278-4554>

REFERENCES

- Abubakar, A., Sewanyana, D., & Newton, C. R. (2016). A systematic review of research on autism spectrum disorders in sub-Saharan Africa. *Behavioural Neurology*, 2016, 3501910.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed, Text Revision ed.). American Psychiatric Publishing.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Publishing.
- Amr, M., Ali, W. B., Hablas, H., Raddad, D., El-Mehesh, F., El-Gilany, A. H., & Al-Shamy, H. (2012). Sociodemographic factors in Arab children with autism spectrum disorders. *Pan African Medical Journal*, 13, 65.
- 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.
- Asparouhov, T., Muthén, B., & Morin, A. J. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeier et al. *Journal of Management*, 41(6), 1561–1577.
- Auyeung, B., Baron-Cohen, S., Wheelwright, S., & Allison, C. (2008). The autism spectrum quotient: Children's version (AQ-Child). *Journal of Autism and Developmental Disorders*, 38(7), 1230–1240.
- Breidbord, J., & Croudace, T. J. (2013). Reliability generalization for childhood autism rating scale. *Journal of Autism and Developmental Disorders*, 43(12), 2855–2865.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford Press.
- Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology*, 34, 155–175.
- Carruthers, S., Kinnaird, E., Rudra, A., Smith, P., Allison, C., Auyeung, B., Chakrabarti, B., Wakabayashi, A., Baron-Cohen, S., Bakolis, I., & Hoekstra, R. A. (2018). A cross-cultural study of autistic traits across India, Japan and the UK. *Molecular Autism*, 9(1), 52.
- Cascio, M. A. (2015). Cross-Cultural Autism Studies, Neurodiversity, and Conceptualizations of Autism. *Culture, Medicine and Psychiatry*, 39(2), 207.
- Chakraborty, S., Thomas, P., Bhatia, T., Nimgaonkar, V. L., & Deshpande, S. N. (2015). Assessment of severity of autism using the Indian scale for assessment of autism. *Indian Journal of Psychological Medicine*, 37(2), 169–174.
- Chlebowski, C., Robins, D. L., Barton, M. L., & Fein, D. (2013). Large-scale use of the modified checklist for autism in low-risk toddlers. *Pediatrics*, 131(4), e1121–e1127.
- Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., & Schwartz, S. H. (2014). Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: A cross-country illustration with a scale to measure 19 human values. *Frontiers in Psychology*, 5, 982.
- 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.
- Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, 43(1), 121–149.
- Doernberg, E., & Hollander, E. (2016). Neurodevelopmental Disorders (ASD and ADHD): DSM-5, ICD-10, and ICD-11. *CNS Spectrums*, 21(4), 295–299.
- DuBay, M., & Watson, L. R. (2019). Translation and cultural adaptation of parent-report developmental assessments: Improving rigor in methodology. *Research in Autism Spectrum Disorders*, 62(1), 55–65.
- Durkin, M. S., Elsabbagh, M., Barbaro, J., Gladstone, M., Happe, F., Hoekstra, R. A., Lee, L.-C., 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.
- Elsabbagh, M., Divan, G., Koh, Y. J., Kim, Y. S., Kauchali, S., Marcin, 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.
- 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.
- Fischer, R., & Karl, J. A. (2019). A primer to (cross-cultural) multi-group invariance testing possibilities in R. *Frontiers in Psychology*, 10, 1507.
- Flores-Rodriguez, Y., & Albores-Gallo, L. (2016). Validity of Mexican CARS Version. *Journal of the American Academy of Child & Adolescent Psychiatry*, 10(55), S110.
- Fombonne, E., Marcin, C., Bruno, R., Tinoco, C. M., & Marquez, C. D. (2012). Screening for Autism in Mexico. *Autism Research*, 5(3), 180–189.
- Freeth, M., Sheppard, E., Ramachandran, R., & Milne, E. (2013). A cross-cultural comparison of autistic traits in the UK, India and Malaysia. *Journal of Autism and Developmental Disorders*, 43(11), 2569–2583.
- García-López, C., & Narbona, J. (2014). Clinical usefulness of IDEA and CARS: Concordance with DSM-IV-TR in children and adolescents with suspicion of PDD. *Anales de pediatria*, 80(2), 71–76.
- Harrison, A. J., Long, K. A., Tommet, D. C., & Jones, R. N. (2017). Examining the role of race, ethnicity, and gender on social and behavioral ratings within the Autism Diagnostic Observation Schedule. *Journal of Autism and Developmental Disorders*, 47(9), 2770–2782.
- He, J., & van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online readings in psychology and culture*, 2(2), 8.
- Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of Personality and Social Psychology*, 82(6), 903–918.
- 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, 34.
- Juneja, M., Mishra, D., Russell, P. S., Gulati, S., Deshmukh, V., Tudu, P., Sagar, R., Silberberg, D., Bhutani, V. K., Pinto, J. M., Durkin, M., Pandey, R. M., Nair, M. C. K., Arora, N. K., & Inclen Study Group. (2014). INCLen diagnostic tool for autism

- spectrum disorder (INDT-ASD): Development and validation. *Indian Pediatrics*, 51(5), 359–365.
- Kim, S. H., Kim, Y. S., Koh, Y. J., Lim, E. C., Kim, S. J., & Leventhal, B. L. (2016). Often asked but rarely answered: Can Asians meet DSM-5/ICD-10 autism spectrum disorder criteria? *Journal of Child and Adolescent Psychopharmacology*, 26(9), 835–842.
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205–223.
- 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.
- 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.
- 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.
- Mandy, W., Charman, T., Puura, K., & Skuse, D. (2014). Investigating the cross-cultural validity of DSM-5 autism spectrum disorder: Evidence from Finnish and UK samples. *Autism*, 18(1), 45–54.
- Matson, J. L., Boisjoli, J., & Wilkins, J. (2007). *Baby and infant screen for children with aUtism traits (BISCUIT)*. Disability Consultants, LLC.
- Matson, J. L., Matheis, M., Burns, C. O., Esposito, G., Venuti, P., Pisula, E., Misiak, A., Kalyva, E., Tsakiris, V., Kamio, Y., & Ishitobi, M. (2017). Examining cross-cultural differences in autism spectrum disorder: A multinational comparison from Greece, Italy, Japan, Poland, and the United States. *European Psychiatry*, 42(1), 70–76.
- Matson, J. L., Worley, J. A., Fodstad, J. C., Chung, K. M., Suh, D., Jhin, H. K., Ben-Itzhake, E., Zachor, D. A., & Furniss, F. (2011). A multinational study examining the cross-cultural differences in reported symptoms of autism spectrum disorders: Israel, South Korea, the United Kingdom, and The United States of America. *Research in Autism Spectrum Disorders*, 5(4), 1598–1604.
- 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.
- Moon, S. J., Hwang, J. S., Shin, A. L., Kim, J. Y., Bae, S. M., Sheehy-Knight, J., & Kim, J. W. (2019). Accuracy of the Childhood Autism Rating Scale: A systematic review and meta-analysis. *Developmental medicine & child neurology*, 61(9), 1030–1038.
- 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.
- Muthén, B., & Asparouhov, T. (2012). Bayesian SEM: a more flexible representation of substantive theory. *Psychological Methods*, 17, 313–335.
- Muthén, B., & Asparouhov, T. (2013). BSEM measurement invariance analysis. Mplus Web Notes: No.17. www.statmodel.com.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide* (7th ed.). Muthén and 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.
- Norbury, C. F., & Sparks, A. (2013). Difference or disorder? Cultural issues in understanding neurodevelopmental disorders. *Developmental Psychology*, 49(1), 45–58.
- Park, E. Y., & Kim, J. (2016). Factor structure of the Childhood Autism Rating Scale as per DSM-5. *Pediatrics International*, 58(2), 139–145.
- Randall, M., Egberts, K. J., Samtani, A., Scholten, R. J., Hoof, 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*, 24(7), CD009044.
- 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.
- 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.
- 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.
- Samms-Vaughan, M. E. (2014). The status of early identification and early intervention in autism spectrum disorders in lower-and middle-income countries. *International Journal of Speech-Language Pathology*, 16(1), 30–35.
- Santos, T. H. F., Barbosa, M. R. P., Pimentel, A. G. L., Lacerda, C. A., Balestro, J. I., Amato, C. A. D. L. H., & Fernandes, F. D. M. (2012). Comparing the use of the Childhood Autism Rating Scale and the Autism Behavior Checklist protocols to identify and characterize autistic individuals. *Jornal da sociedade Brasileira de fonoaudiologia*, 24(1), 104–106.
- Sappok, T., Brooks, W., Heinrich, M., McCarthy, J., & Underwood, L. (2017). Cross-cultural validity of the social communication questionnaire for adults with intellectual developmental disorder. *Journal of Autism and Developmental Disorders*, 47(2), 393–404.
- Schopler, E., Reichler, R. J., DeVellis, R. F., & Daly, K. (1980). Toward objective classification of childhood autism: Childhood autism rating scale (CARS). *Journal of Autism and Developmental Disorders*, 10(1), 91–103.
- Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). *The childhood autism rating scale* (2nd ed.). Western Psychological Services.
- Seddig, D., & Leitgöb, H. (2018). Approximate measurement invariance and longitudinal confirmatory factor analysis: concept and application with panel data. *Survey Research Methods*, 12(1), 29–41.
- Skuse, D., Warrington, R., Bishop, D., Chowdhury, U., Lau, J., Mandy, W., & Place, M. (2004). The developmental, dimensional and diagnostic interview (3di): a novel computerized assessment for autism spectrum disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 43(5), 548–558.
- Stevanovic, D. (2018). Autism Spectrum Disorder International Consortium (ASDIC): Moving towards sustainable and quality of life-improving services for autism in low and middle income regions. *IACAPAP Bulletin*, 51, 90–91.
- Stevanović, D. (2021). Quantitative Checklist for Autism in Toddlers (Q-CHAT): A psychometric study with Serbian Toddlers. *Research in Autism Spectrum Disorders*, 83, 101760.
- Stevanovic, D., Jafari, P., Knez, R., Franic, T., Atilola, O., Davidovic, N., Bagheri, Z., & Lacic, A. (2017). Can we really use available scales for child and adolescent psychopathology across cultures? A systematic review of cross-cultural measurement invariance data. *Transcultural Psychiatry*, 54(1), 125–152.
- Stevanovic, D., Urbán, R., Atilola, O., Vostanis, P., Balhara, Y. S., Avicenna, M., Kandemir, H., Knez, R., & Petrov, P. (2015). Does the Strengths and Difficulties Questionnaire-self report yield invariant measurements across different nations? Data from the International Child Mental Health Study Group. *Epidemiology and Psychiatric Sciences*, 24(4), 323–334.

- van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthe'n, B. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology, 4*, 1–15.
- 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.
- 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.
- World Health Organization (WHO). (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. World Health Organization.
- World Health Organization (WHO). (2018). International statistical classification of diseases and related health problems (11th Revision; ICD 11). <https://icd.who.int/browse11/l-m/en>.
- Yaylaci, F., & Miral, S. (2017). A comparison of DSM-IV-TR and DSM-5 diagnostic classifications in the clinical diagnosis of autistic spectrum disorder. *Journal of Autism and Developmental Disorders, 47*(1), 101–109.
- Zachor, D., Yang, J. W., Itzhak, E. B., 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.
- Zaroff, C. M., & Uhm, S. Y. (2012). Prevalence of autism spectrum disorders and influence of country of measurement and ethnicity. *Social Psychiatry and Psychiatric Epidemiology, 47*(3), 395–398.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Stevanovic, D., Costanzo, F., Fucà, E., Valeri, G., Vicari, S., Robins, D. L., Samms-Vaughan, M., Ozek Erkurun, H., Yaylaci, F., Deshpande, S. N., Deshmukh, V., Arora, N. K., Albores-Gallo, L., Garcia-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>