

# Can we really use available scales for child and adolescent psychopathology across cultures? A systematic review of cross-cultural measurement invariance data

**Dejan Stevanovic**

Clinic for Neurology and Psychiatry for Children and Youth

**Peyman Jafari**

Shiraz University of Medical Sciences

**Rajna Knez**

Clinical Hospital Centre Rijeka

**Tomislav Franic**

University of Split

**Olayinka Atilola**

Lagos State University College of Medicine Ikeja

**Nikolina Davidovic**

University of Split

**Zahra Bagheri**

Shiraz University of Medical Sciences

**Aneta Lakic**

Clinic for Neurology and Psychiatry for Children and Youth

## Abstract

In this systematic review, we assessed available evidence for cross-cultural measurement invariance of assessment scales for child and adolescent psychopathology as an indicator

---

## Corresponding author:

Dejan Stevanovic, Clinic for Neurology and Psychiatry for Children and Youth, Dr. Subotic 6 Belgrade, 11000 Serbia.

Email: [dejanstevanovic@eunet.rs](mailto:dejanstevanovic@eunet.rs)

of cross-cultural validity. A literature search was conducted using the Medline, PsychInfo, Scopus, Web of Science, and Google Scholar databases. Cross-cultural measurement invariance data was available for 26 scales. Based on the aggregation of the evidence from the studies under review, none of the evaluated scales have strong evidence for cross-cultural validity and suitability for cross-cultural comparison. A few of the studies showed a moderate level of measurement invariance for some scales (such as the Fear Survey Schedule for Children-Revised, Multidimensional Anxiety Scale for Children, Revised Child Anxiety and Depression Scale, Revised Children's Manifest Anxiety Scale, Mood and Feelings Questionnaire, and Disruptive Behavior Rating Scale), which may make them suitable in cross-cultural comparative studies. The remainder of the scales either showed weak or outright lack of measurement invariance. This review showed only limited testing for measurement invariance across cultural groups of scales for pediatric psychopathology, with evidence of cross-cultural validity for only a few scales. This study also revealed a need to improve practices of statistical analysis reporting in testing measurement invariance. Implications for future research are discussed.

### **Keywords**

cross-cultural comparison, cross-cultural validity, measurement invariance

### **Introduction**

Childhood psychopathology is increasingly recognized as an important issue in global childhood morbidity (Palfrey, Tonniges, Green, & Richmond, 2005) due to the high and increasing contribution of psychopathology to disease burden among children (Simpson, Bloom, Cohen, Blumberg, & Bourdon, 2005; Smit et al., 2009). This observation has spurred a deluge of epidemiological research establishing the prevalence and characteristics of childhood psychopathology across the globe (e.g., Merikangas, Nakamura, & Kessler, 2009) and has increased global attention to child and adolescent mental health (CAMH) initiatives (World Health Organization, 2003). Attention to childhood psychopathology has led to the development of different assessment scales. The most recent purposive review located 103 published scales for assessing childhood psychopathology (Verhulst & van der Ende, 2006). The reliability and validity of these scales has been examined and many are currently used in research and clinical settings across the world.

Cross-cultural differences in childhood psychopathology continue to pose a challenge to the use of these scales in cross-cultural studies given that the prevalence rates and characteristics of childhood psychopathology differ across cultural/ethnic groups (e.g., Achenbach, Rescorla, & Ivanova, 2012; Canino & Alegria, 2008). For example, in a recent study using the Strengths and Difficulties Questionnaire (SDQ), a 2.8-fold difference in the rates of general psychopathology was observed among adolescents across several countries (Atilola, Balhara, Stevanovic, Avicenna, & Kandemir, 2013). There are many potential sources for

differences across cultures when a quantitative scale is used among children. These include: inherent cross-cultural/ethnic differences due to economic, social, and cultural factors (e.g., Camras & Fatani, 2006; Hackett & Hackett, 1999; Lehman, Chiu, & Schaller, 2004; Mabe & Josephson, 2004; Nikapota & Rutter, 2008); variations in evaluation methods used; variations in the level of child development; and differences in the expression of specific psychopathology (e.g., Achenbach et al., 2012; Goodman et al., 2012; Heiervang, Goodman, & Goodman, 2008).

More distinctly, differences in prevalence rates and psychopathological expressions might be imposed by the theoretical construct of the assessment method used (i.e., construct validity). The assessment method may not necessarily operate in the same way and its underlying construct might not have the same theoretical structure for different cultural/ethnic groups (i.e., lack of measurement invariance), leading to biased estimations (Borsboom, 2006; Dimitrov, 2010). The prevailing assumption among researchers using health assessment scales is that if the theoretical construct (i.e., underlying factorial structure) of a scale developed in one language is replicated across different language groups, this will guarantee that the scale will operate equivalently across these groups and as such is suitable for cross-cultural/ethnic comparisons (e.g., Byrne & Watkins, 2003). However, a prerequisite for cross-cultural/ethnic comparisons is that the theoretical construct is measured in each culture in the same way—that is, that construct equivalence is achieved for the scale representing the theoretical construct when tested simultaneously across cultural/ethnic groups (He & van de Vijver, 2012).

Therefore, in order to compare estimates by one scale across various cultures/ethnic groups, it needs to be demonstrated that its factorial structure is invariant across different ethnic/cultural groups (i.e., cross-cultural factorial invariance; Borsboom, 2006; Byrne & Watkins, 2003; Dimitrov, 2010; Gregorich, 2006; Milfont & Fisher, 2010). Establishing the cross-cultural validity of scales used in CAMH research will improve the accuracy of comparative estimation of regional burden of childhood psychopathology and the tracking of progress of interventions in multinational contexts. In addition, valid cross-cultural research has been identified as one of the gaps in global CAMH research and has been suggested as a key agenda for advancing the utility of CAMH research in multinational contexts (Atilola, 2015).

One important question that is yet to be empirically answered is: How many of the over 100 scales currently used in quantitative CAMH research have cross-cultural validity? A first step in answering this question is to determine how many of the scales have been tested for cross-cultural validity. The second step is to determine how many, of those that have been tested, have been found to have cross-cultural validity and at what level of evidence. Answering these questions can not only guide cross-cultural CAMH researchers around the world on the suitability of available scales for such research, but can also set future directions for cross-cultural validation of CAMH scales. Accordingly, in this paper, we present a systematic review of data from studies that have tested original and different language versions of available scales for pediatric psychopathology for cross-cultural validity.

## Method

### *Search strategy and study selection*

A literature search was conducted using the Medline, PsychInfo, Scopus, Web of Science, and Google Scholar databases to identify studies on cross-cultural measurement invariance. The general eligibility criteria were (a) the study sample included children and/or adolescents; (b) a scale that assessed an aspect of child and/or adolescent psychopathology was evaluated for measurement invariance across at least two ethnic/cultural groups; and (c) the study provided details on the method used and results of measurement invariance testing. The final search was conducted on September 30, 2014. The following search terms with their variations were used: scale name or psychopathological symptoms were combined with “measurement invariance” or “measurement equivalence” or “factorial invariance” or “factorial equivalence” or “differential item functioning” or “cross-ethnic” or “cross-racial” or “cross-cultural” or cross-national” or “cross-country.” The age filter (children/adolescents up to 18 years) was used during the search, while no language or publication date limitations were imposed.

Two coders (DS and PJ) extracted data from selected studies including: the scale name, age group and population, country, language version, cultural/ethnic groups evaluated, measurement invariance method used, and the main outcome for cross-cultural measurement invariance. All search results were combined into a single master database and duplicates removed. Figure 1 presents the PRISMA flow diagram for the current review (Preferred Reporting Items for Systematic Reviews and Meta-Analyses [PRISMA]; Moher, Liberati, Tetzlaff, Altman, & the PRISMA Group, 2009).

### *Definition and determination of cross-cultural validity*

Evaluation of cross-cultural factorial invariance of a scale is usually aimed at testing *measurement invariance*, that is, how the items measure the latent construct of the scale across cultural/ethnic groups, and *structural invariance*, how the latent factors are distributed and related in the separate populations (Dimitrov, 2010; Meredith & Teresi, 2006). Measurement invariance generally refers to the invariant operations of the items or the extent to which the content of each item is being perceived and interpreted in exactly the same way across different cultural/ethnic groups. In other words, the item does not exhibit differential item functioning (DIF) across the groups if taken from the perspective of item-response theory (IRT). Structural equivalence refers to the underlying theoretical structure of the measuring scale or the extent to which the latent factors are distributed and related in the same way in the separate populations (Dimitrov, 2010).

There are several methods for establishing measurement invariance or for testing DIF. The most frequently used methods are multigroup confirmatory factor analysis (MG-CFA) based on structural equation modeling (SEM), and DIF detection

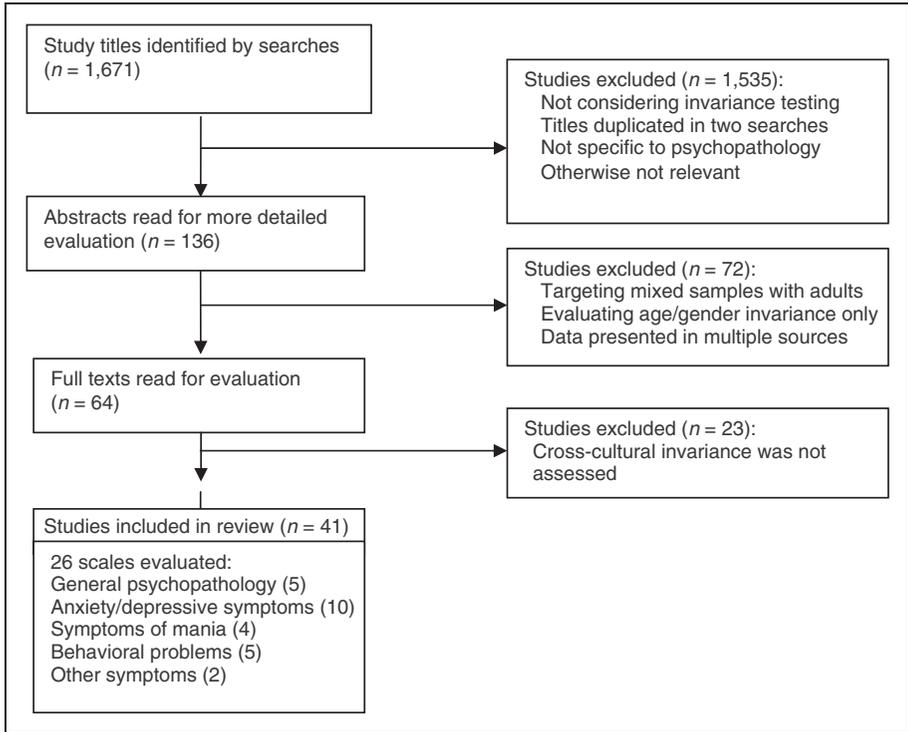


Figure 1. Flow diagram of study selection.

with IRT. Both approaches test for the equality of item-level and subscale-level true scores for persons with identical latent scores, irrespective of group membership (Raju, Laffitte, & Byrne, 2002). The third method frequently used is ordinal logistic regression (OLR) to detect DIF, which is principally based on observed scale scores and not on the latent scores as the previous two. Furthermore, it is also possible to use exploratory factor analysis to assess factorial invariance (i.e., factorial similarity) using the coefficient of congruence ( $rc$ ) and the salient variable similarity index ( $s$ ) which is derived from the relation between pairs of factor loadings for corresponding factors and denotes shared variance between factors (Cattell, 1978; Reynolds & Carson, 2005). Finally, the multiple indicators multiple causes, a method also based on SEM, is used for measurement invariance testing (MIMIC; Joreskog & Goldberger, 1975). MIMIC models can test if members belonging to different groups vary in the probability of endorsing an item after being equated on the underlying latent trait that the item is intended to measure (Joreskog & Goldberger, 1975).

In the present review, a systematic approach was applied to assess the strength of the evidence for the cross-cultural validity of CAMH scales, based on the

methods used to establish cross-cultural measurement invariance. In the case of MG-CFA used, several types of measurement invariance form a nested hierarchy, including dimensional, configural, metric, scalar, and error factorial (Byrne & Watkins, 2003; Gregorich, 2006). When testing one scale across different groups, *dimensional invariance* means that the same number of common factors is present; *configural invariance* means that the same items are associated with the same factors; *metric* (i.e., weak measurement) invariance means that the common factors have the same meaning; *scalar* (i.e., strong measurement) invariance means that the intercepts or threshold of the items are equivalent; *error* (i.e., strict measurement) invariance means that regression residual variances for all items are equal. At least strong measurement invariance must exist in order to allow latent means comparisons across groups and to claim that a scale is suitable for cross-cultural comparisons.

In situations where there is no perfect type of measurement invariance (i.e., full measurement invariance), but neither is there complete noninvariance, it is possible to talk about *partial measurement invariance* (Byrne & Watkins, 2003; Dimitrov, 2010). In the case of partial measurement invariance, only those items that meet criteria for a strong measurement invariance model should be included in composite measures when scores for the scales are to be compared cross-culturally (Gregorich, 2006). To claim that a scale is suitable for cross-cultural comparisons when considering the presence of DIF through common IRT or ORL parameterizations, there must be no or few DIF items associated with negligible individual- or group-level impact (Meredith & Teresi, 2006). According to IRT and OLR, two types of DIF, uniform and nonuniform, can be detected. Uniform DIF is evident when the difference in item response probabilities is constant across the scale. Nonuniform DIF occurs when the direction of DIF differs in different parts of the construct scale. Finally, in the case of use of the coefficient of congruence for measurement invariance testing, an  $rc$  value of 0.90 or higher is an arbitrary figure that is generally used to indicate invariance across groups (Reynolds & Carson, 2005).

### **Best evidence synthesis**

We combined results for cross-cultural measurement invariance based on the assumptions found in the selected studies in order to quantify levels of evidence for cross-cultural validity for each scale. Considering that there are no available standards to quantify levels of evidence based on measurement invariance, we followed general suggestions for measurement invariance testing (Byrne & Watkins, 2003; Dimitrov, 2010; Meredith & Teresi, 2006; Raju et al., 2002) and propositions by the Cochrane Back Review Group (Furlan, Pennick, Bombardier, van Tulder, & Editorial Board, Cochrane Back Review Group, 2009). For each scale, we reported level of evidence for cross-cultural validity as “no,” “conflicting,” “weak,” “moderate,” and “strong” evidence for use (for details of criteria

**Table 1.** Level of evidence for cross-cultural validity.

None	Non-invariance, dimensional, configural, or weak measurement invariance demonstrated or DIF evident
Conflicting*	In 2 or more samples/studies conflicting findings exist on measurement invariance
Weak	For 2 ethnic/cultural groups in 1 study demonstrated full/partial strong measurement invariance or no/few DIF evident
Moderate	For 3–4 ethnic/cultural groups in 1 study or for 2 ethnic/cultural groups in multiple studies demonstrated full/partial strong measurement invariance and/or no or few DIF evident
Strong	For 5 or more ethnic/cultural groups in 1 study or for 3–4 ethnic/cultural groups in multiple studies demonstrated full/partial strong measurement invariance and/or no or few DIF evident

\*In the case that 3 or more studies are available for one scale, but a greater majority of the studies demonstrated full/partial strong measurement invariance or no/few DIF, level of evidence would be estimated based on these studies, but graded one level below.

see Table 1). If two or more studies evaluating similar populations were available for one scale, we combined data from each study to estimate levels of evidence for that scale.

## Results

Of 1,671 studies identified, 41 studies were reported on cross-cultural measurement invariance of 26 different scales (Figure 1). From all included studies, 36 reports were available for 2–4 ethnic/racial groups from the same country using the original scale or its translation, mostly in the United States of America (USA), and 12 reports for 2–30 cross-country groups using the original scale or different language versions. Table 2 presents details of all included studies and respective scales.

For symptom-specific scales, there were moderate levels of evidence for cross-cultural validity for the Fear Survey Schedule for Children-Revised (FSSC-R) and the Multidimensional Anxiety Scale for Children (MASC) self- and parent report, Revised Children's Manifest Anxiety Scale (RCMAS) self-report, Mood and Feelings Questionnaire (MFQ) self-report, Revised Child Anxiety and Depression Scale (RCADS) self-report, and the Disruptive Behavior Rating Scale (DBRS) parent report.

Considering the Strengths and Difficulties Questionnaire (SDQ) as a scale for general psychopathology, conflicting evidence was found from six studies for measurement invariance of its self-report version, while one study reported weak

**Table 2.** Main characteristics of studies evaluating cross-cultural measurement invariance.

Scale	Reference, year	Age group, population	Version	Country, language version, ethnic/cultural groups included	Measurement invariance method*	Main measurement invariance outcome	Evidence for cross-cultural validity
<b>Anxiety symptoms</b>							
Fear Survey Schedule for Children-Revised (FSSC-R; Ollendick, 1983)	Varela, Sanchez-Sosa, Biggs, & Luis, 2008	7–16 yrs, general population	Self-, parent report	USA; original English/Spanish; Mexican/Hispanic/European American	MG-CFA	Full strong measurement invariance	Moderate
Multidimensional Anxiety Scale for Children (MASC; March, Parker, Sullivan, Stallings, & Conners, 1997)	Varela et al., 2008	7–16 yrs, general population	Self-, parent report	USA; original English/Spanish; Mexican/Hispanic/European American	MG-CFA	Full strong measurement invariance	Moderate
Revised Children's Manifest Anxiety Scale (RCMAS; Reynolds & Richmond, 2000)	Varela et al., 2008	7–16 yrs, general population	Self-, parent report	USA; original English/Spanish; Mexican/Hispanic/European American	MG-CFA	Full strong measurement invariance	Moderate
	Varela & Biggs, 2006	10–14 yrs, general population	Self-report	USA; original English; Mexican/American/European	MG-CFA	Full weak measurement invariance	
	Pina, Little, Knight, & Silverman, 2009	6–16 yrs, anxiety disorders	Self-report	USA; original English; Latino/White	MG-CFA	Full strict measurement invariance	No
Revised Children's Manifest Anxiety Scale, 2nd ed. (RCMAS-2; Reynolds & Richmond, 2008)	Ang, Lowe, & Yusof, 2011	10–15, general population	Self-report	Singapore; English original; Chinese/Malaysian / Indian	EFA	Configural invariance	

(continued)

**Table 2.** Continued

Scale	Reference, year	Age group, population	Version	Country, language version, ethnic/cultural groups included	Measurement invariance method*	Main measurement invariance outcome	Evidence for cross-cultural validity
Screen for Childhood Anxiety and Related Emotional Disorders (SCARED; Birmaher, Khetarpal, Brent, & Cully, 1997)	Gonzalez, Weersing, Warnick, Scahill, & Woolston, 2012	5–18 yrs, outpatients	Self-, parent report	USA; original English; African /Non-Hispanic White	MG-CFA	Partial weak measurement invariance	Conflicting
Spence Children's Anxiety Scale (SCAS; Spence, 1997)	Skriner & Chu, 2014	11–14 yrs, general population	Self-report	USA; original English; White/African/Hispanic/Asian	MG-CFA	Full strong measurement invariance	Weak
<b>Depressive symptoms</b>							
Children's Depression Inventory (CDI; Kovacs, 1985)	Holly, Little, Pina, & Caterino, 2014	9–12 yrs, general population	Self-report	USA; original English; Hispanic/Non-Hispanic White	MG-CFA	Full strong measurement invariance	Weak
	Wu et al., 2012	7–10 yrs, general population	Self-report	China/Italy; Chinese/Italian; Chinese/Italian	MG-CFA	Partial strict measurement invariance	Conflicting
	Steele et al., 2006	10–15 yrs, general population	Self-report	USA; English original; African/White	MG-CFA	Full weak measurement invariance	Conflicting
Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1991)	Russell, Crockett, Shen, & Lee, 2008	12–18 yrs, general population	Self-report	USA; original English; White/ Chinese/ Filipino	MG-CFA	No dimensional invariance for all groups, partial strict measurement invariance for White-Filipino	Conflicting
	Skriner & Chu, 2014	11–14 yrs, general population	Self-report	USA; original English; White/African/Hispanic/Asian	MG-CFA	Partial strong measurement invariance	

(continued)

Table 2. Continued

Scale	Reference, year	Age group, population	Version	Country, language version, ethnic/cultural groups included	Measurement invariance method <sup>a</sup>	Main measurement invariance outcome	Evidence for cross-cultural validity
	Crockett, Randall, Shen, Russell, & Driscoll, 2005	12–18 yrs, general population	Self-report	USA; original English; European/Mexican/Cuban/Puerto Rican	MG-CFA	Partial weak measurement invariance for European and Mexican. No configural invariance for Cuban and Puerto Rican	
Mood and Feelings Questionnaire (MFQ; Angold, Costello, Messer, & Pickles, 1995)	Banh et al., 2012	11–15 yrs, general population	Self-report	USA; original English; Asian/African/Hispanic/Non-Hispanic White	ORL/IRT	Few items had negligible DIF	Moderate
<b>Anxiety and depressive symptoms</b>							
Revised Child Anxiety and Depression Scale (RCADS; Chorpita, Moffitt, & Gray, 2005)	Latzman et al., 2011	8–18 yrs, general population	Self-report	USA; original English; African/White	MG-CFA	Full strong measurement invariance	Moderate**
	Trent et al., 2013	8–18 yrs, general population	Self-report	USA; English original; African/White	MG-CFA	Partial strong measurement invariance	
<b>Mania symptoms</b>							
Mood Disorders Questionnaire (MDQ; Hirschfeld et al., 2000)	McDonnell, 2010	12–18 yrs, outpatient	Parent report	USA; original English; White/African	ORL	Few items had negligible DIF	Weak
Young Mania Rating Scale (YMRS; Gracious, Youngstrom, Findling, & Calabrese, 2002)	McDonnell, 2010	12–18 yrs, outpatient	Parent report	USA; original English; White/African	ORL	Few items had negligible DIF	Weak

(continued)

**Table 2.** Continued

Scale	Reference, year	Age group, population	Version	Country, language version, ethnic/cultural groups included	Measurement invariance method*	Main measurement invariance outcome	Evidence for cross-cultural validity
Child Mania Rating Scale (CMRS; Pavuluri, Henry, Devineni, Carbray, & Birmaher, 2006)	McDonnell, 2010	12–18 yrs, outpatient	Parent report	USA; original English; White/African	ORL	Few items had negligible DIF	Weak
General Behavior Inventory (GBI; Depue, Krauss, Spoonst, & Arbisi, 1989)	McDonnell, 2010	12–18 yrs, outpatient	Parent report	USA; original English; White/African	ORL	Few items had negligible DIF	Weak
<b>Behavioral problems</b>							
Vanderbilt ADHD Teacher Rating Scale (VADTRS; Wolraich, Feurer, Hannah, Pinnock, & Baumgaertel, 1998)	Wolraich, Lambert, & Baumgaertel, 2003	Age range n/a, general population	Teacher report	Spain/USA/Germany; original English/Spanish/German; Spanish/German/urban American/suburban American	CFA	Dimensional invariance	No
ADHD-TV Rating Scale-School Version (DuPaul et al., 1997)	Reid et al., 1998	5–18 yrs, general population	Teacher report	USA; original English; White/African	MG-CFA	No configural invariance	No
Disruptive Behavior Rating Scale (DBRS; Barkley & Murphy, 1998)	Gomez, 2009	6–12 yrs, general population	Parent report	Australia/Malaysia; English/Malay; Australian/Malaysian	MG-CFA	Partial strict measurement invariance	Moderate**
	Gomez & Vance, 2008	6–12 yrs, general population	Parent report	Malaysia; English/Malay; Chinese/Malaysian	MIMIC	Few differential symptom functioning (DSF)	

(continued)

Table 2. Continued

Scale	Reference, year	Age group, population	Version	Country, language version, ethnic/cultural groups included	Measurement invariance method <sup>a</sup>	Main measurement invariance outcome	Evidence for cross-cultural validity
Behavior Problem Index (BPI; Peterson & Zill, 1986)	Guttmanova, Szanyi, & Cali, 2008	5-11 yrs, general population	Parent report	USA; original English; White/African/Hispanic	MG-CFA	Full strong measurement invariance	Weak
Eyberg Child Behavior Inventory (ECBI; Eyberg & Pincus, 1999)	Burtler, 2013	3-6 yrs, general population	Parent report	USA; original English; Non-Latino White/African	MG-CFA	No configural invariance	No
<b>General psychopathology</b>							
Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997)	van de Looij-jansen, Goedhart, de Wilde, & Treffers, 2011	11-16 yrs, general population	Self-report	The Netherlands; Dutch; Dutch/Surinamese/Antillean/Aruban/Moroccan/Turkish/Capeverdian	MG-CFA	Full strong measurement invariance	Conflicting for the self-report; weak for the parent report; strong for the teacher report
	Essau et al., 2012	12-17 yrs, general population	Self-report	Italy/Germany/Cyprus/England/Sweden; five languages: English/German/Cypriot/Greek/Swedish/Italian	MG-CFA	Configural invariance	

(continued)

Table 2. Continued.

Scale	Reference, year	Age group, population	Version	Country, language version, ethnic/cultural groups included	Measurement invariance method*	Main measurement invariance outcome	Evidence for cross-cultural validity
	Stevanovic et al., 2014	13–18 yrs, general population	Self-report	India/Indonesia/Nigeria/Serbia/Turkey/Bulgaria/Croatia; seven languages; Indian/Indonesian, Nigerian/Serbian/Turkish/Bulgarian/Croatian	MG-CFA	No dimensional invariance	
	J. P. He, Burstein, Schmitz, & Merikangas, 2013	13–18 yrs, general population	Self-report	USA; original English; White/African/Hispanic	MG-CFA	Full strong measurement invariance	
	Richter, Sagatun, Heyerdahl, Oppedal, & Røysamb, 2011	15–16 yrs, general population	Self-report	Norway; Norwegian; Norwegian/Norwegian/Pakistani/Iranian/Turkish/Somali/Vietnamese	MG-CFA	Configural invariance	
	Goodman, Patel, & Leon, 2010	5–16 yrs, general population	Self-, parent, teacher report	England; English original; White/Indian	MG-CFA	Full strict measurement invariance	
	Zwirs et al., 2011	6–10 yrs, general population	Teacher report	The Netherlands; Dutch; Dutch/Moroccan/Turkish/Surinamese	MG-CFA	Full strong measurement invariance	

(continued)

Table 2. Continued

Scale	Reference, year	Age group, population	Version	Country, language version, ethnic/cultural groups included	Measurement invariance method*	Main measurement invariance outcome	Evidence for cross-cultural validity
Achenbach System of Empirically Based Assessment (ASEBA; Achenbach, 1991; Achenbach & Rescorla, 2001)	McDonnell, 2010 Yarnell et al., 2013 Ivanova et al., 2010 Ivanova, Achenbach, Dumenci, et al., 2007 Hartman et al., 1999	12–18 yrs, outpatient 11 yrs, general population 1–5 yrs, general population 6–18 yrs, general population 4–18 yrs, general population	Parent report Parent report Parent report Parent report Parent report	USA; original English; White/African Mauritius; Kreol; Creole/Hindu/Muslim 23 different languages in 23 countries 30 different languages in 30 countries Greece/Portugal/Turkey/Norway/Netherlands/Israel/USA; seven languages; Greek/Portuguese/Turkish/Norwegian/Dutch/Israeli/American	ORL MG-CFA CFA CFA CFA	Few items had negligible DIF Configural invariance Dimensional invariance Dimensional invariance Dimensional invariance	No
Youth Self-Report (YSR)	O'Keefe, Mennen, & Lane, 2006 Lambert, Essau, Schmitt, & Samms-Vaughan, 2007	14–20 yrs, general population 11–18 yrs, general population	Self-report Self-report	USA; original English; White/African/Latino/Asian Germany/Jamaica German/Jamaican German/Jamaican	MG-CFA IRT	Weak measurement invariance No DIF is evident	Moderate

(continued)

**Table 2.** Continued.

Scale	Reference, year	Age group, population	Version	Country, language version, ethnic/cultural groups included	Measurement invariance method*	Main measurement invariance outcome	Evidence for cross-cultural validity
Teacher Report Form (TRF)	Verhulp, Stevens, van de Schoot, & Vollebergh, 2014	14 yrs, immigrant population	Self-report	The Netherlands; Dutch; Dutch/Surinamese/Turkish/Moroccan	MG-CFA	Full strong measurement invariance	
	Hartman et al., 1999	4–18 yrs, general population	Teacher report	Greece/Portugal/Turkey/Norway/Netherlands/Israel/USA; seven languages; Greek/Portuguese/Turkish/Norwegian/Dutch/Israeli/American	CFA	Dimensional invariance	No
Short Form Assessment for Children (SAC; Glisson, Hemmelgarn, & Post, 2002)	Ivanova, Achenbach, Rescorla, et al., 2007	6–15 yrs, general population	Teacher report	20 languages in 20 countries	CFA	Dimensional invariance	
	Tyson & Glisson, 2005	5–18 yrs, juvenile and family courts	Parent report	USA; English original; African/White	MG-CFA	Full strong measurement invariance	Weak

(continued)

Table 2. Continued

Scale	Reference, year	Age group, population	Version	Country, language version, ethnic/cultural groups included	Measurement invariance method*	Main measurement invariance outcome	Evidence for cross-cultural validity
Social Emotional Questionnaire (SEQ; Scholte, van Berckelaer-Onnes, & van der Ploeg, 2008)	Manti, Scholte, & van Berckelaer-Onnes, 2009	4–12 yrs, general population	Parent report	The Netherlands/Greece; Dutch/Greek	MG-CFA	Configural invariance	No
Pediatric Symptom Checklist-17 (PSC-17; Gardner et al., 1999)	Stoppelbein, Greening, Moll, Jordan, & Suozzi, 2012	6–16 yrs, general population/chronic illness	Parent report	USA; original English; African/White	MG-CFA	Partial weak measurement invariance	No
<b>Miscellaneous</b>							
Schizotypal Personality Questionnaire-Brief (SPO-B; Raine & Benishay, 1995)	Ortuño-Sierra et al., 2013	14–19 yrs, general population	Self-report	Spain/Switzerland; Spanish/Swiss; Spanish/Swiss	MG-CFA	Partial strong measurement invariance	Weak**
Youth Psychopathic Traits Inventory (YPI; Andershed, Kerr, Stattin, & Levander, 2002)	Veen et al., 2011	12–18 yrs, juvenile justice institutions	Self-report	The Netherlands; Dutch; Dutch/Moroccans	MG-CFA	Full strong measurement invariance	Weak

Note. \*MG-CFA = multigroup confirmatory factor analysis; EFA = exploratory factor analysis; CFA = confirmatory factor analysis; ORL = ordinal logistic regression; IRT = item-response theory; MIMIC = multiple indicators multiple causes. \*\*The scale has items that are cross-culturally noninvariant.

evidence for the parent report and strong evidence for the teacher report. For the Child Behavior Checklist (CBCL) and Teacher Report Form (TRF), findings from all six available studies showed no evidence of measurement invariance. There was a moderate level of evidence for the Youth Self Report (YSR).

The rest of the scales either showed weak, conflicting, or lack of cross-cultural measurement invariance (Table 2).

## Discussion

Demonstrating cross-cultural measurement invariance for a scale implies generalizability of aspects of its construct validity such that the scores of that scale generalize across different cultural/ethnic groups. This is a prerequisite for cross-cultural comparisons of psychopathology which assume that the scale measures the same theoretical construct in each culture in the same way (J. He & van de Vijver, 2012). In the present systematic review, we have evaluated the suitability of available scales for CAMH for cross-cultural comparisons based on their documented measurement invariance.

This review found that there has been limited testing for measurement invariance across cultural/ethnic groups of scales used to assess pediatric psychopathology, either in their original or translated versions. Although about 100 different scales for child and adolescent psychopathology were published before 2006 (Verhulst & van der Ende, 2006), and certainly many more since, we could locate only 26 scales with some data about cross-cultural measurement invariance. Available studies mostly evaluated scales for anxiety, depressive, mania, and behavioral symptoms (predominantly attention deficit/hyperactivity disorder), and general psychopathology, predominantly in general populations of children and adolescents. A great majority of the data available on cross-cultural measurement invariance was for the original scales developed in the USA, with two to four ethnic groups evaluated. There were only 11 scales with measurement invariance data available from other countries using different language versions, in studies evaluating either ethnic groups in one country or cultural groups across several countries.

The overall evidence suggests that few of the pediatric psychopathology scales evaluated in the present review have strong evidence for cross-cultural validity, which suggests that cross-cultural comparison of childhood psychopathology using currently available scales should be a cautious exercise. Based on the findings of the current study, moderate levels of evidence were found for the FSSC-R, MASC, RCMAS, MFQ, RCADS, and DBRS. Fortunately, these scales measure the common child and adolescent psychopathologies like anxiety, depression, and disruptive behaviors. However, the two studies evaluating the DBRS only for symptoms of attention deficit hyperactivity disorder and oppositional defiant disorder showed some DIF items and these items should not be included in the composite measures when scores for these scales are to be compared cross-culturally (Gregorich, 2006). The same applies to the RCADS, because one study

demonstrated partial strong measurement invariance. The current data for other symptom-specific scales indicate that no or weak evidence exists for their use in cross-cultural comparisons.

Considering general psychopathology scales, there was weak evidence that the Short Form Assessment for Children could be used in cross-cultural comparisons (SAC; Glisson et al., 2002). All available studies for the SDQ showed conflicting evidence for measurement invariance for its self-report, weak for the parent report, but strong for the teacher report. In particular, two well-designed studies including different language versions of the SDQ from 12 countries in total demonstrated cross-cultural measurement noninvariance of the self-report version (Essau et al., 2012; Stevanovic et al., 2014), which could be sufficient to claim that the SDQ self-report is not suitable for use in cross-cultural comparisons. Additionally, the SDQ parent and teacher report in English and Dutch were only evaluated across ethnic groups of two countries. Further testing of the two forms across several countries is needed in order to claim conclusive evidence for their use in cross-cultural comparisons. On the same note, considering the Achenbach System of Empirically Based Assessment (ASEBA; Achenbach et al., 2008), for the CBCL and TRF there was no evidence that the two could be used in cross-cultural comparisons, while there was a moderate level of evidence for the YSR. Thus, this review found that the SDQ and the ASEBA might have cross-cultural measurement noninvariance in assessing general psychopathology and their scores might not generalize across different cultural/ethnic groups or need to be considered in terms of specific population norms if such are available across different cultural/ethnic groups.

The key limitation of our review is the possibility of publication bias, we examined only articles published in English and there may be relevant data published in other languages in local journals or sources not available through the searches carried out. It was observed that authors frequently titled their articles or gave inappropriate keywords in such a way that it may not be apparent that the study assesses measurement invariance. Additionally, we could have missed studies that would be identified by a hand search and accessing grey literature. Furthermore, we used an approach to assess levels of evidence for cross-cultural comparisons that has not been used previously and it has to be further evaluated or another method should be considered.

With regard to the included studies, a great majority of the studies dealt with measurement invariance of the scales within a single country, considering only migrant ethnic groups, and thus their findings might not generalize to ethnic minority children or adolescents in their host nations/countries. In addition, data for only eight scales was available from two or more studies to assess evidence. Finally, in the reviewed papers MG-CFA was the most common method used to assess measurement invariance, but there were weaknesses in the presentation of the results of included studies. While the items in the reviewed instruments were all polytomously scored, the MG-CFA used in five studies (Russell et al., 2008; Steele et al., 2006; Trent et al., 2013; Varela & Biggs, 2006; Veen et al., 2011) assumed that observed items were continuous and normally distributed. This is a common mistake, which

could yield incorrect results in testing measurement invariance (Kim & Yoon, 2011). Moreover, when assessing measurement invariance, configural invariance should be distinguished from dimensional invariance. Dimensional invariance means that an instrument consists of the same number of factors across groups, while configural invariance shows that each domain of interest is measured by the same set of items across groups. In three of the reviewed articles (Ivanova et al., 2010; Ivanova, Achenbach, Dumenci, et al., 2007; Ivanova, Achenbach, Rescorla, et al., 2007), the term configural invariance was used mistakenly instead of dimensional invariance. Although IRT is one of the well-known DIF detection methods, our review showed that only one study (Lambert et al., 2007) specifically used IRT in testing measurement invariance. The main reason is that IRT requires two crucial assumptions including unidimensionality and local independence to estimate the model parameters. For instance, as reviewed in the present study, in testing measurement invariance of five instruments including Young Mania Rating Scale (YMRS), General Behavior Inventory (GBI), Child Mania Rating Scale (CMRS), Mood Disorders Questionnaire (MDQ), and Child Behavior Checklist (CBCL), unidimensionality and local dependence were not strictly met (McDonnell, 2010). Therefore, OLR was used as an alternative to the IRT. These issues show the extent to which the results of DIF analysis could be affected by the use of different statistical methods.

### **Conclusion: Directions for future research**

Considering the above findings and limitations, this review suggests that there is a critical need for more cross-cultural measurement invariance studies on available scales. First, studies evaluating original scales need to include three or more different ethnic/cultural groups, because there would be greater variability in the measuring construct among the groups to be detected and evaluated for cross-cultural measurement invariance if more groups were present. Second, studies evaluating one scale across several countries need to include the original as well as its translations. This is an important aspect of cross-cultural invariance testing, because demonstrating measurement invariance for the construct of the original scale in the country of its origin does not imply that its construct is transferred into its translations, thus it has to be simultaneously tested with all versions for cross-cultural measurement invariance. Third, studies should test all available rating forms for each scale for cross-cultural measurement invariance. Demonstrating cross-cultural invariance for a self-report version of a scale does not guarantee that its proxy-report is cross-culturally invariant, although there could be measurement invariance across informant reports (Dirks et al., 2014). Fourth, there is a need to further evaluate scales for which there is some evidence of invariance in order to provide more clear data on the appropriateness of their cross-cultural use, especially for the entire SDQ and ASEBA measurement systems that are the most frequently used worldwide. Fifth, it is important to evaluate scales for cross-cultural measurement invariance that measure other psychopathological symptoms frequently present in children and adolescents. Sixth, future studies should focus on clinical samples of

children with psychopathology, considering that almost exclusively the focus of available studies was on general populations. Finally, other methods besides MG-CFA should be included more frequently and studies that include methods based on IRT/ORL would be of substantial importance. Based on the fact that some scales showed partial measurement invariance, we believe that more data will be gained by IRT, because only those items that meet criteria for a strong measurement invariance model should be included in composite measures, when scores for the scales are to be compared cross-culturally (Gregorich, 2006).

In summary, this review showed that there has been limited testing for cross-cultural measurement invariance of scales for child and adolescent psychopathology and available data are insufficient to draw conclusions regarding their cross-cultural validity. Based on the evidence, a few of the scales showed a moderate level of measurement invariance (i.e., the Fear Survey Schedule for Children-Revised, Multidimensional Anxiety Scale for Children, Revised Child Anxiety and Depression Scale, Revised Children's Manifest Anxiety Scale, Mood and Feelings Questionnaire, Disruptive Behavior Rating Scale, and Youth Self-Report), which may make them suitable in cross-cultural comparative studies. Nevertheless, more replication studies are needed with available scales that will either consider different language versions or use more rigorous methods for measurement invariance testing. With more data available on cross-cultural measurement invariance and improved practices in methods used, it would be possible to revise available scales for cross-cultural use or to develop new ones.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### **References**

- Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles*. Burlington: University of Vermont, Department of Psychiatry.
- Achenbach, T. M., Becker, A., Döpfner, M., Heiervang, E., Roessner, V., Steinhausen, H. C., & Rothenberger, A. (2008). Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: Research findings, applications, and future directions. *Journal of Child Psychology and Psychiatry*, 49(3), 251–275.
- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles*. Burlington: University of Vermont, Research Center for Children, Youth, and Families.
- Achenbach, T. M., Rescorla, L. A., & Ivanova, M. Y. (2012). International epidemiology of child and adolescent psychopathology I: Diagnoses, dimensions, and conceptual issues. *Journal of the American Academy of Child and Adolescent Psychiatry*, 51, 1261–1272.

- Andershed, H., Kerr, M., Stattin, H., & Levander, S. (2002). Psychopathic traits in non-referred youths: Initial test of a new assessment tool. In E. Blaauw, J. M. Philippa, K. C. M. P. Ferenschild, & B. van Lodesteijn (Eds.), *Psychopaths: Current international perspectives* (p. 131–158). The Hague, the Netherlands: Elsevier.
- Ang, R. P., Lowe, P. A., & Yusof, N. (2011). An examination of the RCMAS-2 scores across gender, ethnic background, and age in a large Asian school sample. *Psychological Assessment, 23*(4), 899–910.
- Angold, A., Costello, E. J., Messer, S. C., & Pickles, A. (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research, 5*(4), 237–249.
- Atilola, O. (2015). Cross-cultural child and adolescent psychiatry research in developing countries. *Global Mental Health, 2*, e5. doi:10.1017/gmh.2015.8
- Atilola, O., Balhara, Y. P. S., Stevanovic, D., Avicenna, M., & Kandemir, H. (2013). Self-reported mental health problems among adolescents in developing countries: Results from an international pilot sample. *Journal of Developmental and Behavioral Pediatrics, 34*, 129–137.
- Banh, M. K., Crane, P. K., Rhew, I., Gudmundsen, G., Vander Stoep, A., Lyon, A., & McCauley, E. (2012). Measurement equivalence across racial/ethnic groups of the Mood and Feelings Questionnaire for Childhood Depression. *Journal of Abnormal Child Psychology, 40*(3), 353–367.
- Barkley, R. A., & Murphy, K. R. (1998). *Attention-deficit hyperactivity disorder: A clinical workbook* (2nd ed.). New York, NY: Guilford.
- Birmaher, B., Khetarpal, S., Brent, D., & Cully, M. (1997). The Screen for Child Anxiety Related Emotional Disorders (SCARED): Scale construction and psychometric characteristics. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*, 545–553.
- Borsboom, D. (2006). When does measurement invariance matter? *Medical Care, 44*, S176–S181.
- Butler, A. M. (2013). Cross-racial measurement equivalence of the Eyberg Child Behavior Inventory factors among low-income young African American and non-Latino White children. *Assessment, 20*(4), 484–495.
- Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology, 34*, 155–175.
- Camras, L. A., & Fatani, S. S. (2006). The development of emotional expressivity and the influence of culture. *International Society for the Study of Behavioural Development Newsletter, 49*, 12–15.
- Canino, G., & Alegria, M. (2008). Psychiatric diagnosis – Is it universal or relative to culture? *Journal of Child Psychology and Psychiatry, 49*, 237–250.
- Cattell, R. B. (1978). *The scientific use of factor analysis in the behavioral and life sciences*. New York, NY: Plenum Press.
- Chorpita, B. F., Moffitt, C. E., & Gray, J. (2005). Psychometric properties of the Revised Child Anxiety and Depression Scale in a clinical sample. *Behaviour Research and Therapy, 43*, 309–322.
- Crockett, L. J., Randall, B. A., Shen, Y. L., Russell, S. T., & Driscoll, A. K. (2005). Measurement equivalence of the Center for Epidemiological Studies Depression Scale for Latino and Anglo adolescents: A national study. *Journal of Consulting and Clinical Psychology, 73*(1), 47–58.

- Depue, R., Krauss, S., Spoont, M., & Arbisi, P. (1989). General behavior inventory identification of unipolar and bipolar affective conditions in a nonclinical university population. *Journal of Abnormal Psychology, 98*, 117–126.
- Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43*, 121–149.
- Dirks, M. A., Weersing, V. R., Warnick, E., Gonzalez, A., Alton, M., Dauser, C., . . . Woolston, J. (2014). Parent and youth report of youth anxiety: Evidence for measurement invariance. *Journal of Child Psychology and Psychiatry, 55*, 284–291.
- DuPaul, G. J., Power, T. J., Anastopoulos, A. D., Reid, R., McGoey, K., & Ikeda, M. (1997). Teacher ratings of ADHD symptoms: Factor structure and normative data. *Psychological Assessment, 9*, 436–444.
- Essau, C. A., Olaya, B., Anastassiou-Hadjicharalambous, X., Pauli, G., Gilvarry, C., Bray, D., . . . Ollendick, T. H. (2012). Psychometric properties of the Strength and Difficulties Questionnaire from five European countries. *International Journal of Methods in Psychiatric Research, 21*(3), 232–245.
- Eyberg, S. M., & Pincus, D. (1999). *Eyberg Child Behavior Inventory and Sutter–Eyberg Student Behavior Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Furlan, A. D., Pennick, V., Bombardier, C., van Tulder, M., & Editorial Board, Cochrane Back Review Group (2009). Updated method guidelines for systematic reviews in the Cochrane Back Review Group. *Spine, 34*, 1929–1941.
- Gardner, W., Murphy, M., Childs, G., Kelleher, E., Pagano, M., Jellinek, M., . . . Sturmer, R. (1999). The PSC-17: A brief pediatric symptom checklist with psychosocial problem subscales. A report from PROS and ASPN. *Ambulatory Child Health, 5*, 225–236.
- Glisson, C., Hemmelgarn, A., & Post, J. A. (2002). The Shortform Assessment for Children (SAC): An assessment and outcome measure for child welfare and juvenile justice. *Research on Social Work Practice, 12*, 82–106.
- Gomez, R. (2009). Invariance of parent ratings of the ADHD symptoms in Australian and Malaysian, and North European Australian and Malay Malaysia Children: A mean and covariance structures analysis approach. *Journal of Attention Disorders, 12*(5), 422–433.
- Gomez, R., & Vance, A. (2008). Parent ratings of ADHD symptoms: Differential symptom functioning across Malaysian Malay and Chinese children. *Journal of Abnormal Child Psychology, 36*(6), 955–967.
- Gonzalez, A., Weersing, V. R., Warnick, E., Scahill, L., & Woolston, J. (2012). Cross-ethnic measurement equivalence of the SCARED in an outpatient sample of African American and non-Hispanic White youths and parents. *Journal of Clinical Child & Adolescent Psychology, 41*(3), 361–369.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*, 581–586.
- Goodman, A., Heiervang, E., Flettlich-Bilyk, B., Alyahri, A., Patel, V., Mullick, M. S., . . . Goodman, R. (2012). Cross-national differences in questionnaires do not necessarily reflect comparable differences in disorder prevalence. *Social Psychiatry and Psychiatric Epidemiology, 47*, 1321–1331.
- Goodman, A., Patel, V., & Leon, D. A. (2010). Why do British Indian children have an apparent mental health advantage? *Journal of Child Psychology and Psychiatry, 51*(10), 1171–1183.

- Gracious, B., Youngstrom, E., Findling, R., & Calabrese, J. (2002). Discriminative validity of a parent version of the Young Mania Rating Scale. *Journal of American Academy of Child and Adolescent Psychiatry, 41*, 1350–1359.
- Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care, 44*, S78–94.
- Guttmanova, K., Szanyi, J. M., & Cali, P. W. (2008). Internalizing and externalizing behavior problem scores: Cross-ethnic and longitudinal measurement invariance of the Behavior Problem Index. *Educational and Psychological Measurement, 68*(4), 676–694.
- Hackett, R., & Hackett, L. (1999). Child psychiatry across cultures. *International Review of Psychiatry, 11*, 225–235.
- Hartman, C. A., Hox, J., Auerbach, J., Erol, N., Fonseca, A. C., Mellenbergh, G. J., . . . Sergeant, J. A. (1999). Syndrome dimensions of the Child Behavior Checklist and the Teacher Report Form: A critical empirical evaluation. *Journal of Child Psychology and Psychiatry, 40*(7), 1095–1116.
- He, J., & van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture, 2*. doi:10.9707/2307-0919.1111
- He, J. P., Burstein, M., Schmitz, A., & Merikangas, K. R. (2013). The Strengths and Difficulties Questionnaire (SDQ): The factor structure and scale validation in US adolescents. *Journal of Abnormal Child Psychology, 41*(4), 583–595.
- Heiervang, E., Goodman, A., & Goodman, R. (2008). The Nordic advantage in child mental health: Separating health differences from reporting style in a cross-cultural comparison of psychopathology. *Journal of Child Psychology and Psychiatry, 49*, 678–685.
- Hirschfeld, R. M. A., Williams, J. B. W., Spitzer, R. L., Calabrese, J. R., Flynn, L., Keck, P. E. Jr., . . . Zajecka, J. (2000). Development and validation of a screening instrument for bipolar spectrum disorder: The Mood Disorder Questionnaire. *American Journal of Psychiatry, 157*(11), 1873–1875.
- Holly, L. E., Little, M., Pina, A. A., & Caterino, L. C. (2014). Assessment of anxiety symptoms in school children: A cross-sex and ethnic examination. *Journal of Abnormal Child Psychology, 43*(2), 297–309.
- Ivanova, M. Y., Achenbach, T. M., Dumenci, L., Rescorla, L. A., Almqvist, F., Weintraub, S., . . . Verhulst, F. C. (2007). Testing the 8-syndrome structure of the Child Behavior Checklist in 30 societies. *Journal of Clinical Child and Adolescent Psychology, 36*(3), 405–417.
- Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Dumenci, L., Almqvist, F., Bathiche, M., . . . Verhulst, F. C. (2007). Testing the Teacher's Report Form syndromes in 20 societies. *School Psychology Review, 36*(3), 468–483.
- Ivanova, M. Y., Achenbach, T. M., Rescorla, L. A., Harder, V. S., Ang, R. P., Bilenberg, N., . . . Verhulst, F. C. (2010). Preschool psychopathology reported by parents in 23 societies: Testing the seven-syndrome model of the Child Behavior Checklist for ages 1.5–5. *Journal of the American Academy of Child & Adolescent Psychiatry, 49*(12), 1215–1224.
- Joreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of American Statistical Association, 70*, 631–639.
- Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling, 18*, 212–228.

- Kovacs, M. (1985). The Children's Depression Inventory (CDI). *Psychopharmacology Bulletin*, 21, 995–998.
- Lambert, M. C., Essau, C. A., Schmitt, N., & Samms-Vaughan, M. E. (2007). Dimensionality and psychometric invariance of the Youth Self-Report Form of the Child Behavior Checklist in cross-national settings. *Assessment*, 14(3), 231–245.
- Latzman, R. D., Naifeh, J. A., Watson, D., Vaidya, J. G., Heiden, L. J., Damon, J. D.,... Young, J. (2011). Racial differences in symptoms of anxiety and depression among three cohorts of students in the southern United States. *Psychiatry: Interpersonal & Biological Processes*, 74(4), 332–348.
- Lehman, D., Chiu, C., & Schaller, M. (2004). Psychology and culture. *Annual Review of Psychology*, 55, 689–717.
- Mabe, P. A., & Josephson, A. M. (2004). Child and adolescent psychopathology: Spiritual and religious perspectives. *Child and Adolescent Psychiatric Clinics of North America*, 13, 111–125.
- Manti, E., Scholte, E. M., & van Berckelaer-Onnes, I. A. (2009). A cross-cultural comparison of childhood developmental disorders between schoolchildren in the Netherlands and Greece. *European Journal of Special Needs Education*, 24, 437–454.
- March, J. S., Parker, J. D. A., Sullivan, K., Stallings, P., & Conners, K. (1997). The Multi-Dimensional Anxiety Scale for Children (MASC): Factor structure, reliability, and validity. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36, 554–565.
- McDonnell, M. A. (2010). *Race, gender and age effects on the assessment of bipolar disorder in youth*. Boston, MA: Northeastern University (Unpublished doctoral dissertation).
- Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, 44, S69–S77.
- Merikangas, K. R., Nakamura, E. F., & Kessler, R. C. (2009). Epidemiology of mental disorders in children and adolescents. *Dialogues in Clinical Neuroscience*, 11, 7–20.
- Milfont, T. L., & Fisher, R. (2010). Testing measurement invariance across groups: Applications for cross-cultural research. *International Journal of Psychological Research*, 3, 111–121.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. & the PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, 151(4), 264–269.
- Nikapota, A., & Rutter, M. (2008). Sociocultural/ethnic groups and psychopathology. In M. D. Rutter, V. M. Bishop, D. S. Pine, S. Scott, J. Stevenson, E. Taylor, & A. Thapar (Eds.), *Rutter's child and adolescent psychiatry* (5th., pp. 199–211). Oxford, UK: Blackwell.
- O'Keefe, M., Mennen, F., & Lane, C. J. (2006). An examination of the factor structure for the Youth Self Report on a multiethnic population. *Research on Social Work Practice*, 16(3), 315–325.
- Ollendick, T. H. (1983). Reliability and validity of the Revised Fear Survey Schedule for Children (FSSC-R). *Behaviour Research and Therapy*, 21, 395–399.
- Ortuño-Sierra, J., Badoud, D., Knecht, F., Paino, M., Eliez, S., Fonseca-Pedrero, E.,... Debbané, M. (2013). Testing measurement invariance of the Schizotypal Personality Questionnaire-Brief scores across Spanish and Swiss adolescents. *PloS One*, 8(12), e82041.

- Palfrey, J. S., Tonniges, T. F., Green, M., & Richmond, J. (2005). Introduction: Addressing the millennial morbidity—the context of community. *Pediatrics*, *115*, 1121–1123.
- Pavuluri, M. N., Henry, D. B., Devineni, B., Carbray, J. A., & Birmaher, B. (2006). Child Mania Rating Scale (CMRS): Development, reliability and validity. *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*(5), 550–560.
- Peterson, J. L., & Zill, N. (1986). Marital disruption, parent–child relationships, and behavior problems in children. *Journal of Marriage and Family*, *48*, 295–307.
- Pina, A. A., Little, M., Knight, G. P., & Silverman, W. K. (2009). Cross-ethnic measurement equivalence of the RCMAS in Latino and White youth with anxiety disorders. *Journal of Personality Assessment*, *91*(1), 58–61.
- Radloff, L. S. (1991). The use of the Center for Epidemiological Studies Depression Scale in adolescents and young adults. *Journal of Youth and Adolescence*, *20*, 149–166.
- Raine, A., & Benishay, D. (1995). The SPQ-B: A brief screening instrument for schizotypal personality disorder. *Journal of Personality Disorders*, *9*, 346–355.
- Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, *87*, 517–529.
- Reid, R., DuPaul, G. J., Power, T. J., Anastopoulos, A. D., Rogers-Adkinson, D., Noll, M. B., & Riccio, C. (1998). Assessing culturally different students for attention deficit hyperactivity disorder using behavioral rating scales. *Journal of Abnormal Child Psychology*, *26*, 187–198.
- Reynolds, C. R., & Carson, A. D. (2005). Methods for assessing cultural bias in tests. In C. L. Frisby, & C. R. Reynolds (Eds.), *Comprehensive handbook of multicultural school psychology* (pp. 795–823). Hoboken, NJ: Wiley.
- Reynolds, C. R., & Richmond, B. O. (2000). *Revised Children's Manifest Anxiety Scale: Manual*. Los Angeles, CA: Western Psychological Services.
- Reynolds, C. R., & Richmond, B. O. (2008). *Revised Children's Manifest Anxiety Scale, second edition (RCMAS-2): Manual*. Los Angeles, CA: Western Psychological Services.
- Richter, J., Sagatun, Å., Heyerdahl, S., Oppedal, B., & Røysamb, E. (2011). The Strengths and Difficulties Questionnaire (SDQ)—self-report. An analysis of its structure in a multi-ethnic urban adolescent sample. *Journal of Child Psychology and Psychiatry*, *52*(9), 1002–1011.
- Russell, S. T., Crockett, L. J., Shen, Y. L., & Lee, S. A. (2008). Cross-ethnic invariance of self-esteem and depression measures for Chinese, Filipino, and European American adolescents. *Journal of Youth and Adolescence*, *37*(1), 50–61.
- Scholte, E. M., van Berckelaer-Onnes, I. A., & van der Ploeg, J. D. (2008). A rating scale to screen symptoms of psychiatric disorders in children. *European Journal of Special Needs Education*, *23*, 47–62.
- Simpson, G. A., Bloom, B., Cohen, R. A., Blumberg, S., & Bourdon, K. H. (2005). U.S. children with emotional and behavioral difficulties: Data from the 2001, 2002, and 2003 National Health Interview Surveys. *Advance Data*, *360*, 1–13.
- Skriner, L. C., & Chu, B. C. (2014). Cross-ethnic measurement invariance of the SCARED and CES-D in a youth sample. *Psychological Assessment*, *26*(1), 332–337.
- Smit, F., Cuijpers, P., Oostenbrink, J., Batelaan, N., de Graaf, R., & Beekman, A. (2009). Costs of nine common mental disorders: Implications for curative and preventive psychiatry. *Journal of Mental Health Policy and Economics*, *9*, 193–200.

- Spence, S. H. (1997). Structure of anxiety symptoms among children: A confirmatory factor-analytic study. *Journal of Abnormal Psychology, 106*(2), 280–297.
- Steele, R. G., Little, T. D., Ilardi, S. S., Forehand, R., Brody, G. H., & Hunter, H. L. (2006). A confirmatory comparison of the factor structure of the Children's Depression Inventory between European American and African American youth. *Journal of Child and Family Studies, 15*(6), 773–788.
- Stevanovic, D., Urbán, R., Atilola, O., Vostanis, P., Singh Balhara, Y. P., Avicenna, M., . . . Petrov, P. (2014). 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, 30*, 1–12.
- Stoppelbein, L., Greening, L., Moll, G., Jordan, S., & Suozzi, A. (2012). Factor analyses of the Pediatric Symptom Checklist-17 with African-American and Caucasian pediatric populations. *Journal of Pediatric Psychology, 37*(3), 348–357.
- Trent, L. R., Buchanan, E., Ebesutani, C., Ale, C. M., Heiden, L., Hight, T., . . . Young, J. (2013). A measurement invariance examination of the Revised Child Anxiety and Depression Scale in a southern sample: Differential item functioning between African American and Caucasian youth. *Assessment, 20*, 175–187.
- Tyson, E. H., & Glisson, C. (2005). A cross-ethnic validity study of the Shortform Assessment for Children (SAC). *Research on Social Work Practice, 15*, 97–109.
- Van de Looij-Jansen, P. M., Goedhart, A. W., de Wilde, E. J., & Treffers, P. D. (2011). Confirmatory factor analysis and factorial invariance analysis of the Adolescent Self-Report Strengths and Difficulties Questionnaire: How important are method effects and minor factors? *British Journal of Clinical Psychology, 50*(2), 127–144.
- Varela, R. E., & Biggs, B. K. (2006). Reliability and validity of the Revised Children's Manifest Anxiety Scale (RCMAS) across samples of Mexican, Mexican American, and European American children: A preliminary investigation. *Anxiety, Stress & Coping, 19*(1), 67–80.
- Varela, R. E., Sanchez-Sosa, J. J., Biggs, B. K., & Luis, T. M. (2008). Anxiety symptoms and fears in Hispanic and European American children: Cross-cultural measurement equivalence. *Journal of Psychopathology and Behavioral Assessment, 30*(2), 132–145.
- Veen, V. C., Stevens, G. W., Andershed, H., Raaijmakers, Q. A., Doreleijers, T. A., & Vollebergh, W. A. (2011). Cross-ethnic generalizability of the three-factor model of psychopathy: The Youth Psychopathic Traits Inventory in an incarcerated sample of native Dutch and Moroccan immigrant boys. *International Journal of Law and Psychiatry, 34*(2), 127–130.
- Verhulp, E. E., Stevens, G. W., van de Schoot, R., & Vollebergh, W. A. (2014). Using the Youth Self-Report Internalizing Syndrome Scales among immigrant adolescents: Testing measurement invariance across groups and over time. *European Journal of Developmental Psychology, 11*(1), 102–110.
- Verhulst, F. C., & van der Ende, J. (2006). *Assessment scales in child and adolescent psychiatry*. Abingdon, UK: CRC Press.
- Wolraich, M. L., Feurer, I., Hannah, J. N., Pinnock, T. Y., & Baumgaertel, A. (1998). Obtaining systematic teacher report of disruptive behavior disorders utilizing DSM-IV. *Journal of Abnormal Child Psychology, 26*, 141–152.
- Wolraich, M. L., Lambert, E. W., & Baumgaertel, A. (2003). Teachers' screening for attention deficit/hyperactivity disorder: Comparing multinational samples on teacher ratings of ADHD. *Journal of Abnormal Child Psychology, 31*, 445–455.

- World Health Organization. (2003). *Caring for children and adolescents with mental disorders: Setting WHO directions*. Retrieved from [http://www.who.int/mental\\_health/media/en/785.pdf](http://www.who.int/mental_health/media/en/785.pdf)
- Wu, W., Lu, Y., Tan, F., Yao, S., Steca, P., Abela, J. R., & Hankin, B. L. (2012). Assessing measurement invariance of the Children's Depression Inventory in Chinese and Italian primary school student samples. *Assessment, 19*(4), 506–516.
- Yarnell, L. M., Sargeant, M. N., Prescott, C. A., Tilley, J. L., Farver, J. A. M., Mednick, S. A., . . . Luczak, S. E. (2013). Measurement invariance of internalizing and externalizing behavioral syndrome factors in a non-Western sample. *Assessment, 20*(5), 642–655.
- Zwirs, B., Burger, H., Schulpen, T., Vermulst, A. A., HiraSing, R. A., & Buitelaar, J. (2011). Teacher ratings of children's behavior problems and functional impairment across gender and ethnicity: Construct equivalence of the strengths and difficulties questionnaire. *Journal of Cross-Cultural Psychology, 42*(3), 466–481.

**Dejan Stevanovic, MD, PhD**, is a child psychiatrist and specialist in neuropsychology affiliated with the Clinic for Neurology and Psychiatry for Children and Youth, Belgrade, Serbia. His main research interests are cross-cultural psychiatry, neuropsychopharmacology, and pharmacoeconomics. His published works are focussed on psychometrics, cross-cultural scale validation, paediatric quality of life, and anxiety and depression in children and adolescents. He is a cofounder of the International Child Mental Health Study Group (ICMH-SG), a nonprofit organization for child and adolescent mental health research in undeveloped and developing countries.

**Peyman Jafari, PhD**, is Associate Professor of Biostatistics in Shiraz University of Medical Sciences, Shiraz, Iran. He is one of the principal investigators in studies on various aspects of health, particularly on studies dealing with cross-cultural issues. His published works focus on item response theory models, differential item functioning analysis in psychometric studies, and sequential analysis in clinical trials.

**Rajna Knez, MD, PhD**, is Senior Assistant at the School of Medicine, University of Rijeka, Croatia, Department of Psychiatry and Psychological Medicine and a specialist in psychiatry at the Division for Child Psychiatry, Skaraborg Hospital, Skövde, Sweden. Her published works focus on mental health of children and adolescents with physical disorders, child psychotherapy, and depression.

**Tomislav Franic, MD, PhD**, is a child and adolescent psychiatrist at Clinical Hospital Centre Split and Assistant Professor at Medical School University of Split, Croatia. His main fields of interest are suicidality in adolescence, intra-national and international differences in mental health and services, and transition from child and adolescent mental health services (CAMHS) to adult mental health services (AMHS). He is one of Principal Investigators in the Milestone Project which looks at how to improve the transition process and experiences for young people from CAMHS to AMHS across Europe. His publications focus on cross-

cultural validation of scales for childhood psychopathology, culturally nuanced school mental health, and a critical review of widely held paradigms in global mental health research.

**Olayinka Atilola, MD, FMCPsych**, is a lecturer in the Department of Behavioural Medicine at Lagos State University College of Medicine in Nigeria. Dr. Atilola studies cross-cultural mental health research and services among the Yoruba and other ethnic populations in developing countries. He is a Coprincipal Investigator at the Centre for Mental Health Research and Initiative in Lagos, Nigeria. His published works focus on cross-cultural validation of scales for childhood psychopathology, culturally nuanced school mental health, and a critical review of widely held paradigms in global mental health research which are at variance with the contextual realities in developing countries.

**Nikolina Davidović, MA in Psychology**, is a Research Assistant working on the EU-funded MILESTONE Project at the Clinical Hospital Center Split, Department of Psychiatry. The project aims to establish a robust link between child mental health services and adult mental health services to improve experiences and outcomes of transition young people with mental health difficulties are going through. Apart from mental health, her interests lay in adolescent behavior, clinical practice, and cross-cultural studies.

**Zahra Bagheri, PhD**, is Assistant Professor of Biostatistics in the Department of Biostatistics at Shiraz University of Medical Sciences, Shiraz, Iran. She is included as one of the Principal Investigators in studies on various aspects of health. Her published works focus on new statistical methods in quality of life research especially differential item functioning (DIF) analyses.

**Aneta Lakic, MD, PhD**, is a child psychiatrist and neuropsychiatrist affiliated with the Clinic for Neurology and Psychiatry for Children and Youth, Belgrade, Serbia. Her main research interests are mental health in children and adolescents, cognition, and psychopharmacology. Her published works focus on children and adolescents with autism spectrum disorders, scale validation, and forensic psychiatry. She is the President of Association for Child and Adolescent Psychiatry and Allied Professions of Serbia and Montenegro (DEAPS).