Quality of Life Research (2019) 28:829–840
https://doi.org/10.1007/s11136-018-2049-3

Assessing the measurement invariance of the RCADS-25 questionnaire across gender and child–parent dyads in the presence of multilevel data

Zahra Bagheri[1] · Mahsa Akbari[1] · Peyman Jafari[1] · Tania B. Huedo-medina[2]

Accepted: 9 November 2018 / Published online: 24 November 2018

Abstract

Purpose This study aims to assess the measurement invariance of the short version of the Revised Child Anxiety and Depression Scale (RCADS-25) across male and female adolescents along with child self-reports and parent proxy-reports.

Methods The sample consisted of 2039 adolescents and their parents from 51 public and private schools in Shiraz (southern Iran). The measurement invariance of the RCADS-25 was investigated by the multilevel multiple-indicators multiple-causes model. In addition, the reliability of the questionnaire was examined by Cronbach's alpha coefficient, and its validity was evaluated by exploratory and categorical confirmatory factor analysis.

Results Although the RCADS-25 includes five (20%) non-invariant items across male and female adolescents, their impact is negligible at the scale level. In addition, 10 out of 25 items (40%) were non-invariant across self- and proxy-reports, which did not have substantial effect at the scale level.

Conclusion Based on our findings, the instrument can be used, albeit with caution, for comparing depression and anxiety scores between male and female adolescents as well as child self-reports and parent proxy-reports.

Keywords Measurement invariance · RCADS-25 · Adolescents · Child self-reports · Parent proxy-reports · Multilevel

Introduction

In recent years, a considerable volume of literature has developed around the theme of depression and anxiety in children and adolescents. Previous research emphasises that the prevalence of depression and anxiety increases substantially during adolescence [1–3]; approximately 8–27% of children and adolescents suffer from these two devastating disorders at some point in their development [4–6]. Depression and anxiety in adolescents may result in several immediate and long-term dysfunctions, including academic under-achievement, substantial reduction in the perceived quality of life, social problems, substance abuse, suicidal ideation and even rapid propagation into adulthood [7–14]. Another worrisome aspect is the higher rate of depression and anxiety among girls as compared to boys [15].

In most clinical and research settings, recognising children and adolescents with depression/anxiety as well as estimating the severity of these two disorders is conducted based on specific questionnaires [5, 7]. However, boys and girls may have different perceptions of the items in a given questionnaire and may respond to them differently, which in turn may lead to higher scores for girls than boys and consequently a higher prevalence of depression and anxiety among girls. Therefore, from the measurement point of view, it is very important to assess the essential assumption of measurement invariance before comparing boys' and girls' depression/anxiety scores. Measurement invariance of a questionnaire means that members of different groups have similar perceptions of the meanings of the items in the questionnaire, given the same level of underlying depression/anxiety [16]. If measurement invariance is not properly established (i.e., non-invariance occurs), the cross-group comparison of depression and anxiety scores becomes ambiguous. Accordingly, whether the observed disparity in

✉ Peyman Jafari
jafarip@sums.ac.ir

[1] Department of Biostatistics, Shiraz University of Medical Sciences, Shiraz, Iran

[2] Department of Allied Health Sciences, Institute for Collaboration on Health, Intervention, and Policy, University of Connecticut, Storrs, CT, USA

depression or anxiety scores is a true difference in the underlying construct of interest or a synthetic impact of different interpretation of items by boys and girls remains unclear [16]. Earlier studies examining the measurement invariance of the depression/anxiety-related questionnaires such as the Beck Depression Inventory, the Centre for Epidemiologic Studies Depression Scale, the Youth Self-Report (YSR) and the Depressive Symptoms Scale questionnaires across boys and girls have yielded contradictory findings [17–20].

Furthermore, pioneer researchers have suggested that using parent proxy-reports of psychological questionnaires in conjunction with child self-reports can best assess youth depression and anxiety [7, 21], as parents can provide valuable information regarding the depression and anxiety symptoms of their children when the children themselves are unable to complete the questionnaire due to age, education, cognitive problems, physical limitations or the severity of an illness [21, 22]. Recently, a growing number of instruments have been developed to measure children's depression and anxiety based on both children's and parents' point of view. However, previous studies have reported poor to moderate agreements between these two informants at the scale level [21, 23–26]. These disagreements can cloud a clinicians' decision-making strategy due to diverse parent and child reports. According to an epidemiological viewpoint, different prevalence of depression and anxiety may also be estimated based on the reports of different informants. Consequently, in order to recognise whether this discrepancy is due to systematic differences in the construct of interest or a true difference in symptoms children and their parents report, it is crucial to evaluate the measurement invariance of a given questionnaire across child self- and parent proxy-reports. Although the measurement invariance of quality of life questionnaires across child self- and parent proxy-reports had been evaluated in several studies [27–32], this assessment in depression/anxiety questionnaires has been only restricted to the Screen Child Anxiety Emotional Related Disorders (SCARED) questionnaire measuring anxiety in clinically referred samples [33, 34].

The Revised Child Anxiety and Depression Scale (RCADS) is one of the main psychiatric questionnaires which includes both child self-reports and parent proxy-reports versions. It can measure anxiety and depression in children and adolescents and correspond more closely to the current Diagnostics and Statistical Manual of Mental Disorders (DSM-IV) as opposed to other depression/anxiety-related questionnaires [5, 7]. Although the measurement invariance of the RCADS has been evaluated in previous studies across children and adolescents from different cultures and age groups [10, 35], such an explanation has never been provided across genders and parent–child dyads. This highlights the need to conduct more investigation on the measurement invariance of well-known psychiatric questionnaires like RCADS across male and female adolescents as well as children and their parents.

In order to assess measurement invariance, it is of critical importance to select an adequate statistical method that can deal with the complexity of data structure in some sampling techniques. For instance, when children and adolescents are selected from various schools, clinical centres or geographic regions, the hierarchical structure of the data (e.g., students nested within schools) should be considered in measurement invariance testing. Recently, the multilevel multiple-indicators multiple-causes (ML-MIMIC) model has been proposed as a state-of-the-art statistical method to deal with this complexity of data structure [36]. The utilisation of this method, however, has been limited to only a few educational and social studies, and it has received scant attention in psychological and clinical studies. Hence, the primary aim of this research is to utilise the ML-MIMIC technique for assessing measurement invariance of the Persian version of the 25-item RCADS across boys and girls along with child self-reports and parent proxy-reports.

Methods

Participants

The school-based sample included 2039 adolescents and their parents from 51 public and private schools across four educational districts located in diverse socioeconomic areas in the city of Shiraz, southern Iran. The students were in grades 7–12, including 959 (47%) boys and 1080 (53%) girls. A two-stage cluster sampling technique was applied to collect the data from September to December 2015. In the first stage, a random sample of schools was selected from each educational district; 51 out of 56 selected schools collaborated to participate in the study. Then, one class from each grade of each school was chosen. Finally, two trained researchers distributed both the child self- and parent proxy-reports of the Persian version of the RCADS-25 in addition to the informed parental consent form to all students in each class. They explained the objective of the study and also gave instructions about how the students and their parents should complete the questionnaire. Both children and parents who intended to participate in the study filled out their questionnaires at home and returned them to school within 3 days. The total response rate was 68% (2039 out of 3000). 1000 of the students were in secondary schools (grades 7–9) and 1039 were in high schools (grades 10–12). Student distribution was 17.1%, 17.7%, 16.3%, 19.1%, 15.7% and 14.1% in grades 7, 8, 9, 10, 11 and 12, respectively. The missing percentage for all of the items of the child self-reports and parent proxy-reports versions was less than 5%. The study

was approved by the ethical committee of Shiraz University of Medical sciences.

Measure

The RCADS-25 is the short version of the RCADS-47, which includes both the child self-reports and parent proxy-reports versions. In this questionnaire, 10 items (1, 4, 8, 10, 13, 15, 16, 19, 21 and 24) measure depression and 15 items (2, 3, 5, 6, 7, 9, 11, 12, 14, 17, 18, 20, 22, 23 and 25) measure anxiety [7]. Using standard guidelines, including forward and backward translation, the English version of the questionnaire was translated into Persian with the permission of the copyright holder. The participants were asked to indicate how often each item applied to them according to a four-point Likert scale (from 0 = never to 3 = always). Total scores of anxiety and depression subscales were obtained by summing up the selected response categories; thus, higher scores represented higher levels of depression and anxiety.

Statistical analysis

The demographic characteristics of adolescents and their parents were represented using descriptive statistics such as mean, standard deviation and frequency (percentage). In addition, these characteristics were compared between boys and girls using the independent sample *T* test and Chi-square tests for quantitative and qualitative variables, respectively. The reliability of the questionnaire was assessed by Cronbach's alpha coefficient. A coefficient equal to or greater than 0.7 was considered to be a satisfactory level of reliability. Furthermore, the construct validity of the questionnaire was evaluated by exploratory and categorical confirmatory factor analysis (CCFA). The correlation matrix between the RCADS items and the two hypothesised subscales was calculated in the exploratory factor analysis to examine whether each item has an acceptable loading on the subscales of interest. According to Kline [37], the value of 0.40 was used as a factor loading criterion in the exploratory factor analysis without any cross loading of an item across different dimensions. CCFA was used to investigate the relationship between a set of observed variables (the items of the RCADS-25) and a set of continuous latent variables (depression and anxiety subscales). In the present study, we investigated whether or not the hypothesised two-factor model fit the data well. The goodness of fit of the model was investigated using a number of goodness-of-fit indices, including the root mean square error of approximation (RMSEA), Tucker–Lewis index (TLI) and comparative fit index (CFI). Values of CFI and TLI ≥ 0.90 and RMSEA ≤ 0.08 indicate an acceptable model fit [38].

In order to assess the measurement invariance of the RCADS-25 across boys and girls along with parent–child dyads, the ML-MIMIC model was applied. This model is an extension of the MIMIC model with the addition of a random effect term. MIMIC modelling is one the most commonly used methods to investigate factorial invariance by employing a grouping observed variable (x) to explain group differences. The MIMIC model includes measurement and structural components [39, 40]:

$$y_i = \Lambda\eta_i + \Gamma_y x_i + \varepsilon_i, \tag{1}$$

$$\eta_i = \Gamma_\eta x_i + \zeta i. \tag{2}$$

In the measurement component of the model (Eq. 1), the latent variable (η) (e.g., depression) is represented by the observed indicators (y_i, here the RCADS's items). In this equation, Λ is the factor loading and η_i is the common factor score. In the structural component of the model (Eq. 2), the latent variable (η_i) is regressed on the grouping variable of interest (x_i, e.g., gender). MIMIC inherently presumes that all parameters of the model including factor loading or residual variance are equal between the groups [41]. Instead, factorial invariance of intercept, which is known as uniform invariance, is investigated by introducing the effect of a grouping variable (x) on the observed outcome variable ($\Gamma_y x_i$) over and above the effect of the grouping variable on the latent factor (Γ_η) in Eq. 1. In other words, the regression coefficient of x on y (Γ_y) shows the status of intercept invariance of the corresponding variable y (item) across groups [41].

The MIMIC model is simply extended to multilevel MIMIC by introducing a random effect to the model. In other words, Eq. 1 can be extended to the following model by including subscript j to show the effect of clusters. In this study, when factorial invariance is examined across a grouping observed variable varying at the first level of data like gender and child–parent role, the MIMIC model is written as follows:

$$Y_{ij} = \nu + \Lambda_W \eta_{Wij} + \Gamma_y x_{ij} + \Lambda_B \eta_{Bj} + \varepsilon_{Wij} + \varepsilon_{Bj}, \tag{3}$$

where x_{ij} is an observed variable representing group membership of individual i (within-level) in cluster j (between-level). Γ_y indicates the non-invariance in intercepts (uniform non-invariance) across the groups, which is estimated as the fixed effect since it is assumed that factorial invariance across groups is constant in different clusters. This assumption can be realistic with a reasonably developed measure [41]. Kim et al. [41] suggested that non-uniform invariance or factor loading invariance can also be tested by including an interaction term between the latent factor and the grouping variable of interest. Hence, Model 3 can be extended as follows:

$$Y_{ij} = \nu + \Lambda_W \eta_{Wij} + \Gamma_y x_{ij} + \Gamma_{\eta y} \eta_{Wij} x_{ij} + \Lambda_B \eta_{Bj} + \varepsilon_{Wij} + \varepsilon_{Bj}. \tag{4}$$

He adopted the XWITH procedure in Mplus to create the interaction term; however, other studies showed that this procedure leads to a convergence problem and biased parameter estimates when the observed indicator is not normal [42, 43]. Due to facing the same problem of convergence in the present study, we investigated only uniform invariance. Pioneer researchers have stated that uniform non-invariance is more important than non-uniform one because for items with non-uniform non-invariance, the direction of non-invariance differs along the subscale; consequently, the effect of non-invariance cancels out at the scale level [44] and does not influence the results substantially.

Interested readers are encouraged to consult Davidov et al. and Kim et al. for more technical information related to the process, which is beyond the scope of this article [41, 45].

As uniform invariance of the questionnaire's items was examined one at a time [36, 46], the Bonferroni correction was used to consider the multiple statistical tests (with 25 items) [47].

It should be noted that in the present study, the data structure had three levels, including students nested in classes and nested in schools. However, a two-level model in which students were nested in schools was applied because the variation of the observations was negligible at the class level. Previous research showed that nested datasets do not necessarily require multilevel modelling when there is no variation across some levels of data [48]. The variation is mainly quantified by intraclass correlation (ICC) and design effect statistics. Although a non-zero value of ICC shows that there is a non-zero variation at the level of interest, it does not inevitably indicate the need for multilevel modelling. Accordingly, a design effect value of greater than 2 is introduced as a more reliable criterion [48]. In our dataset, the ICC (design effect) at school level was 0.29 (4.19) and 0.136 (3.49) for depression and anxiety scores, respectively. However, the ICC (design effect) at the class level was 0.008 (1.003) for the depression subscale and 0.015 (1.57) for the anxiety subscale which were negligible.

A distinct advantage of the MIMIC model is its ability to control the effect of confounding variables while assessing measurement invariance [40]. To investigate measurement invariance across boys and girls, the effects of parents' education and job were controlled since these two covariates differed significantly between boys and girls in our study. To examine measurement invariance across child self-reports and parent proxy-reports, the effects of age, gender and education, which were significantly different between children and their parents, were controlled.

Moreover, in order to assess whether or not the generalisability of the invariance results was affected by differences in the perceptions of lower or higher age groups of students, a sub-analysis of measurement invariance was also conducted separately in the two age groups. The first age group consisted of students in secondary schools, and the second one included high school students.

In order to compare the anxiety and depression scores of boys and girls along with self- and proxy-reports, the student *t* test was applied and the Cohen's *d* effect size was calculated. The effect size values < 0.20, 0.20–0.49, 0.50–0.79 and > 0.80 are considered negligible, small, moderate and large, respectively [49]. Moreover, to investigate the impact of non-invariance, the items with uniform non-invariance were omitted, and the corresponding subscale scores were then recomputed. SPSS 15 was applied to perform the exploratory factor analysis and t-test, and the Mplus 6.1 software was used to perform the CCFA and ML-MIMIC analyses.

Results

Table 1 represents the summary statistics of adolescents and their parents. As shown, male and female adolescents did not differ significantly in terms of age, parent's age and father's education level. However, a significant difference was observed between boys and girls in terms of mother's education, mother's job and father's job.

The results of the exploratory factor analysis with Varimax rotation are summarised in Table 2. As indicated in both child self-reports and parent proxy-reports versions, all of the items in the depression subscale had factor loadings greater than 0.4, which implies that they have been properly loaded on this subscale; while items 11, 14 and 18, which should have been loaded on the anxiety subscale, had greater factor loadings on the depression subscale in both self- and proxy-reports. Nevertheless, the results of CCFA supported the fit of the two-factor model after considering the correlation between items 3 and 9, for child self-reports (RMSEA = 0.056, CFI = 0.94, TLI = 0.93) and parent proxy-reports (RMSEA = 0.056, CFI = 0.92, TLI = 0.91). The Cronbach's alpha coefficient value of the child self-reports was 0.89 for the whole questionnaire, and it was 0.90 and 0.91 for the depression and anxiety subscales, respectively. The Cronbach's alpha coefficient of the parent proxy-reports was 0.88 for the whole questionnaire, and it was 0.93 and 0.95 for the depression and anxiety subscales, respectively.

Table 3 shows the parameter estimates, standard errors and *p* values for each item in the child self-reports of the RCADS-25 questionnaire that were obtained from the ML-MIMIC model to identify non-invariant items across boys and girls in the whole dataset as well as the subsamples of secondary and high school students, separately. It should be noted that the effect of mother's education, mother's job and father's job as confounding variables were controlled in the model. In this analysis, each item was assessed for

Table 1 Demographic characteristics of adolescents and their parents by adolescent's gender

	Girls Mean ± SD	Boys Mean ± SD	*p* value
Age	14.82 ± 1.54	14.84 ± 1.53	0.74†
Mother age	39.95 ± 5.81	40.47 ± 5.94	0.09†
Father age	45.69 ± 6.35	46.1 ± 6.18	0.31†
	N (%)	N (%)	
Mother education			
Illiterate	99 (11.4)	65 (10.4)	0.001‡
Primary	91 (10.5)	70 (11.2)	
Diploma	478 (55.3)	295 (47)	
University degree	197 (22.8)	197 (31.4)	
Father education			
Illiterate	47 (11.5)	50 (9.8)	0.24‡
Primary	75 (18.4)	79 (15.5)	
Diploma	162 (39.8)	198 (38.7)	
University degree	123 (30.2)	184 (36)	
Mother job			
Self-employment	32 (3.7)	20 (3.2)	0.002‡
Government-related employee	102 (11.8)	111 (17.9)	
Retired	11 (1.3)	16 (2.6)	
Housewife	721 (83.3)	473 (76.3)	
Father job			
Self-employment	184 (45.5)	206 (40.3)	0.01‡
Government-related employee	121 (30)	175 (34.2)	
Retired	35 (8.7)	71 (13.9)	
Others	64 (15.8)	59 (11.5)	

†*p* value based on independent *t* test

‡*p* value based on Chi-square test

invariance one at a time; therefore, 0.002 (0.05/25) was considered as the significance level rather than 0.05 based on the Bonferroni correction. As indicated, in the whole dataset, only item 15 (I cannot think clearly) was detected as uniform non-invariance across boys and girls in the depression subscale. However, in the anxiety subscale, four items—6 (I am afraid of being in crowded places), 11 (I suddenly become dizzy or faint without reason), 12 (I have to do some things over and over again) and 23 (I have to do some things in just the right way to stop bad things from happening)—were distinguished as uniform non-invariance. Furthermore, a close inspection of β coefficients in Table 3 suggested that the boys reported more problems in their answer to items 12, 15 and 23 compared to their female counterparts, as depicted by the negative values of β coefficients. The opposite pattern was observed for items 6 and 11, in which the positive values of β coefficients revealed that the boys reported fewer problems for these items as compared to the girls.

The results of sub-analysis of measurement invariance among secondary school students showed that item 6 in the anxiety subscale was distinguished as uniform non-invariance (with negative β coefficient) across boys and girls. Nevertheless, in high school students, all items were invariant across the groups.

Table 4 shows the results of the ML-MIMIC model for assessing the measurement invariance of the RCADS-25 questionnaire across child self- and parent proxy-reports, while controlling the effect of age, gender and education as confounding variables. The results were represented in the whole dataset along with the subsample of secondary and high school students. In the depression subscale, items 1 (My child feels sad or empty), 10 (My child has problems with his/her appetite) and 19 (My child feels like he/she doesn't want to move) were identified with uniform non-invariance across children and their parents in the whole dataset. While, in the anxiety subscale seven items, including 3 (My child feels afraid of being alone at home), 6 (My child is afraid of being in crowded places), 9 (My child feels scared to sleep on his/her own), 11 (My child suddenly becomes dizzy or faint when there is no reason for this), 14 (My child suddenly starts to tremble or shake when there is no reason for this), 15 (My child cannot think clearly) and 23 (My child has to

Table 2 Factor loadings of (rotated) two-factor solution in child self-reports and parent proxy-reports of RCADS-25

	Child self-report		Parent proxy-reports	
	Depression	Anxiety	Depression	Anxiety
q_1: I feel sad or empty	0.68	0.20	0.62	0.20
q_2: I worry when I think I have done poorly at something	0.22	0.49	0.07	0.48
q_3: I would feel afraid of being on my own at home	− 0.04	0.67	0.03	0.63
q_4: Nothing is much fun anymore	0.67	0.07	0.61	0.10
q_5: I worry that something awful will happen to someone in my family	0.14	0.52	0.18	0.55
q_6: I am afraid of being in crowded places	0.16	0.35	0.11	0.37
q_7: I worry what other people think of me	0.28	0.51	0.31	0.45
q_8: I have trouble sleeping	0.43	0.29	0.33	0.32
q_9: I feel scared if I have to sleep on my own	− 0.03	0.69	0.02	0.64
q_{10}: I have problems with my appetite	0.34	0.17	0.36	0.12
q_{11}: I suddenly become dizzy or faint when there is no reason for this	0.56	**0.19**	0.56	**0.18**
q_{12}: I have to do some things over and over again	0.19	0.43	0.12	0.43
q_{13}: I have no energy for things	0.63	0.11	0.68	0.13
q_{14}: I suddenly start to tremble or shake when there is no reason for this	0.51	**0.21**	0.31	**0.26**
q_{15}: I cannot think clearly	0.55	0.23	0.52	0.26
q_{16}: I feel worthless	0.63	0.28	0.63	0.23
q_{17}: I have to think of special thoughts to stop bad things from happening	0.24	0.40	0.23	0.42
q_{18}: I think about death	0.49	**0.29**	0.39	**0.28**
q_{19}: I feel like I don't want to move	0.58	0.04	0.64	0.02
q_{20}: I worry that I will suddenly get a scared feeling when there is nothing to be afraid of	0.22	0.62	0.31	0.59
q_{21}: I am tired a lot	0.71	0.14	0.71	0.12
q_{22}: I feel afraid that I will make a fool of myself in front of people	0.39	0.55	0.44	0.47
q_{23}: I have to do some things in just the right way to stop bad things from happening	0.37	0.44	0.26	0.51
q_{24}: I feel restless	0.67	0.28	0.61	0.32
q_{25}: I worry that something bad will happen to me	0.41	0.56	0.43	0.51

Extraction method: Alpha factoring with Varimax rotation

Values in bold indicate items did not load onto their designated subscales

do some things in just the right way to stop bad things from happening) were distinguished as uniform non-invariance across these two groups.

The positive value of β coefficient in items 3, 6, 9, 10, 11, 19 and 23 indicated that the parents reported more problems for their children's symptoms as opposed to their children. In contrast, the parents reported fewer problems for items 1, 14 and 18, in comparison to their children, due to the negative values of β coefficients.

Moreover, in the subsample of secondary school students items 4, 14, 18 and 23 were identified with uniform non-invariance across child self-reports and parent proxy-reports with negative β coefficient for items 14 and 18, as opposed to the positive β coefficient for items 4 and 23. In the subsample of high school students items 3, 6, 9, 10, 14, 18 and 19 were distinguished as uniform non-invariance. The β coefficient was positive for items 3, 6, 9, 10 and 19, while it was negative for items 14 and 18.

Table 5 displays the anxiety and depression scores before and after the removal of the non-invariant items. Removing non-invariant items led to a slight reduction in the depression and anxiety scores in both girls and boys. Although before and after applying the removal strategy, the girls had higher depression and anxiety scores compared with the boys; these differences were small according to the effect size values (ranging from 0.22 to 0.37). In addition, after removing the non-invariant items across children and their parents, the subscale scores were reduced in both groups; nevertheless, the significant difference between these two groups did not alter. The magnitude of the effect size also indicated that there were small differences between children and their parents before and after the removal strategy (ranging from 0.28 to 0.43).

Discussion

This study aimed to examine the measurement invariance of the RCADS-25 questionnaire across male and female adolescents along with child self- and parent proxy-reports using the ML-MIMIC model. Our findings revealed that, in the

Table 3 The results of ML-MIMIC models for assessing the measurement invariance of the RCADS across gender in the whole dataset as well as subsample of secondary and high school students

	Whole dataset		Secondary students		High school students	
	β (SE)	p value	β (SE)	p value	β (SE)	p value
Depression						
q_1	− 0.051 (0.136)	0.706	0.062 (0.224)	0.782	0.316 (0.395)	0.423
q_4	0.197 (0.103)	0.056	− 0.120 (0.185)	0.516	0.196 (0.215)	0.363
q_8	0.108 (0.095)	0.256	0.154 (0.248)	0.536	0.516 (0.205)	0.012
q_{10}	0.084 (0.127)	0.507	0.468 (0.213)	0.028	− 0.098 (0.359)	0.784
q_{13}	− 0.322 (0.182)	0.077	− 0.200 (0.305)	0.513	− 0.256 (0.384)	0.506
q_{15}	0.371 (0.109)	0.001*	0.206(0.243)	0.396	0.001(0.240)	0.995
q_{16}	− 0.023 (0.163)	0.886	0.425 (0.314)	0.175	− 0.166 (0.193)	0.388
q_{19}	− 0.143 (0.116)	0.219	− 0.217 (0.149)	0.145	0.022 (0.425)	0.959
q_{21}	− 0.188 (0.096)	0.050	− 0.376 (0.234)	0.108	− 0.495 (0.260)	0.057
q_{24}	− 0.038 (0.106)	0.717	− 0.100 (0.180)	0.579	0.069 (0.246)	0.778
Anxiety						
q_2	0.145 (0.117)	0.210	0.282 (0.146)	0.053	0.374 (0.185)	0.043
q_3	− 0.196 (0.350)	0.575	− 0.451 (0.223)	0.043	− 0.153 (0.168)	0.362
q_5	0.262 (0.104)	0.011	0.393 (0.399)	0.325	0.073 (0.217)	0.737
q_6	− 0.742 (0.121)	< 0.001*	− 1.028 (0.275)	< 0.001*	− 0.243 (0.381)	0.525
q_7	− 0.020 (0.170)	0.907	− 0.096 (0.270)	0.721	− 0.479 (0.242)	0.048
q_9	− 0.416 (0.283)	0.141	− 0.445 (0.270)	0.099	− 0.163 (0.193)	0.398
q_{11}	− 0.479 (0.157)	0.001*	− 0.304 (0.296)	0.304	− 0.504 (0.316)	0.111
q_{12}	0.564 (0.150)	< 0.001*	0.835 (0.340)	0.014	0.322 (0.187)	0.086
q_{14}	− 0.336 (0.186)	0.070	0.174 (0.302)	0.565	− 1.000 (0.317)	0.002
q_{17}	0.181 (0.132)	0.171	0.360 (0.256)	0.160	0.709 (0.234)	0.002
q_{18}	− 0.027 (0.143)	0.851	− 0.024 (0.226)	0.914	− 0.008 (0.142)	0.956
q_{20}	− 0.171 (0.132)	0.196	− 0.085 (0.302)	0.778	0.131 (0.290)	0.652
q_{22}	− 0.122 (0.129)	0.347	− 0.280 (0.230)	0.225	− 0.081 (0.298)	0.785
q_{23}	0.448 (0.106)	< 0.001*	0.398 (0.283)	0.160	0.371 (0.196)	0.059
q_{25}	0.033 (0.133)	0.805	− 0.282 (0.277)	0.308	0.286(0.162)	0.076

Girls is the reference group

*Significant p values based on Bonferroni correction

whole dataset, five out of 25 items (20%) were detected as uniform non-invariance across genders, which suggested that girls and boys had different perceptions of these items. Interestingly, the boys reported more problems for items 15 (I cannot think clearly), 12 (I have to do some things over and over again) and 23 (I have to do some things in just the right way to stop bad things from happening) compared to the girls. This discrepancy could be attributed to the differences in male and female brain processes. For instance, it has been shown that when males are deeply involved in a task they are not able to focus on their surroundings or other tasks, while females are able to transition between tasks more promptly [50]. Moreover, previous research showed that girls use the ruminative coping technique more than boys [1, 51]; hence, they may not have to do a task over and over again as boys may have to. In contrast to our findings, a number of previous studies reported that girls suffering from depression and anxiety experience more concentration problems than boys [52, 53]. This contradiction may be because the sample examined in the current study was comprised of a general population of schoolchildren and adolescents, while previous studies focused on children or adolescents who were clinically diagnosed as depressed or anxious. In addition, gender differences were investigated in their studies at the scale level and not at the item level.

Moreover, our findings revealed that, in the whole dataset, the girls reported more problems for items 11 (I suddenly become dizzy or faint when there is no reason for this) and 6 (I am afraid of being in crowded places). A possible explanation for the observed disparity between girls and boys for item 11 is that the onset of menstruation is a primary cause of numerous health problems in girls, such as fainting or a sense of dizziness [54, 55]. Furthermore, girls start dieting at an earlier age in response to beauty ideals, which in turn may result in eating disorders; hence, they may suffer from more physical health problems than their male counterparts [54]. The inconsistency in the perception of item 6 between

Table 4 The results of ML-MIMIC models for assessing the measurement invariance of the RCADS across parent–child dyads in the whole dataset as well as subsample of secondary and high school students

	Whole dataset		Secondary students		High school students	
	β (SE)	p value	β (SE)	p value	β (SE)	p value
Depression						
q_1	− 0.353 (0.081)	< 0.001*	− 0.401 (0.146)	0.006	− 0.276 (0.118)	0.019
q_4	0.252 (0.073)	0.001*	0.545 (0.103)	< 0.001*	0.072 (0.101)	0.474
q_8	− 0.134 (0.060)	0.026	− 0.295 (0.174)	0.090	− 0.103 (0.111)	0.353
q_{10}	0.393 (0.053)	< 0.001*	0.330 (0.107)	0.002	0.488 (0.068)	< 0.001*
q_{13}	0.211 (0.065)	0.001	0.223 (0.091)	0.015	0.046 (0.106)	0.664
q_{15}	− 0.181 (0.074)	0.015	− 0.194 (0.175)	0.265	− 0.304 (0.128)	0.018
q_{16}	− 0.230 (0.084)	0.006	− 0.469 (0.210)	0.026	− 0.009 (0.131)	0.947
q_{19}	0.266 (0.076)	< 0.001*	− 0.174 (0.140)	0.214	0.415 (0.109)	< 0.001*
q_{21}	0.059 (0.061)	0.336	0.221 (0.174)	0.203	0.062 (0.080)	0.439
q_{24}	− 0.034 (0.073)	0.645	0.204 (0.127)	0.108	− 0.062 (0.124)	0.620
Anxiety						
q_2	− 0.031 (0.064)	0.630	0.156 (0.097)	0.108	− 0.252 (0.103)	0.015
q_3	0.313 (0.062)	< 0.001*	0.076 (0.090)	0.398	0.494 (0.115)	< 0.001*
q_5	− 0.058 (0.046)	0.208	− 0.048 (0.088)	0.584	− 0.082 (0.077)	0.286
q_6	0.333 (0.074)	< 0.001*	0.241 (0.148)	0.105	0.476 (0.144)	0.001*
q_7	0.044 (0.054)	0.407	− 0.026 (0.083)	0.755	0.115 (0.106)	0.278
q_9	0.357 (0.053)	< 0.001*	0.273 (0.131)	0.037	0.376 (0.090)	< 0.001*
q_{11}	0.190 (0.053)	< 0.001*	0.274 (0.095)	0.004	0.209 (0.109)	0.055
q_{12}	− 0.094 (0.072)	0.193	0.041 (0.135)	0.761	0.004 (0.125)	0.976
q_{14}	− 1.058 (0.084)	< 0.001*	− 0.792 (0.144)	< 0.001*	− 1.073 (0.142)	< 0.001*
q_{17}	0.043 (0.064)	0.504	0.033 (0.111)	0.763	0.080 (0.113)	0.476
q_{18}	− 0.662 (0.053)	< 0.001*	− 0.616 (0.120)	< 0.001*	− 0.685 (0.072)	< 0.001*
q_{20}	− 0.023 (0.064)	0.715	− 0.145 (0.121)	0.230	− 0.019 (0.114)	0.877
q_{22}	0.182 (0.065)	0.005	0.040 (0.133)	0.764	0.334 (0.108)	0.002
q_{23}	0.327 (0.061)	< 0.001*	0.427 (0.135)	0.001*	0.129 (0.087)	0.139
q_{25}	0.123 (0.073)	0.091	0.127 (0.133)	0.326	− 0.005 (0.127)	0.967

Children is the reference group

*Significant p values based on Bonferroni correction

boys and girls may be due to a greater social sensitivity in girls as opposed to boys [56].

Another important finding was that in the whole dataset, children and their parents had different perceptions for 10 out of 25 items (40%) of the RCADS-25. In contrast, two previous studies examining measurement invariance of the SCARED questionnaire across children and their parents confirmed the measurement invariance of the questionnaire across the groups [33, 34]. This inconsistency may be due to the fact that these studies assessed the clinically referred samples of children and adolescents.

In our sample, parents reported more problems for items 3, 6, 9, 10, 11, 19 and 23. This may be due the fact that the symptoms of sleep patterns, activity or appetite are more openly observable [24–26], so parents may pay more attention to these symptoms. On the other hand, social desirability of children may partially account for the lower rating of social avoidance or their fears, as measured by items 3, 6 and 9 [23]. The present study is consistent with previous studies in that children reported more problems for internalising symptoms measured by items 1, 14 and 18 than their parents. The most probable explanation for this discrepancy between children and their parents is that because parents do not have access to the inner distress of their children, they are unable to recognise such symptoms [21, 23, 26, 57].

It is interesting to note that when the measurement invariance of the RCADS-25 across boys and girls was separately investigated in the two subsamples of secondary and high school students, only one item was detected as uniform non-invariance among secondary students. A possible explanation for this might be that boys and girls in similar age groups had more agreement on the perceptions of the items; whereas a wide age range might lead to finding more non-invariant items. Another possible explanation is that the smaller sample size in each subsample might reduce the power of the model for detecting non-invariant items. Additionally, fewer non-invariant items were detected across child–parent dyads in each of the two subsamples as

Table 5 Mean comparison of the RCADS-25 subscales between boys and girls as well as children and their parents before and after eliminating DIF items

	Girls Mean ± SD	Boys Mean ± SD	*p* value, *d**
Scores without eliminating DIF items			
Anxiety	11.38 ± 7.17	9.84 ± 6.32	< 0.001, 0.22
Depression	6.33 ± 5.13	4.65 ± 4.30	< 0.001, 0.35
Scores with eliminating DIF items			
Anxiety	9.21 ± 5.77	7.96 ± 5.11	< 0.001, 0.23
Depression	5.78 ± 4.72	4.18 ± 3.97	< 0.001, 0.37
	Children Mean ± SD	Parents Mean ± SD	*p* value, *d**
Scores without eliminating DIF items			
Anxiety	10.84 ± 6.87	8.06 ± 5.90	< 0.001, 0.43
Depression	5.70 ± 4.87	4.41 ± 4.32	< 0.001, 0.28
Scores with eliminating DIF items			
Anxiety	6.99 ± 4.45	5.24 ± 3.79	< 0.001, 0.42
Depression	4.12 ± 3.65	3.07 ± 3.17	< 0.001, 0.31

*Cohen's *d* value

compared to the whole dataset. Yet, there was more discrepancy between children's and parents' perceptions in the older age group than in the younger one (seven non-invariant items among high school students in comparison to five non-invariant items among secondary students). This can be attributed to the puberty in late adolescence, which is an influential factor in child–parent disagreement and adolescents' strive for autonomy and individuation [58].

It should be noted that, in a questionnaire, the presence of non-invariant items per se does not imply that the measure cannot be applied to cross-group comparison. Therefore, after detecting non-invariance, it is of critical importance to assess whether or not the effect of a given non-invariant item is important enough to be transferred to the scale level. In the present study, a removing and retaining technique is used to examine whether the score of the observed subscales changed significantly with and without non-invariant items. Removing the non-invariant items detected across boys and girls led to a slight reduction in the depression and anxiety scores, with a small magnitude of the effect size, implying invariance across genders. Although removing non-invariant items distinguished across children and their parents resulted in the reduction of depression and anxiety subscale scores, the small magnitude of the effect size before and after the removal strategy suggested that these differences were not substantial.

Another important finding is that girls suffered from higher levels of depression and anxiety in comparison to boys, which is in line with previous findings [3, 15, 17, 53, 59, 60]. This can be contributed to not only environmental, cultural, biological and hormonal factors, but also the sex-role orientation, interpersonal orientation, rumination and more contradictory social expectations placed on girls. In addition, higher levels of social desirability for boys lead to gender differences in terms of the presence and severity of depression and anxiety [2, 3, 51, 53, 54, 59, 61, 62].

A key strength of the current study is that the hierarchical nature of data was taken into account by applying the ML-MIMIC technique. This issue is important due to the fact that when students are nested into schools, the critical assumption of independency of observations does not hold due to some shared qualities. Overlooking this dependency while applying traditional statistical methods based on the independency assumption results in inaccurate parameter estimates and misleading conclusions about establishing measurement invariance [36, 63]. Finch and French [36] reported that when the hierarchical structure of data in measurement invariance analysis is ignored, the likelihood of finding non-invariant items decreases. Moreover, the effect of confounding variables can be taken into account in the ML-MIMIC model. It has been established that if the effect of confounding variables is not considered in the measurement invariance analysis, both the non-invariance detection procedure and subsequent group comparison may be distorted. Another positive aspect of our study is that a large sample size from 51 public and private schools, including students from diverse socioeconomic backgrounds was selected in Shiraz, one of the five biggest cities of Iran. Therefore, it can be claimed that the sample was a representative sample of adolescents in the urban area of southern Iran.

However, the generalisability of our findings is subject to certain limitations. First, the sample was selected from a general population of school adolescents; hence, the

proportion of adolescents who might be clinically diagnosed as depressed or anxious was most likely very small. Therefore, further studies need to be carried out in order to evaluate whether or not Iranian male and female adolescents with depression and anxiety disorders perceive the meaning of the RCADS-25 questionnaire in the same way. Second, the invariance of the RCADS-25 was only conducted across sex and child–parent groups; future studies should examine the measurement invariance of this instrument across different age groups (e.g., children and adolescents) and cultures as well. Another limitation is that only uniform invariance was investigated. Therefore, it would be more fruitful to use alternative methods like the multilevel ordinal logistic model or multilevel multi-group confirmatory factor analysis to test both uniform and non-uniform invariance in future studies. Finally, according to Kim et al. [41], we used the same two-factor structure model at both levels (with anxiety and depression as factors and their corresponding items), mainly because the variation at the school level was not of substantial interest in the present study but was regarded as nuisance variation. However, the results would be different if we applied another factor structure model such as the partially saturated models suggested by Ryu [64].

Conclusion

In summary, the findings of this research provide evidence that boys and girls as well as children and their parents have different perceptions and understandings of certain items in the RCADS-25 questionnaire. However, further analysis indicates that the impact of non-invariance is negligible at the scale level across boys and girls along with children and their parents. This implies that the instrument can be used with caution to compare depression and anxiety between male and female adolescents as well as child self-reports and parent proxy-reports scores. As researchers have recently made efforts to choose a questionnaire that can function identically across different groups, this research can serve as a base for future studies and provide important insights for instrument developers to refine the instrument by rewording, removing and calibrating non-invariant items to develop an invariant instrument.

Acknowledgements This work was supported by a Grant (Number 94-01-01-9208) from Shiraz University of Medical Sciences Research Council. The article was extracted from Mahsa Akbari's Master of Science thesis.

Compliance with ethical standards

Conflict of interest The authors hereby declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the ethical and research committee of our institution, Shiraz University of Medical Sciences and also with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Compas, B. E., Orosan, P. G., & Grant, K. E. (1993). Adolescent stress and coping: implications for psychopathology during adolescence. *Journal of Adolescence, 16*(3), 331–349.
2. Lewinsohn, P. M., Clarke, G. N., Seeley, J. R., & Rohde, P. (1994). Major depression in community adolescents: Age at onset, episode duration, and time to recurrence. *Journal of the American Academy of Child and Adolescent Psychiatry, 33*(6), 809–818.
3. McCauley Ohannessian, C., Lerner, R. M., Lerner, J. V., & von Eye, A. (1999). Does self-competence predict gender differences in adolescent depression and anxiety? *Journal of Adolescence, 22*(3), 397–411.
4. Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry, 60*(8), 837–844.
5. Ebesutani, C., Reise, S. P., Chorpita, B. F., Ale, C., Regan, J., Young, J., Higa-McMillan, C., & Weisz. J. R. (2012). The revised child anxiety and depression scale-short version: Scale reduction via exploratory bifactor modeling of the broad anxiety factor. *Psychological Assessment, 24*(4), 833–845.
6. Lewinsohn, P. M., Zinbarg, R., Seeley, J. R., Lewinsohn, M., & Sack, W. H. (1997). Lifetime comorbidity among anxiety disorders and between anxiety disorders and other mental disorders in adolescents. *Journal of Anxiety Disorders, 11*(4), 377–394.
7. Ebesutani, C., Tottenham, N., & Chorpita, B. (2015). The Revised Child Anxiety and Depression Scale—parent version: Extended applicability and validity for use with younger youth and children with histories of early-life caregiver neglect. *Journal of Psychopathology and Behavioral Assessment, 37*(4), 705–718.
8. Kendall, P. C., Safford, S., Flannery-Schroeder, E., & Webb, A. (2004). Child anxiety treatment: Outcomes in adolescence and impact on substance use and depression at 7.4-year follow-up. *Journal of Consulting and Clinical Psychology, 72*(2), 276–287.
9. Langley, A. K., Bergman, R. L., McCracken, J., & Piacentini, J. C. (2004). Impairment in childhood anxiety disorders: Preliminary examination of the child anxiety impact scale-parent version. *Journal of Child and Adolescent Psychopharmacology, 14*(1), 105–114.
10. Mathyssek, C. M., Olino, T. M., Hartman, C. A., Ormel, J., Verhulst, F. C., & Van Oort, F. V. (2013). Does the Revised Child Anxiety and Depression Scale (RCADS) measure anxiety symptoms consistently across adolescence? The TRAILS study. *International journal of methods in psychiatric research, 22*(1), 27–35.
11. Woodward, L. J., & Fergusson, D. M. (2001). Life course outcomes of young people with anxiety disorders in adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*(9), 1086–1093.
12. Garber, J., & Weersing, V. R. (2010). Comorbidity of anxiety and depression in youth: Implications for treatment and prevention. *Clinical Psychology, 17*(4), 293–306.

13. Rohde, P., Lewinsohn, P. M., & Seeley, J. R. (1994). Are adolescents changed by an episode of major depression? *Journal of the American Academy of Child and Adolescent Psychiatry, 33*(9), 1289–1298.
14. Weissman, M. M., Wolk, S., Wickramaratne, P., Goldstein, R. B., Adams, P., Greenwald, S., Ryan, N. D., Dahl, R. E., & Steinberg, D. (1999). Children with prepubertal-onset major depressive disorder and anxiety grown up. *Archives of General Psychiatry, 56*(9), 794–801.
15. Halbreich, U., & Kahn, L. S. (2007). Atypical depression, somatic depression and anxious depression in women: Are they gender-preferred phenotypes? *Journal of Affective Disorders, 1021*(1), 245–258.
16. Teresi, J. A., & Fleishman, J. A. (2007). Differential item functioning and health assessment. *Quality of Life Research, 16*(1), 33–42.
17. Bares, C., Andrade, F., Delva, J., Grogan-Kaylor, A., & Kamata, A. (2012). Differential item functioning due to gender between depression and anxiety items among Chilean adolescents. *The International Journal of Social Psychiatry, 58*(4), 386–392.
18. Brunet, J., Sabiston, C. M., Chaiton, M., Low, N. C. P., Contreras, G., Barnett, T. A., & O'Loughlin, J. L. (2014). Measurement invariance of the depressive symptoms scale during adolescence. *BMC Psychiatry, 14*(1), 1–10.
19. Byrne, B. M., Baron, P., & Balev, J. (1996). The Beck depression inventory: Testing for its factorial validity and invariance across gender for Bulgarian non-clinical adolescents. *Personality and Individual Differences, 21*(5), 641–651.
20. Roberts, R. E., Andrews, J. A., Lewinsohn, P. M., & Hops, H. (1990). Assessment of depression in adolescents using the Center for Epidemiologic Studies Depression scale. *Psychological Assessment, 2*(2), 122–128.
21. Choudhury, M. S., Pimentel, S. S., & Kendall, P. C. (2003). Childhood anxiety disorders: Parent–child (dis)agreement using a structured interview for the DSM-IV. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*(8), 957–964.
22. Eiser, C., & Morse, R. (2001). Can parents rate their child's health-related quality of life? Results of a systematic review. *Quality of Life Research, 10*(4), 347–357.
23. De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*(4), 483–509.
24. Grills, A. E., & Ollendick, T. H. (2003). Multiple informant agreement and the anxiety disorders interview schedule for parents and children. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*(1), 30–40.
25. Maurizi, L. K., Gershoff, E. T., & Aber, J. L. (2012). Item-level discordance in parent and adolescent reports of parenting behavior and its implications for adolescents' mental health and relationships with their parents. *Journal of Youth and Adolescence, 41*(8), 1035–1052.
26. Salbach-Andrae, H., Klinkowski, N., Lenz, K., & Lehmkuhl, U. (2009). Agreement between youth-reported and parent-reported psychopathology in a referred sample. *European Child & Adolescent Psychiatry, 18*(3), 136–143.
27. Bagheri, Z., Jafari, P., Tashakor, E., Kouhpayeh, A., & Riazi, H. (2014). Assessing whether measurement invariance of the KIDSCREEN-27 across child–parent dyad depends on the child gender: A multiple group confirmatory factor analysis. *Global Journal of Health Sciences, 6*(5), 142–153.
28. Jafari, P., Allahyari, E., Salarzadeh, M., & Bagheri, Z. (2016). Item-level informant discrepancies across obese-overweight children and their parents on the PedsQL™ 4.0 instrument: An iterative hybrid ordinal logistic regression. *Quality of Life Research, 25*(1), 25–33.
29. Jafari, P., Bagheri, Z., Hashemi, S. Z., & Shalileh, K. (2013). Assessing whether parents and children perceive the meaning of the items in the PedsQLTM 4.0 quality of life instrument consistently: A differential item functioning analysis. *Global Journal of Health Sciences, 5*(5), 80–88.
30. Jafari, P., Sharafi, Z., Bagheri, Z., & Shalileh, S. (2014). Measurement equivalence of the KINDL questionnaire across child self-reports and parent proxy-reports: A comparison between item response theory and ordinal logistic regression. *Child Psychiatry and Human Development, 45*(3), 369–376.
31. Lin, C. Y., Luh, W. M., Cheng, C. P., Yang, A. L., Su, C. T., & Ma, H. I. (2013). Measurement equivalence across child self-reports and parent proxy-reports in the Chinese version of the pediatric quality of life inventory version 4.0. *Child Psychiatry and Human Development, 44*(5), 583–590.
32. Huang, I. C., Shenkman, E. A., Leite, W., Knapp, C. A., Thompson, L. A., & Revicki, D. A. (2009). Agreement was not found in adolescents' quality of life rated by parents and adolescents. *Journal of Clinical Epidemiology, 62*(3), 337–346.
33. Dirks, M. A., Weersing, V. R., Warnick, E., Gonzalez, A., Alton, M., Dauser, C., Scahill, L., & Woolston, J. (2014). Parent and youth report of youth anxiety: Evidence for measurement invariance. *Journal of child psychology and psychiatry, 55*(3), 284–291.
34. Behrens, B., Swetlitz, C., Kircanski, K., Pine, D. S., & Pagliaccio, D. (2017). Measurement invariance of the scared questionnaire: Examining parent and child report of anxiety symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry, 56*(10), S242.
35. Trent, L. R., Buchanan, E., Ebesutani, C., Ale, C. M., Heiden, L., Hight, T. L., Damon, J. D.,… 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*(2), 175–187.
36. Finch, H., & Frenc, B. (2010). Detecting differential item functioning of a course satisfaction instrument in the presence of multilevel data. *Journal of The First-Year Experience & Students in Transition, 22*(1), 27–47.
37. Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.
38. Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233–255.
39. Joreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70*(351), 631–639.
40. Bye, B. V., Gallicchio, S. J., & Dykacz, J. M. (1985). Multiple-indicator, multiple-cause models for a single latent variable with ordinal indicators. *Sociological Methods and Research, 13*(4), 487–509.
41. Kim, E. S., Yoon, M., Wen, Y., Luo, W., & Kwok, O. (2015). Within-level group factorial invariance with multilevel data: Multilevel factor mixture and multilevel MIMIC models. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(4), 603–616.
42. Cham, H., West, S. G., Ma, Y., & Aiken, L. S. (2012). Estimating latent variable interactions with non-normal observed data: A comparison of four approaches. *Multivariate Behavioral Research, 47*(6), 840–876.
43. Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 65*(4), 457–474.
44. Shenkman, E. A. (2011). Differential item functioning in quality of life measure between children with and without special healthcare needs. *Value in Health, 14*(6), 872–883.

45. Davidov, E., Dülmer, H., Schlüter, E., Schmidt, P., & Meuleman, B. (2012). Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance. *Journal of Cross-Cultural Psychology, 43*(4), 558–575.
46. Finch, H. (2005). The MIMIC model as a method for detecting DIF: Comparison with Mantel-Haenszel, SIBTEST, and the IRT likelihood ratio. *Applied Psychological Measurement, 29*(4), 278–295.
47. Kirk, R. E. (1995). *Experimental design: Procedures for the behavioral sciences*. Pacific Grove: Brooks/Cole.
48. Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48*(1), 85–112.
49. Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Lawrence Erlbaum Associates.
50. Zaidi, Z. F. (2010). Gender differences in human brain: A review. *The Open Anatomy Journal, 2*, 37–55.
51. Nolen-Hoeksema, S., & Girgus, J. S. (1994). The emergence of gender differences in depression during adolescence. *Psychological Bulletin, 115*(3), 424–443.
52. Bennett, D. S., Ambrosini, P. J., Kudes, D., Metz, C., & Rabinovich, H. (2005). Gender differences in adolescent depression: Do symptoms differ for boys and girls? *Journal of Affective Disorders, 89*(1–3), 35–44.
53. Verhoeven, M., Sawyer, M. G., & Spence, S. H. (2013). The factorial invariance of the CES-D during adolescence: Are symptom profiles for depression stable across gender and time? *Journal of Adolescence, 36*(1), 181–190.
54. Bisegger, C., Cloetta, B., von Rueden, U., Abel, T., Ravens-Sieberer, U., & European Kidscreen Group. (2005). Health-related quality of life: Gender differences in childhood and adolescence. *Sozial- und Präventivmedizin, 50*(5), 281–291.
55. Viira, R., & Koka, A. (2012). Gender differences in health-related quality of life among estonian adolescents: A 6-month follow-up. *Acta Kinesiologiae Universitatis Tartuensis, 18*, 84–93.
56. Zahn-Waxler, C., Shirtcliff, E. A., & Marceau, K. (2008). Disorders of childhood and adolescence: Gender and psychopathology. *Annual Review of Clinical Psychology, 4*, 275–303.
57. Wren, F. J., Berg, E. A., Heiden, L. A., Kinnamon, C. J., Ohlson, L. A., Bridge, J. A., Birmaher, B., & Bernal, M. P. (2007). Childhood anxiety in a diverse primary care population: Parent–child reports, ethnicity and SCARED factor structure. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*(3), 332–340.
58. Laursen, B., Coy, K. C., & Collins, W. A. (1998). Reconsidering changes in parent–child conflict across adolescence: A meta-analysis. *Child Development, 69*(3), 817–832.
59. Galambos, N. L., Leadbeater, B. J., & Barker, E. T. (2004). Gender differences in and risk factors for depression in adolescence: A 4-year longitudinal study. *International Journal of Behavioral Development, 28*(1), 16–25.
60. Pine, D. S., Cohen, P., Gurley, D., Brook, J., & Ma, Y. (1998). The risk for early-adulthood anxiety and depressive disorders in adolescents with anxiety and depressive disorders. *Archives of General Psychiatry, 55*(1), 56–64.
61. Svedberg, P., Eriksson, M., & Boman, E. (2013). Associations between scores of psychosomatic health symptoms and health-related quality of life in children and adolescents. *Health and Quality of Life Outcomes, 11*, 176.
62. Kendall, P. C., Kortlander, E., Chansky, T. E., & Brady, E. U. (1992). Comorbidity of anxiety and depression in youth: Treatment implications. *Journal of Consulting and Clinical Psychology, 60*(6), 869–880.
63. Muthen, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research, 22*(3), 376–398.
64. Ryu, E. (2014). Factorial invariance in multilevel confirmatory factor analysis. *The British Journal of Mathematical and Statistical Psychology, 67*(1), 172–194.