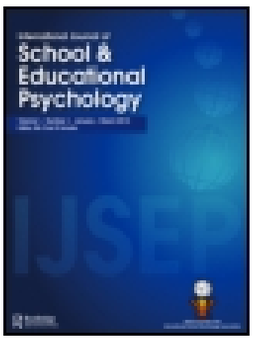

International Journal of School & Educational Psychology

ISSN: 2168-3603 (Print) 2168-3611 (Online) Journal homepage: https://www.tandfonline.com/loi/usep20

Testing measurement invariance and multilevel factor structures of bullying among Iranian adolescents

Maysam Rezapour, Narges Khanjani & Moghadameh Mirzai

To cite this article: Maysam Rezapour, Narges Khanjani & Moghadameh Mirzai (2019): Testing measurement invariance and multilevel factor structures of bullying among Iranian adolescents, International Journal of School & Educational Psychology, DOI: 10.1080/21683603.2019.1605952

To link to this article: https://doi.org/10.1080/21683603.2019.1605952

Published online: 26 Apr 2019.

INTERNATIONAL JOURNAL OF SCHOOL & EDUCATIONAL PSYCHOLOGY
https://doi.org/10.1080/21683603.2019.1605952

Testing measurement invariance and multilevel factor structures of bullying among Iranian adolescents

Maysam Rezapour [a], Narges Khanjani[a], and Moghadameh Mirzai[b]

[a]Neurology Research Center, Kerman University of Medical Sciences, Kerman, Iran; [b]Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

ABSTRACT

School bullying is a significant public health problem and a widespread type of school violence. Empirically validated evaluation scales are needed for use in research and practice to explain the nature and scope of adolescent bullying. The current study investigated the multilevel factor structure and measurement invariance of the Olweus bullying, victimization, and perpetration scales across gender and school types on a random sample of 1,558 Iranian adolescences. Although results indicated adequate goodness of fit for the one-factor and four-factor structures in both scales, it seems like the four-factor structure is more appropriate for use in within-level (students) and the one-factor for between-level (school). In addition, measurement invariance analyses showed that configural, metric, and scalar invariance were established across gender and school types. These findings provide psychometrical support for the use of the Olweus bullying, victimization, and perpetration scales in Iranian adolescents.

KEYWORDS

Bullying victimization scale; bullying perpetration scale; Iran; measurement invariance; multilevel confirmatory factor analysis; school

Bullying among adolescents is a significant public health problem and a widespread type of school violence (Gladden, Vivolo-Kantor, Hamburger, & Lumpkin, 2014) that has been consistently linked to serious social, emotional, and psychological problems for both the perpetrators and victims (Gini & Pozzoli, 2009; Mishna, McLuckie, & Saini, 2009). Thus, empirically validated evaluation tools are needed to detect bullying perpetration and victimization among adolescents, in research and practice. This is a fundamental step to prevent bullying behaviors and reduce further adverse psychological consequences.

Bullying: Its definition and types

Bullying is defined as a subtype of intentional aggressive behavior that is repeated against a victim who cannot readily defend himself or herself; and as an important aspect of bullying, there is a certain imbalance of power or strength between the bully and the victim (Olweus, 2010). Bullying includes many forms, such as physical (e.g., hitting, pushing, and kicking), verbal (e.g., name-calling and teasing in a hurtful way), relational or social (e.g., social exclusion and spreading rumors), and cyber, which is a more modern form (Gladden et al., 2014). Cyber bullying can happen through mobile phones, Internet, e-mail, online social networking (telegram or Instagram), or creating nasty websites (Monks & Coyne, 2011). Although this framework is accepted in extant research, there is variability in bullying concepts across cultures (Smith, Cowie, Olafsson, & Liefooghe, 2002). Thus, it is necessary to evaluate the framework of bullying in different countries and languages; also, to extend and develop our knowledge about bullying in these cultures. Iran, as a Middle Eastern country, has an Islamic education system and can be considered as an example of a different context. Few studies about school bullying have been conducted in Iran (Mohebbi, Mirnasab, & Wiener, 2016; Rezapour, Khanjani, & Soori, 2019; Soori, Rezapour, & Khodakarim, 2014).

School context in Iran

There are several aspects of Iranian schools that might affect the perception of Iranian students about bullying behaviors. First, students within the educational system have limited access to the Internet or mobile phones. Second, all schools are segregated based on gender, because of Islamic beliefs and moral values. But in Iranian universities, male and female students are together in one classroom. Studies conducted about bullying up to now have been done in mixed-gender (or coeducational) schools and there are few

CONTACT Narges Khanjani n_khanjani@kmu.ac.ir Department of Epidemiology and Biostatistics, School of Public Health, Kerman University of Medical Sciences, Kerman, Iran.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/usep.

studies about bullying in single-gender environments (Pahlke, Hyde, & Allison, 2014). Third, there is a variety of school types in Iran. There are 47 types of schools in Iran's current education system, in which public (governmental), private (nongovernmental), and gifted schools are the most common types. These three school types comprise 90% of the schools in Iran (Minister of Education, 2016). In Iran, public schools are funded by the government and education is free for domestic students up to 19 years of age. However, parents have to pay for things like school uniforms, stationery, exam fees, and some course-related costs. Gifted schools are funded by the government, but receive a small amount of parent funding for super-educational programs, and their students get accepted through entrance exams. These are special schools whose teachers and trainers are selected from authoritative and experienced people with high academic education. Private schools are not government funded. They charge fees to operate. Rezapour, Khanjani, and Mirzai (2019) showed that students in private schools reported more positive perceptions about school environment subdomains.

Bullying measurement

On the other hand, different methodologies (Vivolo-Kantor, Martell, Holland, & Westby, 2014) make it difficult to provide a valid, widely accepted instrument for measuring bullying. Various approaches have been employed for determining the victims and perpetrators of bullying, including peer nomination by classmates, teacher or parent reports, and self-report. Each method has its strengths and weaknesses (Cornell & Bandyopadhyay, 2010). Self-report surveys are the most frequently used measures to estimate the prevalence of bullying in schools. Although self-report has the limitations of under- or overestimation, it is a rapid and efficient approach for gathering data from large numbers of students (Felix, Sharkey, Green, Furlong, & Tanigawa, 2011). In addition to this, selecting an appropriate tool from the various available instruments is hard. A recent review showed that more than 40 instruments are available for measuring bullying involvement (Vessey, Strout, DiFazio, & Walker, 2014).

The revised Olweus Bully/Victim Questionnaire (OBVQ; Olweus, 1996) is the most widely used multidimensional tool designed by now that measures bullying involvement experiences (Vessey et al., 2014). The OBVQ is based on the basic definition of bullying, which involves three criteria—intentionality, repetitiveness, and power imbalance—and assesses both perpetration and victimization using multiple items for various forms (i.e., verbal, physical, relational, cyber; Olweus, 2013). Data obtained from this tool constitute one of the evaluation requirements for the Olweus Bullying Prevention Program (OBPP; Olweus & Limber, 2010). However, there are several methodological gaps about its psychometric properties.

Research gaps

First, although bullying behaviors (perpetration and victimization) are extensively examined in multidimensional structures (Betts, Houston, & Steer, 2015; Eastman, Moore, Cecilione, Hettema, & Roberson-Nay, 2018; Harbin, Kelley, Piscitello, & Walker, 2019; Marsh et al., 2011), almost all previous studies about the psychometric properties of the Olweus victimization and perpetration of bullying scales considered items in a single construct (unidimensional structure) for both scales (perpetration and victimization; Bevans, Bradshaw, & Waasdorp, 2013; Breivik & Olweus, 2015; Gonçalves et al., 2016; Gothwal, Sumalini, Irfan, Giridhar, & Bharani, 2013; Guilheri, Cogo-Moreira, Kubiszewski, Yazigi, & Andronikof, 2015; Khawar, Malik, & Batool, 2015; Kyriakides, Kaloyirou, & Lindsay, 2006; Lee & Cornell, 2009; Roberson & Renshaw, 2018), except one study conducted by Rezapour, Soori, and Khodakarim (2013) in Iran. However, Rezapour et al. (2013) examined the multidimensional structures (verbal, relational, physical) of the Olweus bullying victimization and perpetration scales without considering the cyber dimension in both scales, on 830 public school students.

Second, since some features of school, like the school management, the teachers and staff, and the social characteristics of the pupils, are similar in one school, and pupils are clustered within schools, observations are not independent, and ignoring these clusters in analysis can lead to biased standard errors, parameter estimates, and fit statistics (Julian, 2001). Multilevel modeling techniques consider the dependency of observations at different levels, and multilevel confirmatory factor analysis (MLCFA) is one of these techniques. A strength of MLCFA is that, by subdividing the variance of scores into within and between-level components, the reliability of each factor can be obtained at each level. All previous psychometric studies on the Olweus bullying victimization and perpetration scales (Bevans et al., 2013; Breivik & Olweus, 2015; Gonçalves et al., 2016; Gothwal et al., 2013; Guilheri et al., 2015; Khawar et al., 2015; Kyriakides et al., 2006; Lee & Cornell, 2009; Roberson & Renshaw, 2018) ignore the clustering of individual respondents within classes or schools. In Iranian school contexts, the Olweus bullying victimization and perpetration scales have been accessed using conventional confirmatory factor analysis (CFA), but without accounting for the clusters of students in theirs schools (Rezapour et al., 2013).

Third, the response options of bullying survey questions are often ordinal and have a multivariate nonnormal distribution; therefore, using estimation methods defined for normally distributed and continuous data may lead to biased parameter estimates, which has happened in previous studies (Khawar et al., 2015; Kyriakides et al., 2006; Rezapour et al., 2013).

Finally, although extant research shows gender differences (Smith et al., 2002; Soori et al., 2014), and school types (Lehman, 2015; Topçu, Erdur-Baker, & Çapa-Aydin, 2008) affect the prevalence of various forms of bullying victimization and perpetration, few studies have evaluated the invariance of bullying and victimization factor structures across gender (Antoniadou, Kokkinos, & Markos, 2016; Marsh et al., 2011) and school types. If boys and girls do not perceive these constructs in a similar way, any observed difference in scores for gender are likely to be meaningless, and may be due to faulty measurements. This view also employs two school types. However, until now only two studies have evaluated the measurement invariance of the factor structures of the Olweus bullying victimization and perpetration scales across gender (Guilheri et al., 2015; Roberson & Renshaw, 2018) and no study has evaluated it by school type. Guilheri et al. (2015) assessed measurement invariance by gender in the Olweus bullying victimization and perpetration scales among 802 school children aged 9–12 (mean age = 10.3) in France, and Roberson and Renshaw (2018) did the same among a representative sample of U.S. youth in grades 5 to 10, in which configural and scalar invariance for gender was achieved. Until now, no study has examined the measurement invariance of the factor structures of bullying behaviors across gender and school types in Iran. On the other hand, establishing measurement invariance across gender and school types for victimization and perpetration domains provides significant information in understanding the constructs of bullying in Iran and helps develop interventional programs.

The present study

The present study addressed the research gaps mentioned in five ways among Iranian pupils. First, we evaluated and compared the psychometric properties of the Olweus bullying victimization and perpetration scales according to previous theories and research in unidimensional (one-factor) and multidimensional (four-factor) structures. Second, we investigated both student-level and school-level dimensions of the Olweus bullying victimization and perpetration scales by multilevel confirmatory factor analysis. Third, the current study used an estimator that was adjusted to multivariate nonnormal distribution. Fourth, we presented factor reliabilities (internal consistency) at both the student and school levels of both scales. Fifth, we assessed the measurement invariance of the multidimensional factor structures in both scales (victimization and perpetration) across gender and school types.

Methods

Participants

A stratified three-stage cluster sampling design was used to recruit 1,558 pupils aged 14–17 in Mazandran province, north of Iran. The study sample was drawn from 7 cities that were selected randomly from 22 cities in Mazandran province. From each city, 6 schools based on two strata, gender and school types (public, gifted, private schools), were randomly selected. In stage 3, one or two classes were selected randomly from the total classes, and all students voluntarily completed the scales. Eventually, 68 classes with a size range from 6 to 35 students from 42 schools with a size range from 35 to 435 students (mean = 229, *SD* = 89) were selected. The average number of students clustered within schools (42 schools in the sample) was 37 students. In the sample, 45.4% of the participants were boys, and 40.3% and 26.8% of the participants were from public and private schools, respectively.

Measures

The Persian-Olweus Bullying Questionnaire (P-OBQ) is the standardized form of the Olweus Bullying Questionnaire (OBQ) validated among Iranian pupils (Rezapour et al., 2013). In previous study, this questionnaire was validated through several steps including standard forward and backward translation of the OBQ into Persian, content validity, and traditional exploratory and confirmatory factor analysis, without considering the cyber forms. The test–retest (after two weeks) reliability coefficients were from 0.63 to 0.92. This questionnaire has two sections that include the "victimization of bullying" scale and the "perpetration of bullying" scale; and each scale, separately, included 9 items for measuring the frequency of various forms (verbal by three items, relational by two items, physical by three items, and cyber by one item) of bullying. We expanded one item in the cyber part to three items and asked the questions about bullying behaviors that happened in telegram, Instagram, and websites. Eventually there were 11 items in each scale. Response options were *never, only once or twice, 2 or 3 times a month, about once a week*, or *several times a week*. The English version of the P-OBQ can be found in Appendix A, and the Persian version of the P-OBQ in Appendix B.

Procedures

This study was approved by the ethics committee of the Kerman University of Medical Sciences (Ethics Code: IR.KMU.AH.REC.1395.89). Also, the Security Office of the Educational Authority of Mazandaran Province, Iran, approved this study; and informed consent was obtained from the parents, students, and teachers of the selected schools. Data were collected from January 25 until March 12, 2017. Pupils completed both scales about bullying victimization and perpetration in their classrooms during normal school hours.

Data analysis

Confirmatory factor analysis

First, conventional CFA was performed on the 11 items from each of the scales of bullying (the victimization of bullying scale and the perpetration of bullying scale). One-factor and four-factor (verbal, relational, physical, and cyber) structures were examined. Then multilevel confirmatory factor analysis (MLCFA) was conducted to account for the clustered nature of the data (within-level or level 1 for students, and between-level or level 2 for schools). MLCFA evaluated the same factor structures (the same number of factors) at within (student) and between (school) levels in both scales. The weighted least squares means and variance (WLSMV) estimator was used for analysis of all structures, which included robust standard errors and adjustment to the $\chi 2$ test statistic, due to unbalanced group sizes and categorical indicators (Heck & Thomas, 2015).

Intraclass correlation coefficients (ICC) were computed via variance-components models for each of the items in both scales for the school level. ICCs can range from 0 to 1 and higher values of the ICC indicate a greater proportion of school-level variance (between) and likely greater bias, if the clustered nature of the data is ignored. Although there are no clear guidelines for appropriate values of ICC that multilevel analyses should be performed, authors think ICCs smaller than 0.05 may benefit less (Brown, 2015). However, Muthén (1999) states that if the design effect is greater than 2, this indicates clustering in the data needs to be taken into account during estimation. Because of the small number of cities included in this study (seven cities), and also one or two classes per school, city and classes were not considered as a level. Therefore, only two levels were considered: between-level (schools) and within-level (students).

The model fitness was assessed using a combination of fit indices including the comparative fit index (CFI), the Tucker Lewis Index (TLI), the standardized root mean square residual (SRMR), the root mean square error of approximation (RMSEA), weighted root mean square residual (WRMR), and the relative chi-square that is the ratio of chi-square to degrees of freedom. Excellent fits were judged by CFI, and TLI values greater than 0.95 and SRMR values less than or equal to 0.06 (Hu & Bentler, 1999) and structures with CFIs and TLIs greater than or equal to 0.90 and an RMSEA less than or equal to 0.08 are considered adequate fits (Hu & Bentler, 1999).WRMR less than 1.0 indicates good model fit for categorical data (Yu, 2002). Also, although nonsignificant chi-square tests indicate a good fit, they are sensitive to large sample sizes and should be interpreted considering the degrees of freedom (Kline, 2015).

In the second step of analysis, two forms of reliability were computed for each factor in the factor structures in both scales. First, the Cronbach's alpha was calculated at the within-level (student-level). Cronbach's alpha does not account for the clustered data structure and does not explain the expected internal consistency at the between-level (school-level) . Second, the Spearman Brown reliability was calculated for each factor in the factor structures in both scales at between-level (school-level; Hox, 2010). The Spearman-Brown Formula to estimate factor reliabilities at the school level is: [k (ICC)]/[(k−1) (ICC) +1], where k is the average number of students (respondents) per school, which was 37 students in this study. The intra-class correlation (ICC) was calculated by constraining equal factor loadings at the within (student) and between (school) level $[\sigma^2_B/(\sigma^2_B + \sigma^2_W)]$ (Muthén, 1991).

Measurement invariance

The third step of analysis tested the multigroup invariance of suggested final factor structures of the first step in both scales. Multigroup measurement invariance investigated whether the latent constructs could be interpreted in the same way across different groups and is known as measurement equivalence. The clustering data structure was accounted for using pooled within-group covariance matrix modeling (Stapleton, 2013). The groups were gender and school types. Multigroup invariances of structures in both scales were carried out in three stages (configural, metric, and scalar) using the MODEL = CONFIGURAL METRIC SCALAR command in the MODEL option of the ANALYSIS command in Mplus 7.4 by WLSMV estimator (Muthén & Muthén, 1998-2015). In the first stage, configural invariance was tested, and provided baseline fit indices to detect invariance by comparing

with the fit statistics of succeeding, more restrictive models. Configural invariance indicates that the same factor structure exists across groups. In the second stage, metric invariance was tested (Pendergast, von der Embse, Kilgus, & Eklund, 2017). Metric (or weak) invariance indicates that the factor loadings of the scale are equivalent across groups (Pendergast et al., 2017). In the third stage, scalar invariance was estimated. Scalar (or strong) invariance indicates that the origin (or threshold for categorical and intercept for continued data) of the scale is similar across groups (Pendergast et al., 2017). Though there are no commonly accepted methods for assessing the change in fit-indices in multigroup CFA, the change in CFI index (ΔCFI) is generally used for evaluating measurement invariance that should be less than 0.01 (Byrne & Van de Vijver, 2010). Also, some authors have used ΔRMSEA < 0.05 for evaluating invariance (Savickas & Porfeli, 2012). Although $\Delta\chi^2$ statistic is the most widely used to inspect invariance, the test is influenced by large samples and it has been shown that, if ΔCFI < 0.01 and sample size is higher than 200, any differences between groups might be negligible (Meade, Johnson, & Braddy, 2008). Thus, although the χ2 change tests were reported, we ignored it in interpreting the invariance of the current study. All analyses were conducted using Mplus version 7.4, and the sampling weights were incorporated into analyses.

Results

Descriptive statistics

Thirty-six cases were excluded for analysis of the victimization of bullying scale (2.3%), and 28 cases were excluded from the perpetration of bullying scale (1.8%), because of incomplete and missing values, and the size of the final sample was 1,522 pupils for the victimization of bullying scale and 1,530 pupils for the perpetration of bullying scale. Correlations, means, standard deviations, skewness, kurtosis, and intraclass correlations (ICCs) of items from both scales are presented in Table 1. Items were correlated in a range from 0.27 to 0.89 in within-level and 0.01 to 1.00 in between-level for the victimization scale; and in a range from 0.33 to 0.81 in within-level and 0.11 to 1.00 in between-level for the perpetration scale. The variability between and within schools on each item was examined by intraclass correlations (ICCs) for each of scale. The ICCs for items in the victimization of bullying scale ranged from 0.03 (for item 2) to 0.20 (for item 3); and ICCs for items of perpetration of bullying scale ranged from 0.01 (for item 22) to 0.32 (for item 14; Table 1).

Prior to conducting MLCFA for each scale, conventional CFA was performed with one-factor and four-factor structures. Table 2 displays the fit indices of one- and four-factor structures in conventional CFA and MLCFA for both scales. Fit indices for both scales in conventional CFA showed that there was a good overall fit for each factor structure. Also, for MLCFA, the RMSEA, CFI, and TLI of each factor structure had an acceptable fit for both scales. SRMR for the victimization scale at both levels (within and between) was acceptable. While for the perpetration of bullying scale the SRMR at within-level (students) in both structures was acceptable, at between-level (schools) it was not good. The RMSEA index of MLCFA indicates how well a given model approximates the true model. If it has values less than 0.05 for both scales and in both structures, this is known as excellent.

Factor loadings

Standardized factor loadings of MLCFA for both factor structures that have the same number of factors at within (student) and between (school) levels in both scales are shown in Table 3. Since the residual variances are often small on the between-level of multilevel models, Muthén (2005) recommends that, if the residual variances are negative and small in between levels, it can be fixed to zero to allow the model to converge. In Table 3, in one- and four-factor structures of both scales, items that had a small negative residual variance estimate at the school level were fixed at zero.

At within-level (students) of both scales, the standardized factor loadings for all items at both factor structures were near to each other (ranged from 0.59 to 0.90 for one-factor structure and from 0.64 to 0.93 for four-factor structure) and statistically significant ($p < .05$). But at between-level (schools) for the victimization scale, the standardized factor loadings for all items at both factor structures ranged from 0.23 to 1.00, and the standardized factor loading for item 10 was not significant ($p > .05$). Also, the standardized factor loadings at level 2 of the perpetration of bullying scale in all structures ranged from 0.50 to 1.00, and the standardized factor loadings for items 21 and 22 in both factor structures and items 13 and 14 in four-factor structures were not significant ($p > .05$).

Interfactor correlations of the four-factor structure (verbal, relational, physical, and cyber) at within and between levels in both scales were higher than 0.61and are displayed in Table 4. In the four-factor structure at within-level of the victimization scale, the smallest correlation was between the verbal and cyber factors

Table 1. The descriptive statistics (mean, standard deviation, skewness, kurtosis, school intraclass correlation (coefficients of items) and correlations of the bullying victimization and perpetration scales of between-level (upper diagonal) and within-level.

Victimization scale											
	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11
Item 1	-	0.92*	0.95*	0.84*	1.00*	1.00*	0.99*	0.91*	1.00*	0.02	0.92*
Item 2	0.47*	-	1.00*	1.00*	0.90*	1.00*	1.00*	0.88*	1.00*	0.42*	0.99*
Item 3	0.43*	0.51*	-	0.91*	1.00*	0.95*	1.00*	1.00*	1.00*	0.43*	1.00*
Item 4	0.41*	0.47*	0.44*	-	0.72*	0.94*	0.82*	0.89*	1.00*	0.09	0.99*
Item 5	0.32*	0.42*	0.50*	0.37*	-	1.00*	1.00*	1.00*	1.00*	0.14	0.96*
Item 6	0.32*	0.51*	0.53*	0.46*	0.64*	-	1.00*	0.96*	1.00*	0.13	1.00*
Item 7	0.54*	0.43*	0.39*	0.37*	0.27*	0.42*	-	0.04	0.02	0.02	0.01
Item 8	0.45*	0.41*	0.49*	0.35*	0.40*	0.56*	0.55*	-	1.00*	0.38	0.86*
Item 9	0.27*	0.45*	0.55*	0.41*	0.56*	0.65*	0.36*	0.51*	-	0.07	1.00*
Item 10	0.48*	0.54*	0.59*	0.62*	0.39*	0.60*	0.33*	0.30*	0.71*	-	0.32
Item 11	0.29*	0.51*	0.48*	0.36*	0.43*	0.58*	0.47*	0.56*	0.89*	0.65*	-
Mean	1.80	1.32	1.19	1.44	1.12	1.12	1.36	1.26	1.01	1.03	1.05
SD	1.30	0.84	0.61	0.93	0.48	0.49	0.88	0.82	0.50	0.29	0.33
Skewness	1.54	3.16	4.14	2.52	5.06	5.24	2.93	3.59	6.06	10.12	8.56
Kurtosis	1.00	9.78	18.90	6.01	29.05	31.00	8.37	12.35	39.07	110.54	83.32
ICC_school	0.13	0.03	0.20	0.09	0.11	0.08	0.11	0.12	0.06	0.07	0.08
Perpetration scale											
	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20	Item 21	Item 22
Item 12	-	0.91*	1.00*	0.88*	0.93*	1.00*	1.00*	1.00*	0.90*	0.80*	1.00*
Item 13	0.50*	-	0.93*	0.19*	0.85*	0.61*	0.83*	1.00*	0.60*	0.11	1.00*
Item 14	0.49*	0.39*	-	1.00*	1.00*	1.00*	0.97*	0.97*	0.96*	1.00*	1.00*
Item 15	0.44*	0.61*	0.33*	-	1.00*	0.95*	0.94*	0.62*	1.00*	1.00*	–0.50*
Item 16	0.46*	0.65*	0.44*	0.73*	-	1.00*	1.00*	0.91*	1.00*	1.00*	1.00*
Item 17	0.47*	0.63*	0.59*	0.76*	0.81*	-	1.00*	1.00*	1.00*	1.00*	1.00*
Item 18	0.58*	0.46*	0.41*	0.49*	0.55*	0.56*	-	0.95*	1.00*	0.86*	1.00*
Item 19	0.55*	0.50*	0.49*	0.58*	0.58*	0.64*	0.58*	-	0.92*	1.00*	1.00*
Item 20	0.49*	0.62*	0.45*	0.66*	0.79*	0.74*	0.50*	0.60*	-	1.00*	1.00*
Item 21	0.52*	0.56*	0.39*	0.66*	0.60*	0.73*	0.42*	0.42*	0.80*	-	0.99*
Item 22	0.60*	0.58*	0.55*	0.72*	0.76*	0.76*	0.50*	0.61*	0.67*	0.70*	-
Mean	1.51	1.23	1.30	1.11	1.08	1.10	1.28	1.20	1.08	1.03	1.03
SD	1.01	0.65	0.70	0.50	0.43	0.49	0.74	0.75	0.43	0.31	0.29
Skewness	2.24	3.80	3.03	5.75	6.55	5.62	3.45	4.13	6.59	10.61	10.44
Kurtosis	4.35	16.35	10.41	36.06	46.77	34.03	12.66	16.62	48.08	118.68	117.44
ICC_school	0.15	0.02	0.32	0.07	0.11	0.13	0.15	0.18	0.08	0.03	0.01

Note. Between-level: school; Within-level: students. Full item descriptions are found in the Appendix.
Polychoric correlations, suitable for use with categorical indicators using weighted least squares with mean and variance corrected (WLSMV) estimation. *p value < 0.05.

Table 2. Fit indices for conventional confirmatory factor analysis and multilevel confirmatory factor analyses (MLCFA) of the bullying victimization and perpetration scales.

	χ^2	*df*	RMSEA	CFI	TLI	SRMR (within)	SRMR (between)	WRMR
Victimization scale								
Conventional confirmatory factor analysis								
One-factor structure	170.51	44	0.043	0.971	0964	-	-	1.196
Four-factor structure	72.64	38	0.024	0.992	0.998	-	-	0.74
Multilevel confirmatory factor analyses (MLCFA)								
One-factor structure	291.59	88	0.039	0.91	0.888	0.092	0.070	1.197
Four-factor structure	157.68	76	0.026	0.964	0.948	0.064	0.066	0.790
Perpetration scale								
Conventional confirmatory factor analysis								
One-factor structure	114.27	44	0.032	0.984	0.980	-	-	0.983
Four-factor structure	85.20	38	0.028	0.989	0.985	-	-	0.80
Multilevel confirmatory factor analyses (MLCFA)								
One-factor structure	145.61	88	0.021	0.979	0.974	0.061	0.221	0.770
Four-factor structure	101.88	76	0.015	0.991	0.987	0.051	0.195	0.585

Note. All χ2s are statistically significant ($p < .001$). RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker Lewis index; SRMR = standardized root mean square residual (at the both the within and between levels); WRMR = weighted root mean square residual; *df* = degrees of freedom.

Table 3. Standardized factor loadings for each structure at both levels (within and between) for each scale (victimization and perpetration scales) by multilevel confirmatory factor analysis.

			Structure: four-factor							
	Structure: one-factor		Verbal	Relational	Physical	Cyber	Verbal	Relational	Physical	Cyber
	λ-w	*λ-b*	*λ-w*	*λ-w*	*λ-w*	*λ-w*	*λ-b*	*λ-b*	*λ-b*	*λ-b*
Victimization of Bullying Scale										
Item 1	0.59	0.96	0.67				0.96			
Item 2	0.66	0.98		0.73				1.00[a]		
Item 3	0.70	0.99[a]			0.74				1.00[a]	
Item 4	0.60	0.88		0.64				0.94		
Item 5	0.64	0.96			0.68				0.97	
Item 6	0.75	1.00[a]			0.79				1.00	
Item 7	0.62	1.00	0.72				1.00[a]			
Item 8	0.67	0.96	0.76				0.96			
Item 9	0.85	1.00[a]				0.93				0.98
Item 10	0.77	0.24*				0.85				0.23*
Item 11	0.90	0.99				0.92				0.95
Perpetration of Bullying Scale										
Item 12	0.69	1.00[a]	0.74				1.00[a]			
Item 13	0.72	0.83		0.74				0.50*		
Item 14	0.60	1.00[a]			0.61				1.00[a]	
Item 15	0.80	0.86		0.82				0.51*		
Item 16	0.87	0.99			0.88				0.99	
Item 17	0.91	0.99			0.92				0.98	
Item 18	0.70	1.00[a]	0.74				1.00[a]			
Item 19	0.74	0.96	0.80				0.96			
Item 20	0.84	0.95				0.91				1.00[a]
Item 21	0.77	0.88*				0.82				0.94*
Item 22	0.83	0.71*				0.88				0.79*

Note. λ_w = Standardized factor loadings for within-level, λ_b = Standardized factor loadings for between-level; *Standardized factor loadings not statistically significant, $p > 0.05$; [a]residual variance fixed to zero at the between-level as recommended by Muthén (2005); full item descriptions are found in the Appendix.

(0.61). While in within-level of the perpetration scale, interfactor correlations of the verbal with cyber factors had the smallest correlation (0.78). Interfactor correlations of both scales in between-level were high and statistically significant (ranges of 0.75 to 1.00).

With regard to high between-level (school) correlations in both scales, we suggest an alternative structure for both scales, which is the four-factor structure at within-level, and a single factor structure at between-level (Figures 1 and 2). The advantage of the one-factor

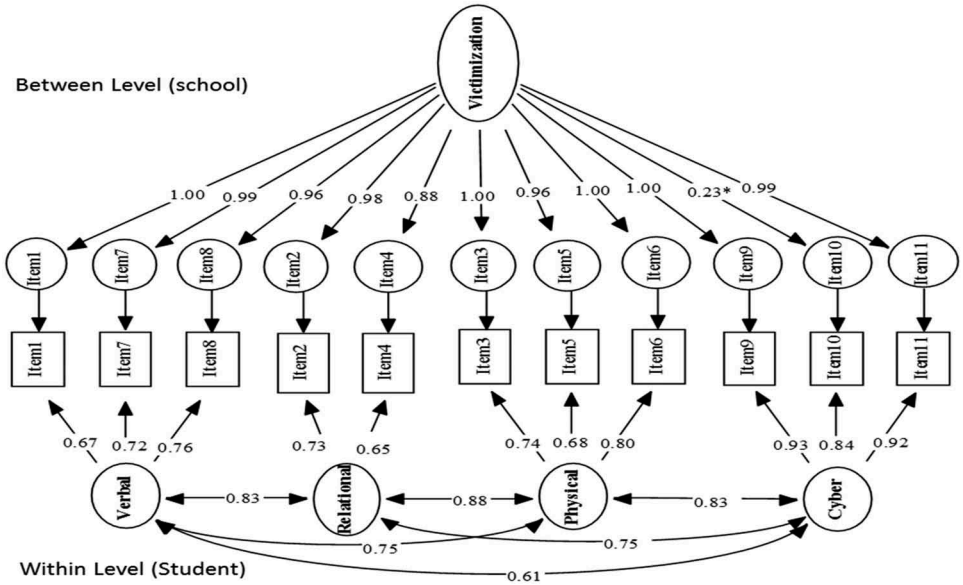

Figure 1. MCFA of four-factor structure in within-level and one-factor structure in between-level for victimization of bullying scale. *p value > 0.05. Fit indices; χ2 = 169.48, df = 86, RMSEA = 0.025, CFI = 0.963, TLI = 0.953, $SRMR_w$ = 0.64, $SRMR_b$ = 0.071, WRMR = 0.791.

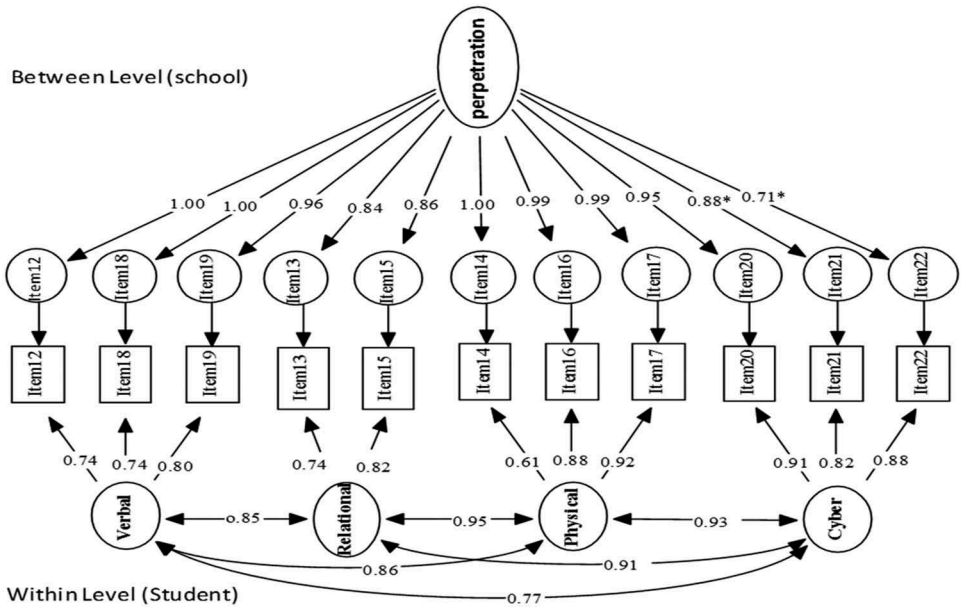

Figure 2. MCFA of four-factor structure in within-level and one-factor structure in between-level for perpetration of bullying scale. *p-value > 0.05. Fit indices; χ2 = 109.51, df = 85, RMSEA = 0.014, CFI = 0.991, TLI = 0.989, $SRMR_w$ = 0.51, $SRMR_b$ = 0.221, WRMR = 0.589.

solution at between-level was more parsimonious than the four-factor structure.

Reliability

In order to compare the factor variances at the students (within) and schools (between) levels with one another, the factor loadings were constrained to invariants at both levels (Mehta & Neale, 2005). The amount of variance in each level, ICC for each factor, and multilevel reliability for four-factor structures in both scales are shown in Table 4. Student-level Cronbach's alpha across both scales for four-factor structures were acceptable (ranged from 0.58 to 0.69 for victimization and ranged from 0.58 to 0.70 for perpetration). School-level Spearman-Brown reliability across both scales for four-

Table 4. Interfactor correlations for MLCFA, Variances, Intraclass Correlation Coefficients (ICC), and reliabilities of factors in bullying victimization and perpetration scales, separately.

	Within-level (students)				Between-level (schools)					Reliability	
	Verbal	Relational	Physical	Cyber	Verbal	Relational	Physical	Cyber	ICC	Between-level	Within-level
Victimization Scale											
Verbal	1.00	0.83	0.76	0.61	1.00	0.87	1.00	0.81	0.21	0.91	0.67
Relational	-	1.00	0.89	0.75	-	1.00	0.85	0.75	0.08	0.75	0.60
Physical	-	-	1.00	0.83	-	-	1.00	0.80	0.18	0.89	0.69
Cyber	-	-		1.00	-	-	-	1.00	0.03	0.52	0.58
Variance	0.83	1.10	1.25	6.65	0.22	0.09	0.27	0.19	-	-	
Perpetration Scale											
Verbal	1.00	0.85	0.86	0.78	1.00	0.93	0.97	0.95	0.17	0.88	0.68
Relational	-	1.00	0.96	0.92	-	1.00	1.00	0.91	0.01	0.27	0.58
Physical	-	-	1.00	0.93	-	-	1.00	1.00	0.22	0.91	0.70
Cyber	-	-	-	1.00	-	-	-	1.00	0.05	0.66	0.65
Variance	1.83	1.19	0.64	3.94	0.39	0.014	0.18	0.22	-	-	

Note. All correlations were statistically significant (all $p < 0\ .05$).

factor structures were also acceptable, except for cyber victimization (α = 0.52) and relational perpetration factors (α = 0.27; Table 4). Student-level Cronbach's alphas for victimization and perpetration scales (one-factor structure) were 0.75 and 0.83 respectively. The ICCs for the victimization and the perpetration scales were 0.12 and 0.10, and the Spearman-Brown reliability at school-level was 0.83 and 0.80, respectively (not shown in tables).

Measurement invariance

Multiple-group confirmatory factor analysis (MG-CFA) was used to test measurement invariance in the four-factor structure (Figures 1 and 2) across gender and school types in each scale (Table 5). Findings from the test for configural invariance (unconstrained model) show that both scales have equivalent factor structures across gender and school types. After constraints were added to test the metric model, the very small changes in CFI and RMSEA (ΔCFI and ΔRMSEA) between the unconstrained (configural) and constrained (metric) models achieved metric invariance across gender and school types. After adding constraints to the threshold (for categorical data), the scalar test for measurement invariance was conducted. Changes less than 0.01 in CFI and less than 0.05 in RMSEA between the metric and scalar models show scalar invariance across gender and school types (MG-CFA to test measurement invariance in the one-factor structure across gender and school types in each scale is presented in "Appendix C" and shows that configural, metric, and scalar invariance were established across gender and school types).

Discussion

In the current study, we examined the factor structure of a widespread self-report measure of perpetration and victimization of bullying, as well as the measurement invariance of this measure across gender and school types among Iranian adolescents. The results of the MLCFA showed that, although both structures when unidimensional (one-factor) and multidimensional (four-factor) have adequate psychometric properties, the four-factor structure in within-level and one-factor structure in between-level was the best model for both scales.

The multilevel reliability of the total scales in terms of their internal consistency was acceptable (0.83 for perpetration and 0.80 for victimization); and this was true for each factor in both scales. The internal consistency of verbal, relational, and physical factors in victimization and verbal, physical, and cyber factors in the perpetration scale were acceptable, while school-level Spearman-Brown reliability of the cyber factor in the victimization scale and the relational factor in perpetration scale were not acceptable (0.27 for relational perpetration and 0.52 for cyber victimization). This refers to the low ICCs of these factors. Thus, we recommend that the items of these dimensions (relational perpetration and cyber victimization) are not appropriate to measure at school-level. Overall, ICCs of the relational and cyber factors compared to verbal and physical factors in both scales were lower. Thus, variations of relational and cyber factors are explained more at the individual level, which may refer to the personal, individual, and even genetic characteristics of students. But, variations of verbal and physical factors are explained more at the school level (between level), which may refer to high cultural and environmental commonalities among students.

Table 5. Fit indices for invariance tests of the four-factor structure in both the bullying victimization and perpetration scales across gender and school types in the four-factor structure.

	χ^2	*df*	$\Delta\chi^2$	*df*	*p* value	CFI	ΔCFI	RMSEA	ΔRMSEA	TLI
Victimization Scale (*N* = 1522)										
Gender (boys, girls)										
Configural	225.8	76	-	-	-	0.957	-	0.051	-	0.937
Metric	258.2	83	65.9	7	< 0.001	0.949	0.008	0.052	0.001	0.933
Scalar	286.8	112	56.6	29	0.001	0.949	0.000	0.045	0.007	0.950
School types (public, gifted, private)										
Configural	414.4	114	-	-		0.941	-	0.072	-	0.914
Metric	423.2	128	33.3	14	0.003	0.941	0.000	0.067	0.007	0.925
Scalar	434.0	186	76.0	58	0.056	0.957	−0.016	0.051	0.016	0.957
Perpetration Scale (*N* = 1530)										
Gender (boys, girls)										
Configural	198.5	76	-	-	-	0.976	-	0.046	-	0.965
Metric	235.8	83	81.3	7	< 0.001	0.970	0.006	0.049	−0.003	0.960
Scalar	267.9	112	59.1	29	0.001	0.969	0.001	0.043	0.006	0.970
School types (public, gifted, private)										
Configural	224.5	114	-	-	-	0.981	-	0.043	-	0.972
Metric	254.7	128	59.9	14	< 0.001	0.978	0.003	0.044	−0.001	0.972
Scalar	326.1	186	107.7	58	< 0.001	0.976	0.002	0.038	0.006	0.979

Note. χ2 = chi-square test; *df* = degrees of freedom; Δ = change in values between competing models (constrained and unconstrained models); CFI = comparative fit index; RMSEA = root mean square error of approximation; TLI = Tucker Lewis Index; change in χ2 was calculated using the MODEL = CONFIGURAL METRIC SCALAR command in Mplus 7.4.

Although, internal consistency is affected from the number of items (only two items for the relational factor and three items for the cyber factor), in our study, when the multilevel structure of the data (i.e., students nested within schools) was ignored, the internal consistency of the cyber factor in the victimization scale and the relational factor in the perpetration scale were acceptable. Low ICCs in the cyber factor may be because of the independency of cyber bullying from the school setting. But, additional research is needed to further explore the reason of the low ICCs of relational factors in Iranian schools. A recent systematic review about investigating the contextual-level risk factors of school bullying showed that the range of ICC at school-level was 0.9% to 5.6% for bullies and 0.6% to 13% for victims (Azeredo, Rinaldi, De Moraes, Levy, & Menezes, 2015).

Consistent with previous studies (Breivik & Olweus, 2015; Gonçalves et al., 2016; Gothwal et al., 2013; Guilheri et al., 2015: Khawar et al., 2015; Kyriakides et al., 2006; Lee & Cornell, 2009), findings of the current study supported a unidimensional construct for both scales. Also, consistent with the study done by Rezapour et al. (2013), findings of the current study supported the multidimensional construct of the Olweus Bullying Questionnaire.

Each of these structures (one-factor and four-factor) in both scales can be explained in two perspectives, statistical and conceptual. Since both structures in both scales showed an acceptable statistical fit, using each of these structures for various studies depends on the nature of the research. If the nature of the research is epidemiologic, we suggest using the one-factor structure, because using this structure can show the incidence and prevalence trend of total bullying behaviors over the years. But if the nature of the research is investigating the consequences of each form of bullying, it is better to use the four-factor structure, because in the four-factor structure, it is assumed that each of these forms of bullying (verbal, physical, relational, and cyber) are not similar in regard to their psychological and clinical burden, circumstances of occurrence, or duration of involvement (Cornell, Sheras, & Cole, 2006).

Consistent with the current study, Marsh et al. (2011) conducted CFA and showed the intercorrelations between verbal, relational, and physical forms of victimization ranged from 0.84 to 0.83, and for perpetration of bullying ranged from 0.72 to 0.83. In addition, findings of our study showed high positive and statistically significant intercorrelations between the cyber factors and traditional factors (verbal, relational, and physical) in within-level for both scales that ranged from 0.61 to 0.93, while Antoniadou et al. (2016) showed moderate, positive, and statistically significant intercorrelations between cyber factors and traditional factors among Greek adolescents (0.43 for cyber victimization and traditional victimization and 0.46 for cyber perpetration and traditional perpetration).

The scales showed configural, metric, and scalar invariance across genders (boys, girls) and school types (public, gifted, private). Thus, boys and girls as

well as students of public, private, and gifted schools have a similar understanding of what bullying is, but they do differ in the types they use and experience. These findings are consistent with results obtained by Guilheri et al. (2015) in France and Roberson and Renshaw (2018) in the United States by the Olweus bullying victimization and perpetration scales; and are also consistent with results obtained by Marsh et al. (2011) in Australia and Antoniadou et al. (2016) in Greece by other measures of bullying victimization and perpetration.

Limitations

The present study had several limitations. Firs,t the study relied on pupils' reports about involvement in bullying and did not use other data collection approaches such as school discipline records or principal-reports and peer-reports to reduce potential bias from self-reports. However, the measurement invariance of both scales shows the good quality of the data. In addition, because bullying behaviours are dynamic between bullies and victims (victimization of bullying is a complement to perpetration of bullying), the observed patterns of factor structures in the victims and bullies are expected to be the same and, in this situation, measurement bias is low. In this study, this consistency in the patterns of factor structures among the victims of bullies was also observed. Second, the sample was only from one province in northern Iran, and additional research from other provinces of Iran is needed to evaluate the generalizability of our findings. These findings may not be generalized to students across the nation. Third, although an overriding contribution of the current study was accounting for the clusters of students within the schools in the analyses and doing appropriate modeling of factors at both individual and school levels, and because it also was the first study to do multilevel factor structures on the Olweus questionnaire, this study had a moderate sample size in the second level to support the model in between-level. Muthén and Muthén (2007) have stated that 30 to 50 samples in between-level is the minimal range recommended for multilevel factor analysis, and thus a larger sample of schools in between-levels in future studies would provide higher power to discriminate between alternative structures. Fourth, the stability and reliability of the scales over time (test–retest reliability) were not evaluated. Fifth, although in the present study multidimensional structures was suggested for each scale, with regard to the high intercorrelation between factors of each scale and low reliability for relational perpetration and cyber victimization factors, we suggested Exploratory Structural Equation Modeling (ESEM) to measure the structures and estimate reliability, because cross-loadings of each item on all factors in the ESEM within the same scale typically reduce high interfactor correlations (Marsh et al., 2011). Sixth, with regard to the low reliability of relational perpetration, we need qualitative research for building paradigms about this dimension of bullying in Iran and developing appropriate items. And finally, future research investigating the concurrent and predictive validity of the Olweus bullying victimization and perpetration scales, with physical and psychological consequences and different aspects of school climate, is also needed to be conducted among Iranian adolescences.

Conclusion and implications

The adaptation and validation of a school bullying measure is an important contribution to education and health research and preventive programs in Iran. Considering the absence of systematic preventive programs for bullying in Iran, these findings provide evidence to support the use of the Olweus bullying victimization and perpetration scales for assessment of bullying behaviors at both student and school levels, as well as planning preventive interventions among Iranian adolescences. When estimating the prevalence of bullying victimization and perpetration forms from the responses to the Olweus bullying victimization and perpetration scales, we can use the cutoff point of "2 or 3 times a month" as an operational criteria provided by Solberg and Olweus (2003) for dividing students into involved and not-involved in bullying. Also the one-factor structure of both scales would be appropriate for decision makers and policy makers, because it shows the general prevalence of these phenomena. But the multidimensional structures are more useful for therapists, psychiatrists, and clinical psychologists. Furthermore, the MLCFA showed the importance of using proper analysis to study the factor structure of a multilevel construct such as bullying behaviors, and explained the variation of this phenomenon in within-level and between-level. Results of MLCFA in the current study are valuable because the Olweus bullying victimization and perpetration scales are developed at student-level, which may not establish their psychometric properties when scores are aggregated at school-level. The MLCFA can help to prevent the potential measurement errors of the developed tools in student-level that are used to interpret at the school-level, and vice versa (Konold & Cornell, 2015). In addition, the obtained evidence of the measurement invariance of self-report measures in perpetration and victimization can compare scores on each scale across gender and school types, and suggests developing similar preventive interventions across these groups.

Acknowledgments

Researchers thank all of the schools, staff, and students who participated and the Security Office of the Education Department at Mazandaran Province. Authors also thank Dr. Amanda J. Nguyen, from the Curry School of Education, University of Virginia, for reading and commenting on the draft.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Kerman University of Medical Sciences [Grant No: 95000607];

Notes on contributors

Maysam Rezapour is a PhD candidate in Epidemiology at the Department of Epidemiology and Biostatistics at Kerman University of Medical Sciences, Iran. His current research interests include adolescent violence, school health, and social epidemiology.

Narges Khanjani is an Environmental Epidemiologist currently working at the Department of Epidemiology and Biostatistics, School of Public Health, Kerman University of Medical Sciences (KMU), Kerman, Iran. She is also a member of the Iranian Epidemiology Association (http://www.irea.ir/) and a member of the Iranian Association for Environmental Health (www.iaeh.org). She is an adjunct research fellow at the Monash Center for Occupational and Environmental Health, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia. She is also a member of the International Society for Environmental Epidemiology (ISEE), http://www.iseepi.org/ and the International Epidemiological Association (IEA), http://www.ieaweb.org/.

Moghadameh Mirzai, PhD, is an Assistant Professor of Biostatistics at the Department of Epidemiology and Biostatistics, Kerman University of Medical Sciences, Iran. Her current research interest is Multilevel Analyses.

ORCID

Maysam Rezapour http://orcid.org/0000-0002-6039-3286

References

Antoniadou, N., Kokkinos, C. M., & Markos, A. (2016). Development, construct validation and measurement invariance of the Greek cyber-bullying/victimization experiences questionnaire (CBVEQ-G). *Computers in Human Behavior*, *65*, 380–390. doi:10.1016/j.chb.2016.08.032

Azeredo, C. M., Rinaldi, A. E. M., De Moraes, C. L., Levy, R. B., & Menezes, P. R. (2015). School bullying: A systematic review of contextual-level risk factors in observational studies. *Aggression and Violent Behavior*, *22*, 65–76. doi:10.1016/j.avb.2015.04.006

Betts, L. R., Houston, J. E., & Steer, O. L. (2015). Development of the multidimensional peer victimization scale–Revised (MPVS-R) and the multidimensional peer bullying scale (MPVS-RB). *The Journal of Genetic Psychology*, *176*(2), 93–109. doi:10.1080/00221325.2015.1007915

Bevans, K. B., Bradshaw, C. P., & Waasdorp, T. E. (2013). Gender bias in the measurement of peer victimization: An application of item response theory. *Aggressive Behavior*, *39*(5), 370–380. doi:10.1002/ab.21486

Breivik, K., & Olweus, D. (2015). An item response theory analysis of the Olweus Bullying scale. *Aggressive Behavior*, *41*(1), 1–13. doi:10.1002/ab.21571

Brown, T. A. (2015). Recent developments involving CFA models. In T. A. Brown (Ed.), *Confirmatory factor analysis for applied research* (2nd ed., pp. 400–430). New York, United States of America: The Guilford Press.

Byrne, B. M., & Van de Vijver, F. J. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, *10*(2), 107–132. doi:10.1080/15305051003637306

Cornell, D. G., & Bandyopadhyay, S. (2010). The Handbook of bullying in schools: An international perspective. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *The assessment of bullying* (pp. 265–276). New York, NY: Routledge.

Cornell, D. G., Sheras, P. L., & Cole, J. C. M. (2006). Assessment of bullying. In M. J. F. Shane & R. Jimerson (Eds.), *The handbook of school violence and school safety: From Research to practice* (pp. 191–209). Mahwah, NJ: Lawrence Erlbaum Associates.

Eastman, M. L., Moore, A. A., Cecilione, J., Hettema, J. M., & Roberson-Nay, R. (2018). Confirmatory factor structure and psychometric properties of the multidimensional peer victimization scale. *Journal of Psychopathology and Behavioral Assessment*, *40*(4), 725–735. doi:10.1007/s10862-018-9678-2

Felix, E. D., Sharkey, J. D., Green, J. G., Furlong, M. J., & Tanigawa, D. (2011). Getting precise and pragmatic about the assessment of bullying: The development of the California bullying victimization scale. *Aggressive Behavior*, *37*(3), 234–247. doi:10.1002/ab.20389

Gini, G., & Pozzoli, T. (2009). Association between bullying and psychosomatic problems: A meta-analysis. *Pediatrics*, *123*(3), 1059–1065. doi:10.1542/peds.2008-1215

Gladden, R. M., Vivolo-Kantor, A. M., Hamburger, M. E., & Lumpkin, C. D. (2014). *Bullying surveillance among youths: uniform defnitions for public health and recommended data elements, version 1.0*. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention and U.S. Department of Education.

Gonçalves, F. G., Heldt, E., Peixoto, B. N., Rodrigues, G. A., Filipetto, M., & Guimarães, L. S. P. (2016). Construct validity and reliability of Olweus Bully/Victim Questionnaire–Brazilian version. *Psicologia: Reflexão e Crítica*, *29*(1), 27. doi:10.1186/s41155-016-0019-7

Gothwal, V. K., Sumalini, R., Irfan, S. M., Giridhar, A., & Bharani, S. (2013). Revised Olweus bully/victim questionnaire: Evaluation in visually impaired. *Optometry and Vision Science*, *90*(8), 828–835. doi:10.1097/OPX.0b013e3182959b52

Guilheri, J., Cogo-Moreira, H., Kubiszewski, V., Yazigi, L., & Andronikof, A. (2015). Bullying in school: Construct validity of the French version of the Olweus revised Bully/victim Questionnaire (rBVQ). *Neuropsychiatrie De l'Enfance Et De l'Adolescence*, *63*(4), 211–217. doi:10.1016/j.neurenf.2015.01.010

Harbin, S. M., Kelley, M. L., Piscitello, J., & Walker, S. J. (2019). Multidimensional bullying victimization scale: Development and validation. *Journal of School Violence*, *18*(1), 146–161. doi:10.1080/15388220.2017.1423491

Heck, R. H., & Thomas, S. L. (2015). Multilevel models with categorical variables. In R. H. Heck & S. L. Thomas (Eds.), *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus* (3nd ed., pp. 318–319). New York, NY: Routledge.

Hox, J. J. (2010). Multivariate Multilevel Regression Models. In J. J. Hox (Ed.), *Multilevel analysis: Techniques and applications* (2nd ed., pp. 188–204). New York, NY: Routledge.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. doi:10.1080/10705519909540118

Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling*, *8*(3), 325–352. doi:10.1207/S15328007SEM0803_1

Khawar, R., Malik, F., & Batool, S. S. (2015). Validation of The Urdu version of Olweus Bully/Victim Questionnaire (OBVQ). *Journal of Pakistan Psychiatric Society*, *12*(1), 24–28.

Kline, R. B. (2015). Global Fit Testing. In R. B. Kline (Ed.), *Principles and practice of structural equation modeling* (pp. 262–299). New York: Guilford publications.

Konold, T. R., & Cornell, D. (2015). Measurement and structural relations of an authoritative school climate model: A multi-level latent variable investigation. *Journal of School Psychology*, *53*(6), 447–461. doi:10.1016/j.jsp.2015.09.001

Kyriakides, L., Kaloyirou, C., & Lindsay, G. (2006). An analysis of the Revised Olweus Bully/Victim Questionnaire using the Rasch measurement model. *British Journal of Educational Psychology*, *76*(4), 781–801. doi:10.1348/000709905X53499

Lee, T., & Cornell, D. (2009). Concurrent validity of the Olweus bully/victim questionnaire. *Journal of School Violence*, *9*(1), 56–73. doi:10.1080/15388220903185613

Lehman, B. (2015). Physical and nonphysical bullying victimization of academically oriented students: The role of gender and school type. *American Journal of Education*, *122*(1), 133–159. doi:10.1086/683294

Marsh, H. W., Nagengast, B., Morin, A. J., Parada, R. H., Craven, R. G., & Hamilton, L. R. (2011). Construct validity of the multidimensional structure of bullying and victimization: An application of exploratory structural equation modeling. *Journal of Educational Psychology*, *103*(3), 701–732. doi:10.1037/a0024122

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*(3), 568–592. doi:10.1037/0021-9010.93.3.568

Mehta, P. D., & Neale, M. C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods*, *10*(3), 259–284. doi:10.1037/1082-989X.10.3.259

Minister of Education. (2016). *School types*. Retrieved from http://oerp.ir/article/1395/419.

Mishna, F., McLuckie, A., & Saini, M. (2009). Real-world dangers in an online reality: A qualitative study examining online relationships and cyber abuse. *Social Work Research*, *33*(2), 107–118. doi:10.1093/swr/33.2.107

Mohebbi, M., Mirnasab, M., & Wiener, J. (2016). Parental and school bonding in Iranian adolescent perpetrators and victims of bullying. *School Psychology International*, *37*(6), 583–605. doi:10.1177/0143034316671989

Monks, C. P., & Coyne, I. (2011). *Bullying in different contexts*. New York: Cambridge University Press.

Muthen, B. O. (2005). *Negative residual variance*. Retrieved from http://www.statmodel.com/discussion/messages/11/555.html?

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, *28*, 338–354. doi:10.1111/j.1745-3984.1991.tb00363.x

Muthen, L. K. (1999). *Intraclass correlations*. Retrieved from http://www.statmodel2.com/discussion/messages/12/18.html

Muthen, L. K., & Muthen, B. (2007). *Multilevel modeling with latent variables using Mplus* (Unpublished manuscript). doi:10.1094/PDIS-91-4-0467B

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Olweus, D. (1996). *The revised Olweus Bully/Victim Questionnaire*. Bergen, Norway: Research Center for Health Promotion, University of Bergen.

Olweus, D. (2010). Understanding and researching bullying some critical issues. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 9–34). New York, NY: Routledge.

Olweus, D. (2013). School bullying: Development and some important challenges. *Annual Review of Clinical Psychology*, *9*, 751–780. doi:10.1146/annurev-clinpsy-050212-185516

Olweus, D., & Limber, S. P. (2010). Bullying in school: Evaluation and dissemination of the Olweus bullying prevention program. *American Journal of Orthopsychiatry*, *80*, 124–134. doi:10.1111/j.1939-0025.2010.01015.x

Pahlke, E., Hyde, J. S., & Allison, C. M. (2014). The effects of single-sex compared with coeducational schooling on students' performance and attitudes: A meta-analysis. *Psychological Bulletin*, *140*(4), 1042–1072. doi:10.1037/a0035740

Pendergast, L. L., von der Embse, N., Kilgus, S. P., & Eklund, K. R. (2017). Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology. *Journal of School Psychology*, *60*, 65–82. doi:10.1016/j.jsp.2016.11.002

Rezapour, M., Khanjani, N., & Mirzai, M. (2019). Exploring associations between school environment and bullying in Iran: Multilevel contextual effects modeling. *Children and Youth Services Review*, *99*, 54–63. doi:10.1016/j.childyouth.2019.01.036

Rezapour, M., Khanjani, N., & Soori, H. (2019). The types of bullying behaviors and its association with general life satisfaction and self-rated health among iranian pupils. *Journal of Research in Health Sciences*, *19*(1), e00436.

Rezapour, M., Soori, H., & Khodakarim, S. (2013). Testing psychometric properties of the perpetration of bullying and victimization scales with Olweus bullying questionnaire in middle schools. *Journal of Safety Promotion and Injury Prevention*, *1*(4), 212–221.

Roberson, A. J., & Renshaw, T. L. (2018). Structural validity of the HBSC bullying measure: Self-report rating scales of youth victimization and perpetration behavior. *Journal of Psychoeducational Assessment*, *36*(6), 628–643. doi:10.1177/0734282917696932

Savickas, M. L., & Porfeli, E. J. (2012). Career adapt-abilities scale: Construction, reliability, and measurement equivalence across 13 countries. *Journal of Vocational Behavior*, *80*(3), 661–673. doi:10.1016/j.jvb.2012.01.011

Smith, P. K., Cowie, H., Olafsson, R. F., & Liefooghe, A. P. (2002). Definitions of bullying: A comparison of terms used, and age and gender differences, in a Fourteen–Country international comparison. *Child Development*, *73*(4), 1119–1133. doi:10.1111/1467-8624.00461

Solberg, M. E., & Olweus, D. (2003). Prevalence estimation of school bullying with the olweus bully/victim questionnaire. *Aggressive Behavior, 29*, 239–268. doi:10.1002/ab.10047

Soori, H., Rezapour, M., & Khodakarim, S. (2014). Epidemiological pattern of bullying among school children in Mazandaran Province-Iran. *Journal of Child and Adolescent Behavior*, *2*(145). doi:10.4172/2375-4494.1000145

Stapleton, L. M. (2013). Multilevel structural equation modeling techniques with complex sample data. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 521–562). Charlotte, NC, US: IAP Information Age Publishing.

Topçu, Ç., Erdur-Baker, Ö., & Çapa-Aydin, Y. (2008). Examination of cyberbullying experiences among turkish students from different school types. *CyberPsychology & Behavior*, *11*(6), 643–648. doi:10.1089/cpb.2007.0161

Vessey, J., Strout, T. D., DiFazio, R. L., & Walker, A. (2014). Measuring the youth bullying experience: A systematic review of the psychometric properties of available instruments. *Journal of School Health*, *84*, 819–843. doi:10.1111/josh.12210

Vivolo-Kantor, A. M., Martell, B. N., Holland, K. M., & Westby, R. (2014). A systematic review and content analysis of bullying and cyber-bullying measurement strategies. *Aggression and Violent Behavior*, *19*, 423–434. doi:10.1016/j.avb.2014.06.008

Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Retrieved from Mplus website http://www.statmodel.com/download/Yudissertation.pdf

Appendix A. The Victimization of Bullying Scale and Perpetration of Bullying Scale, English version.

	Victimization of Bullying Scale
Item1	I was called mean names, was made fun of, or teased in a hurtful way.
Item2	Other pupils left me out of things on purpose, left me out from their group of friends, or completely ignored me.
Item3	I was hit, kicked, pushed, shoved around, or locked indoors.
Item4	Other pupils told lies or spread false rumors about me and tried to make others dislike me.
Item5	I had money or other things taken away from me or damaged.
Item6	I was threatened or forced to do things I didn't want to do.
Item7	I was bullied with mean names or comments about my face.
Item8	I was bullied with mean names, comments, or gestures with asexual meaning.
Item9	I was bullied with exchanging hurtful or threatening texts and pictures on my cell telegram or Instagram.
Item10	I was bullied with creating hurtful websites.
Item11	I was bullied with mean or hurtful messages, calls or pictures, or in other ways on my cell phone.
	Perpetration of Bullying Scale
Item12	I called another pupil(s) mean names, made fun of or teased him or her in a hurtful way.
Item13	I kept him or her out of things on purpose, excluded him or her from my group of friends, or completely ignored him or her.
Item14	I hit, kicked, pushed, and shoved him or her around or locked him or her indoors.
Item15	I spread false rumors about him or her and tried to make others dislike him or her.
Item16	I took money or other things from him or her or damaged his or her belongings.
Item17	I threatened or forced him or her to do things he or she didn't want to do.
Item18	I bullied him or her with mean names or comments about his or her face.
Item19	I bullied him or her with mean names, comments, or gestures with a sexual meaning.
Item20	I bullied him or her with exchanging hurtful or threatening texts and pictures on my cell telegram or Instagram.
Item21	I bullied him or her with creating hurtful websites.
Item22	I bullied him or her with mean or hurtful messages, calls or pictures, or in other ways on my cell phone.
Responses options were "never," "only once or twice," "2 or 3 times a month," "about once a week," or "several times a week."	

Appendix B. The Bullying Victimization and Perpetration Scales, Persian version.

مقیاس قربانی شدن از زورگویی
آیا شما از شروع مدرسه تاکنون، توسط دانش آموزان دیگر در مدرسه، به یک یا چند روش زیر مورد زورگویی قرار گرفته اید ؟ لطفاً برای هریک از سوالات یک پاسخ را علامت بزنید
1-مرا با اسم های زشت صدا زدند مسخره ام کردند، یا با یک روش بد سر به سرم گذاشتند
2- دانش آموزان دیگر، مرا به خاطر چیزهایی عمداً از گروه دوستانشان بیرون کردند یا اصلاً تحویلم نگرفتند
3-دانش آموز دیگر با من دعوا افتاد, کتکم زد و مرا هل داد و روی زمین انداخت
4-دانش آموزان دیگر، دروغها یا شایعه های غلط درباره من پخش کردند و سعی کردند تا دیگران مرا دوست نداشته باشند
5-دانش آموزان دیگر، پول یا وسایل مرا، به زور از من گرفتند یا خراب کردند
6- دانش آموزان دیگر مرا ترساندند و با تهدید، مجبور به کارهایی کردند که خودم نمی خواستم
7-دانش آموزان دیگر، مرا با لقب های زشت درباره قیافه‌ام آزار دادند
8-دانش آموزان دیگر، با حرکات زشت جنسی به من توهین کردند
9- دانش آموز(ان) دیگر مرا با تصاویر و پیام های آزاردهنده یا تهدید کننده تلگرامی یا اینستاگرامی مورد توهین و تمسخر قرار دادند
10- دانش آموز (ان) دیگر با فریب و حیله ، در به اشتراک گذاری اطلاعات شخصی دیگران مرا مورد زورگویی قرار دادند
11- دانش آموز(ان) دیگر مرا با ارسال پیامک یا مکالمات تلفنی مورد توهین و تمسخر قرار دادند
مقیاس ارتکاب زورگویی
آیا شما از شروع مدرسه تاکنون ، به یک یا چند روش زیر، دانش آموزان دیگر را در مدرسه تان مورد زورگویی قرار داده اید؟ لطفاً برای هریک از سوالات یک پاسخ را علامت بزنید
12- دانش آموز دیگر را با اسم های زشت صدا زدم و مسخره کردم یا با یک روش بد ، سر به سرش گذاشتم
13- عمداً دانش آموز دیگر را از گروه دوستانم بیرون کردم و یا اصلاً او را تحویل نگرفتم
14- با دانش آموز دیگر دعوا افتادم، کتکش زدم ، هلش دادم و روی زمین انداختم
15- شایعات و دروغ هایی درباره دانش آموز دیگر پخش کردم و سعی کردم تا دیگران از او بد شان بیاید
16- من پول یا وسایل دانش آموز دیگر را به زور گرفتم یا خراب کردم
17- دانش آموز دیگر را با ترساندن وادار به انجام کارهایی کردم که خودش نمی خواست
18- دانش آموز دیگر را با لقب های زشت درباره قیافه اش آزار دادم
19- به دانش آموز دیگر با حرکات زشت جنسی توهین کردم
20- دانش آموز(ان) دیگر را با تصاویر و پیام های آزاردهنده یا تهدید کننده تلگرامی یا اینستاگرامی مورد تمسخر و زورگویی قرار دادم
21- دانش آموز (ان) دیگر را با فریب و حیله ، در به اشتراک گذاری اطلاعات شخصی دیگران مرا مورد زورگویی قرار دادم
22- دانش آموز(ان) دیگر را با ارسال پیامک یا مکالمات تلفنی مورد توهین و زورگویی قرار دادم
گزینه های پاسخ برای هردو مقیاس عبارت بودند از : 1:هرگز، 2:یک یا 2 بار، 3: 2یا 3 بار دریک ماه، 4: حدود یک بار در یک هفته، 5: چندین بار در یک هفته.

Appendix C. Fit indices for invariance tests of the bullying victimization and perpetration scales in different gender and school types in the one-factor structure.

	χ^2	*df*	$\Delta\chi^2$	*df*	*p Value*	RMSEA	ΔRMSEA	CFI	ΔCFI	TLI
Victimization Scale (*N* = 1522)										
Gender (Boys, Girls)										
Configural	292.7	88	-	-	-	0.055	-	0.941	-	0.926
Metric	343.1	98	112	10	< 0.001	0.057	−0.002	0.929	0.012	0.920
Scalar	371.4	130	78.1	32	0.001	0.049	0.008	0.929	0.000	0.941
School Type (Public, Gifted, Private)										
Configural	541.6	132	-	-	-	0.078	-	0.919	-	0.899
Metric	563.3	152	53.1	20	0.002	0.073	0.005	0.919	0.000	0.912
Scalar	535.6	216	86.8	64	0.053	0.054	0.019	0.937	−0.018	0.95
Perpetration Scale (*N* = 1530)										
Gender (Boys, Girls)										
Configural	224.3	88	-	-	-	0.045	-	0.973	-	0.966
Metric	252.3	98	66.9	10	< 0.001	0.045	0.000	0.970	0.003	0.966
Scalar	316.1	130	115.7	32	0.001	0.043	0.002	0.963	0.007	0.969
School Type (Public, Gifted, Private)										
Configural	259.9	132	-	-	-	0.043	-	0.978	-	0.972
Metric	291.3	152	56.6	20	0.002	0.042	0.001	0.976	0.002	0.974
Scalar	374.2	216	129.9	64	0.037	0.038	0.004	0.973	0.003	0.979

Note. χ2 = chi-square test; *df* = degrees of freedom; Δ = change in values between competing models (constrained and unconstrained models); CFI = comparative fit index; RMSEA = root mean square error of approximation; TLI = Tucker Lewis Index; change in χ2 was calculated using using the MODEL = CONFIGURAL METRIC SCALAR command in Mplus 7.4.