frontiers in Education

METHODS
published: 20 July 2021
doi: 10.3389/feduc.2021.659398

Analyses of the Factor Structure and Item Measurement Bias of a School Climate Scale in Chilean Students

Francisca Calderón[1,2]* *and Jorge González*[1,2]

[1]*Department of Statistics, Faculty of Mathematics, Pontificia Universidad Católica de Chile, Santiago, Chile,* [2]*Laboratorio Interdiciplinario de Estadística Social, LIES, Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Santiago, Chile*

School Climate is an essential aspect in every school community. It relates to perceptions of the school environment experienced by various members of the educational system. Research has shown that an appropriate school climate impacts not only on the quality of life of all members in the educational system, but also on learning outcomes and education improvements. This study aims to explore a measure of School Climate on Chilean students. A sample of 176,126 10th grade students was used to investigate the factor structure of the items composing the School Climate construct, and to evaluate the potential presence of Differential Item Functioning between male and female groups. Both explanatory and confirmatory factor analysis as well as Rasch models were used to analyze the scale. Differential item functioning between male and female groups was investigated using the Langer-improved Wald test. The results indicated a multidimensional structure of the School Climate construct and that measurement bias for male and female groups exist in some of the items measuring the construct.

Keywords: differential item functioning, school climate, gender DIF, IDPS Chilean data, factor analysis

OPEN ACCESS

Edited by:
Zi Yan,
The Education University of Hong Kong, China

Reviewed by:
Vanessa Scherman,
University of South Africa, South Africa
Rossana Gomez-Campos,
Catholic University of the Maule, Chile

***Correspondence:**
Francisca Calderón
flcalderon@mat.uc.cl

Specialty section:
This article was submitted to Assessment, Testing and Applied Measurement, a section of the journal Frontiers in Education

Received: *27 January 2021*
Accepted: *05 July 2021*
Published: *20 July 2021*

Citation:
Calderón F and González J (2021) Analyses of the Factor Structure and Item Measurement Bias of a School Climate Scale in Chilean Students. Front. Educ. 6:659398. doi: 10.3389/feduc.2021.659398

1 INTRODUCTION

Achievement tests developed by measurement programs have become relevant worldwide, as many decisions in educational policy are based on the information gathered using these instruments. In addition to measuring cognitive abilities (e.g., mathematics or language), these programs have incorporated measures of non-cognitive indicators such as perceptions and attitudes on academic motivation, school climate, participation and civic education, and healthy lifestyles, among others. This information is usually obtained from the answers to questionnaires administered to different participants of the educational system (i.e., teachers, students, parents). These measurements are useful not only to monitor non-academic aspects that are fundamental to students' general education, but also because they have an impact on learning outcomes and education improvement (McEvoy and Welker, 2000; Tubbs and Garner, 2008; Maxwell et al., 2017) (Høigaard et al., 2015; Cerda et al., 2019; Wang et al., 2020).

In Chile, the Agency of Quality of Education[1] (the Agency in what follows) is the governmental organization responsible for measuring cognitive and non-cognitive aspects in the educational system. Currently, together with the "Education Quality Measurement System" test (*Sistema de Medición de la Calidad de Educación*, SIMCE, by their initials in spanish) (Agencia de Calidad de la

[1]www.agenciaeducacion.cl

Educación, 2015), Chilean students, teachers and parents, express their perceptions and attitudes towards different non-academic aspects, through the Quality and Education Context (QEC) questionnaires (Agencia de Calidad de la Educación, 2017a). These questionnaires collect information on personal and social development indicators (*Indicadores de Desarrollo Personal y Social*, IDPS, for their initials in spanish) and are administered annually on a census application. The QEC questionnaires include dichotomous and polytomous (Likert-type) items related with four indicators: Academic Motivation, School Climate, Participation and Civic education, and Healthy lifestyles. The measured indicators have been defined by the Curriculum and Assessment Unit (Ministerio de Educación de Chile, 2014) and theoretically founded by a literature review in each of the assessed areas (Unidad de Currículum Evaluación - Ministerio de Educación, 2013). The Agency has made an effort to build a conceptual model for the constructs by incorporating multiple factors in each of the measured dimensions and establishing an evaluation frame that allows having operational definitions that contribute to the measurement process of the constructs (Agencia de Calidad de la Educación, 2017a). These multiple factors have been validated not only using appropriate literature reviews but also in internal workshops that considered the conceptual foundation of the constructs, together with the opinion of external experts in the topics. In this paper we focus our attention on the School Climate indicator.

School climate is an important aspect in every school community. Schools aim to identify features of the school environment that can be intervened to improve student outcomes. Schools initiatives seek for improving school climate as a previous step that lead to student academic and social well-being (Durlak et al., 2011). Although there is no clear consensus on a general definition of the construct, school climate is largely recognized as multidimensional, including the academic, community, safety, and institutional environment categorizations (Wang and Degol, 2016).

School climate and school coexistence have been researched from many different angles. For instance, Costa et al. (2020) describe the educational management of school coexistance at the intermediate level. The profile and role of professionals in charge of school coexistence in schools is investigated by Cortez Muñoz et al. (2018). Fierro-Evans and Carbajal-Padilla (2019) operationalize the concept of school climate in three areas of school life: pedagogical-curricular, organizational-administrative and socio-community, and elaborate on the relevance of having an instrument to evaluate school coexistence and its relationship with other variables or aspects. Wang and Degol (2016) analyzed the quality of interpersonal relationships within a school, both between teachers and authorities, and between teachers and students and presented a theoretical framework for the study of school climate that emphazise various features of the construct.

In 2011, the Chilean Law about School Violence (Ley, 2011), established the creation of school boards or committees of peaceful coexistence at educational centers as well as declares the obligation of having a person responsible for peaceful school-wide climate at every educational center. Since then, school climate has gained a protagonist role in educational practices and culture in spite of the wider variety of socioeconomic contexts and school practices and systems (Mineduc, 2011; Mineduc, 2015; Mineduc, 2019).

Research has shown that an appropriate school climate impacts not only on the quality of life of all members in the educational community, but also on learning outcomes and education improvements (see, e.g., Thapa et al., 2013; Zullig et al., 2010; Wang and Degol, 2016; Fan and Williams, 2018, among others). Moreover, evaluation of measurement invariance as a precondition for comparisons across groups becomes relevant, because male and female students' perceptions of school characteristics may differ (Bauer, 2017). In the context of school climate, any differences in item responses should reflect only respondents' perceptions of school climate factors, and not group membership (McGrath et al., 2020). Previous research show that school climate perception exhibits significant differences by gender and grade (Yates, 2001; Yates, 2003).

Some studies have analyzed the factor structure and measurement bias in non-cognitive tests (Prati et al., 2017; Lombardi et al., 2019; Tapia-Fonllem et al., 2020). A review of published studies until 2015 using school climate instruments that have been tested psychometrically is found in Ramelow et al. (2015). Previous research in Chile has analyzed scales adapted from other instruments and used in other countries, using only small samples of students from schools in Santiago, Chile (e.g., Cornejo and Redondo, 2001; López et al., 2014). Other related research including analyses on school-climate scales can be found in Muñoz et al. (2018) and Gálvez-Nieto et al. (2017). However, to date no studies exist that specifically look at the scale used in the Chilean national QEC questionnaires. This research aims to fill in this gap by studying the factor structure and item measurement bias in a School Climate scale of the national non-cognitive assessment carried out through the QEC questionnaires.

Internal consistency reliability was assessed using Cronbach's α. The structure of the School Climate item set was analyzed using both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Rasch analysis was applied to further analyse the quality of the instrument. The combined used of CFA and Rasch analyses have been successfully implemented in validation studies (Immekus et al., 2019; Yan, 2020). Item measurement bias (i.e., diferential item functioning, DIF) (Camilli, 2006) was evaluated using the Langer-improved Wald test for DIF (Langer, 2008; Wood et al., 2013).

2 MATERIALS AND METHODS

2.1 Instrument

The School Climate section of the QEC questionnaire is composed of 46 items and involves multiple factors following the conceptual scheme established by the Agency (see Agencia de Calidad de la Educación, 2017a, Table 3, p. 53). These factors are described as follows: perceptions and attitudes about respectful interactions among members of the educational community, the valuation of diversity, and the absence of discrimination inside the school. Students' perceptions regarding taking care of the school and respect for the environment. Opinions about the

TABLE 1 | Student school climate questionnaire Chile 2017.

Item	Question	Response scale format
	How much do you agree or disagree with each of the following statements	
p04_01	Students in my class keep the classroom clean	Agreement (4)
p04_02	My classmates help to keep the school clean	Agreement (4)
p04_03	Teachers encourage us to take care of things in the school	Agreement (4)
p05_01	The students in my class respect each other	Agreement (4)
p05_02	Teachers treat us all with respect	Agreement (4)
p05_03	The students in my class respect the teachers	Agreement (4)
p05_04	In my school everyone is treated with respect	Agreement (4)
	Have you felt discriminated against for any of the following reasons in your school	
p06_01	Because of your gender	Yes-No
p06_02	Because of some physical characteristic	Yes-No
p06_03	Because of your personality	Yes-No
p06_04	Because of your sexual orientation	Yes-No
p06_05	Because of the way you dress or your hairstyle	Yes-No
p06_06	Because of your religion or beliefs	Yes-No
p06_07	Because of your political ideas	Yes-No
p06_08	Because of your learning pace or style	Yes-No
p06_09	Because of a physical impairment	Yes-No
p06_10	Because of your family's financial situation	Yes-No
p06_11	Because of identifying as belonging to an indigenous community	Yes-No
p06_12	Because you are an immigrant or the child of immigrants	Yes-No
p06_13	Because of having children or being pregnant	Yes-No
p07	Do you know the school's code of conduct?	Yes-No
	How much do you agree or disagree with each of the following statements	
p08_01	The rules of coexistence are known to all students	Agreement (4)
p08_02	Students respect the rules of coexistence	Agreement (4)
p08_03	Sanctions stated in the coexistence manual are enforced	Agreement (4)
p08_04	Penalties are always applied fairly	Agreement (4)
p08_05	Teachers teach us the consequences of not meeting the standards	Agreement (4)
	What do you think of the following actions?	
p09_01	Cutting classes	Seriously (4)
p09_02	Skipping school	Seriously (4)
p09_03	Copying on tests	Seriously (4)
p09_04	Copying homework or tasks (for example, from a classmate, from the Internet, etc.).	Seriously (4)
	How often have the following situations occurred in your school?	
p10_01	Fights between students (for example, screaming, pushing, teasing, combos, etc.).	Frequency (4)
p10_02	Abuse between students (for example, insults, doodles, teasing, etc.).	Frequency (4)
p10_03	Threats or harassment between students	Frequency (4)
	Are you afraid that someone will harm you in the following places in your school?	
p11_01	School entrance and exit	Yes-No
p11_02	Classroom	Yes-No
p11_03	Hallways	Yes-No
p11_04	Schoolyard	Yes-No
p11_05	Bathrooms	Yes-No
	How often have you been bullied in any of the following ways by students from your school?	
p12_01	Physically: Hitting you or breaking your things	Frequency (5)
p12_02	Verbally: Insulting, mocking or threatening you	Frequency (5)
p12_03	Socially: Isolating you, speaking ill of you or humiliating you in front of others	Frequency (5)
p12_04	Electronically: Threatening, humiliating or mocking you with messages on the Internet.	Frequency (5)
	How much do you agree or disagree with each of the following statements	
p13_01	The teachers explain to us what to do in situations of abuse or intimidation among students	Agreement (4)
p13_02	We talked in class about the effects of bullying or intimidation among students	Agreement (4)
	What do you think of the following actions?	
p14_01	Damaging things in the establishment	Seriously (4)
p14_02	Insulting a classmate	Seriously (4)

Agreement (4): Strongly Disagree-Disagree-Agree-Strongly Agree. Frequency (4):Never or almost never - A few times -many times - always or almost always. Frequency (5):Never -A couple of times a year -A couple of times a month -Many times a week-Everyday. Seriously (4):Not serious at all -Somewhat serious -very serious-extremely serious.

existence of clear and well-known rules, demanded and respected by all, and the predominance of effective conflict resolution mechanisms. Attitudes that students have regarding the norms of coexistence and their transgression. The degree of security and physical and psychological violence at school, as well as the existence of prevention mechanisms and action against it. Attitudes against bullying and the factors that affect their physical or psychological integrity. **Table 1** shows the 46 items

TABLE 2 | Distribution of answers in each item category (in percentages).

Item	Strongly disagree	Disagree	Agree	Strongly agree	Missing	Corr item-test
p04_01	15.35	41.20	38.08	5.37	0.58	0.49
p04_02	19.37	45.01	30.90	4.72	0.67	0.52
p04_03	2.96	7.60	48.14	41.30	0.82	0.53
p05_01	7.47	26.10	52.76	13.67	0.58	0.58
p05_02	1.78	8.43	48.07	41.72	0.68	0.58
p05_03	4.26	23.27	55.73	16.74	0.81	0.59
p05_04	9.23	32.85	42.51	15.40	0.70	0.62
Item	**Yes**	**No**			**Missing**	**Corr item-test**
p06_01	3.47	96.53			0.51	0.51
p06_02	17.68	82.32			0.54	0.45
p06_03	21.56	78.44			0.57	0.43
p06_04	2.75	97.25			0.62	0.47
p06_05	10.51	89.49			0.64	0.47
p06_06	4.08	95.92			0.65	0.43
p06_07	4.69	95.31			0.66	0.47
p06_08	10.58	89.42			0.69	0.44
p06_09	1.55	98.45			0.81	0.55
p06_10	2.51	97.49			0.72	0.54
p06_11	1.42	98.58			0.92	0.53
p06_12	1.34	98.66			1.08	0.50
p06_13	0.79	99.21			1.43	0.65
p07	46.94	53.06			1.56	0.38
Item	**Strongly disagree**	**Disagree**	**Agree**	**Strongly Agree**	**Missing**	**Corr item-test**
p08_01	10.81	44.76	36.14	8.29	0.60	0.52
p08_02	6.98	42.49	46.07	4.47	0.78	0.61
p08_03	3.37	13.44	57.37	25.82	0.98	0.57
p08_04	10.00	27.10	45.16	17.73	1.11	0.6
p08_05	4.83	15.50	55.32	24.34	0.93	0.54
Item	**No serious at all**	**Somewhat serious**	**Very serious**	**Extremely serious**	**Missing**	**Corr item-test**
p09_01	4.38	38.01	40.21	17.40	0.59	0.43
p09_02	2.39	9.47	31.86	56.28	0.66	0.44
p09_03	11.24	41.11	33.10	14.55	1.52	0.41
p09_04	21.28	46.59	22.50	9.63	1.82	0.40
Item	**Always or almost always**	**a lot of times**	**Few times**	**Never or almost never**	**Missing**	**Corr item-test**
p10_01	2.93	10.56	46.23	40.29	0.59	0.49
p10_02	15.07	30.11	39.86	14.96	0.70	0.54
p10_03	3.60	10.17	34.09	52.14	0.79	0.54
Item	**Yes**	**No**			**Missing**	**Corr item-test**
p11_01	9.37	90.63			0.73	0.40
p11_02	4.04	95.96			0.84	0.57
p11_03	5.10	94.90			0.87	0.54
p11_04	5.04	94.96			0.80	0.54
p11_05	6.85	93.15			0.78	0.48

Item	Everyday	Many times a week	A couple of times a month	A couple of times a year	Never	Missing	Corr item-test
p12_01	0.31	0.54	1.42	8.10	89.62	0.60	0.56
p12_02	1.06	2.57	4.89	20.83	70.65	0.71	0.56
p12_03	0.95	2.15	4.43	16.80	75.68	2.14	0.52
p12_04	0.47	0.73	1.65	6.90	90.24	0.69	0.56

Item	Strongly disagree	Disagree	Agree	Strongly Agree	Missing	Corr item-test
p13_01	9.40	21.86	48.87	19.87	0.86	0.53
p13_02	11.57	25.44	44.94	18.06	0.99	0.51
Item	**No serious at all**	**Somewhat serious**	**Very serious**	**Extremely serious**	**Missing**	**Corr item-test**
p14_01	2.53	17.44	48.90	31.13	0.83	0.40
p14_02	5.23	25.51	43.56	25.71	0.97	0.39

that compose the scale and the corresponding response scale format used to answer them.

2.2 Data

We use data collected by the Chilean Agency of Quality of Education through the administration of the QEC questionnaires to students in 2017 (Agencia de Calidad de la Educación, 2017b). For the analyses that follow, we use the responses of students on the School Climate indicator.

A total of 200, 525 10th grade students in 2017 answered the questionnaire (proportion of male is 49.75%, and female is 50.25%). Item *p*12_03 (measuring how often the student was socially bullied) presented the highest percentage of missingness (2.14%). The rest of items had less than 2% of missing responses. The final working data set had 176, 126 records and considered only the complete cases. **Table 2** shows the proportions of answers in each item category.

2.3 Statistical Analysis

The adequacy of items measuring the school climate construct was first evaluated using item-test correlations. Internal consistency reliability was assessed using Cronbach's α (Cronbach, 1951).

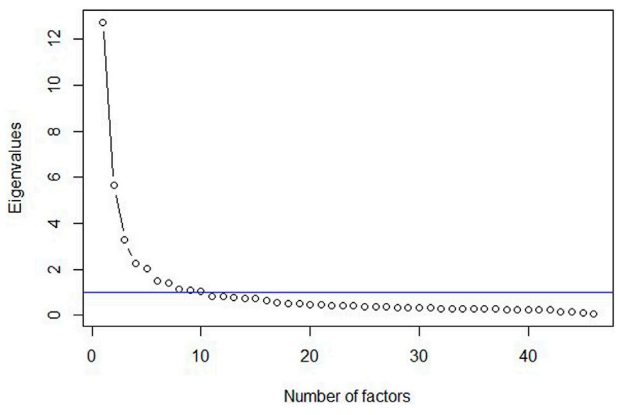

FIGURE 1 | Scree plot for the students QEC questionnaire.

EFA was performed using the correlation matrix of the 46 school climate items. Preliminary analyses to assess the suitability of the data for factor analysis included the Kaiser–Meyer–Olkin measure (Kaiser, 1974a) and Bartlett's test of sphericity (Bartlett, 1951).

The decision on the number of factors to retain was make using a scree-plot (Cattell, 1966), in which the eigenvalues are plotted against their rank, and the number of factors to retain is indicated by the *elbow* of the curve (Joreskog, 2007).

CFA with maximum likelihood (ML) estimation was performed to test the structural validity under two scenarios: 1) a model was fitted to assess the structure found in the EFA analysis; and 2) a second model was fitted to test whether the multiple factors defined by the Agency followed a three-factor model with three main dimensions: Respectful environment, Organized environment, and Safe environment.

The evaluation of model-data fit was analysed using the comparative fit index (CFI; Bentler, 1990), Tucker-Lewis index (TLI; Bentler and Bonett, 1980; Tucker and Lewis, 1973), and the root mean square error of approximation (RMSEA; Steiger, 1990; Steiger and Lind, 1980). RMSEA is an absolute fit index in that it assesses how far a hypothesized model is from a perfect model. On the contrary, CFI and TLI are incremental fit indices that compare the fit of a hypothesized model with that of a baseline model (i.e., a model with the worst fit) (Yan and Yanyun, 2018).

Rasch analyses (Wright and Masters, 1982) were conducted to further analyze the quality of the scale. Different models were evaluated in terms of global measures of fit (e.g., AIC, BIC, log-likelihood); item fits statistics (Infit and Outfit MNSQ); and reliability (i.e., EAP reliabilities).

In order to examine DIF between male and female groups, we used the Langer-improved Wald test (Langer, 2008; Cai et al., 2011; Cai, 2012). The Langer-improved Wald test statistic is implemented in IRTPro (Cai et al., 2011) and flexMIRT (Cai, L., 2017) software.

Reliability, item correlation, exploratory factor analyses and Rasch analyses were performed using R (R Core Team, 2020), CFA was implemented using Mplus version 8.4 (Muthén and

TABLE 3 | Results for EFA with nine factors.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
SS loadings	4.607	4.316	2.681	2.627	2.596	2.213	1.866	1.690	1.413
Proportion Var	0.100	0.094	0.058	0.057	0.056	0.048	0.041	0.037	0.031
Cumulative Var	0.100	0.194	0.252	0.309	0.366	0.414	0.454	0.491	0.522

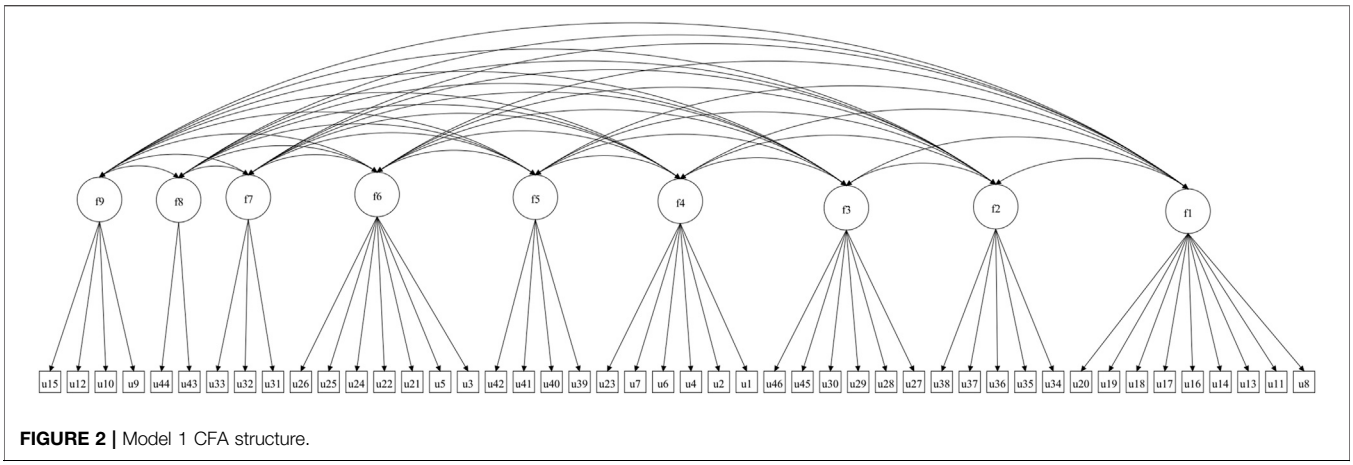

FIGURE 2 | Model 1 CFA structure.

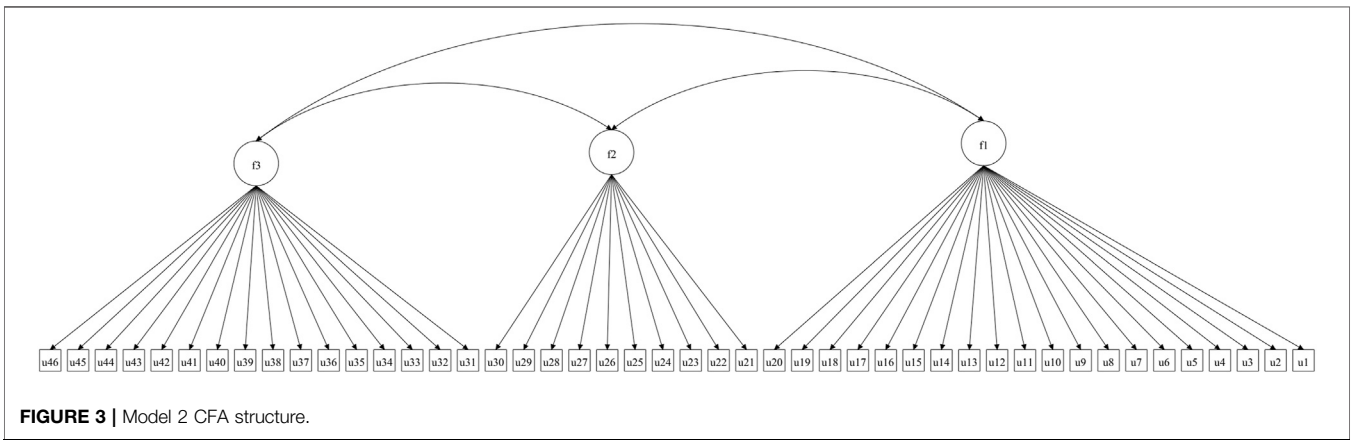

FIGURE 3 | Model 2 CFA structure.

Muthén, 2007), and DIF analyses were carried out using the flexMIRT software (Cai, L., 2017).

3 RESULTS

3.1 Item-Test Correlation and Reliability

Olsson et al. (1982) considered values of polyserial correlation as low ($\rho = 0.25$), moderate ($\rho = 0.50$) and high ($\rho = 0.75$). **Table 2** shows the item-test correlation for each item. The item-test correlations for the polytomously scored items ranged from 0.39 to 0.62 (median 0.53) so no items were deleted from the data set. The biserial correlation was computed for items scored in two categories with values ranging from 0.38 to 0.65 (median 0.48), so all the items were retained for further analyses (Kline, 2005).

Following Nunnally and Bernstein (1994), Cronbach's alpha coefficients $< .50$ are considered insufficient, $.50 - .69$ moderate, $.70 - .79$ satisfactory and $\geq .80$ good. A good measure of reliability was obtained for the school climate indicator (46 items $\alpha = 0.88$) (Cortina, 1993).

3.2 Exploratory Factor Analysis

The Bartlett's test of sphericity was significant ($X^2_{(1035)}$=31024.28, $p < 0.001$). Kaiser (1974b) suggested that KMO > 0.9 were marvelous, in the 0.80*s*, meritorious, in the 0.70*s*, middling, in the 0.60*s*, mediocre, in the 0.50*s*, miserable, and less than 0.5, unacceptable. The Kaiser–Meyer–Olkin value in our analysis was 0.93, indicating that the items were appropriate for factor analysis.

The correlation matrix was computed using the mixedCor function (Revelle, 2020) in R. This function allows the combination of continuous, polytomous and dichotmous variables providing the appropriate correlation in each case (i.e., polychorics for the polytomous items, tetrachorics for the dichotomous items, and the polyserial or biserial correlations for the various mixed variables).

EFA with oblique rotation (CF-quartimax) was applied to reduce the large dimensionality in this data set. **Figure 1** shows the scree plot which suggests a nine-factor structure. The nine-factor solution explained 52.2% of the common variance. The percentage of explained variance by each factor is shown in **Table 3**.

TABLE 4 | Fit indices for CFA analyses.

Model	RMSEA	CFI	TLI
Model 1 (9-dimensional)	0.040	0.928	0.922
Model 2 (3-dimensional)	0.087	0.647	0.629

TABLE 5 | Correlation between the nine latent traits.

	F1	F2	F3	F4	F5	F6	F7	F8
(F1) Tolerance of diversity								
(F2) Perception to be afraid inside school	0.222							
(F3) Attitude towards cheating	0.056	–0.022						
(F4) Respect for others and for school facilities	0.164	0.169	0.215					
(F5) Experience with bullying	0.309	0.317	0.052	0.259				
(F6) Respect for the student code of conduct	0.151	0.102	0.356	0.518	0.157			
(F7) Experience with violence	0.167	0.226	0.116	0.443	0.347	0.276		
(F8) Teacher support in dealing with bullying	0.098	0.066	0.235	0.311	0.108	0.464	0.154	
(F9) Experience with discrimination	0.429	0.278	0.050	0.252	0.475	0.173	0.235	0.121

While some might argue that less than sixty percent of explained variance is not acceptable (Hair et al., 2013, p. 107), and that nine factors add complexity to the practical interpretation, we have combined both empirical and theoretical foundations to report the results of EFA as a model to be tested for validity, not only focusing on the percentage of explained variance and number of factors, but rather attempting to offer a practical explanation and interpretation taking into account the conceptual multi-factor scheme adopted by the Agency for the definition of the school climate construct. As a matter of fact, the nine factors can be named and interpreted as follows (see also **Table 3**):

F1 *Tolerance of diversity* produced an eigenvalue of 12.766 and explained 10.0% of the common variance. F2 *Perception to be afraid inside school* produced an eigenvalue of 5.640 and explained 9.4% of the common variance. F3 *Attitude towards cheating* produced an eigenvalue of 3.269 and explained 5.8% of the common variance. F4 *Respect for others and for school facilities* produced an eigenvalue of 2.235 and explained 5.7% of the common variance. F5 *Experience with bullying* produced an eigenvalue of 2.032 and explained 5.6% of the common variance. F6 *Respect for the student code of conduct* produced an eigenvalue of 1.504 and explained 4.8% of the common variance. F7 *Experience with violence* produced an eigenvalue of 1.397 and explained 4.1% of the common variance. F8 *Teacher support in dealing with bullying* produced an eigenvalue of 1.146 and explained 3.7% of the common variance. And finally, F9 *Experience with discrimination* produced an eigenvalue of 1.098 and explained 3.1% of the common variance.

Summarizing, the results lead to conclude that the School Climate scale appears to be multidimensional. In fact, these results are somewhat expected in that for constructs measured based on a theory that incorporates multiple factors in each of the measured dimensions, a multidimensional structure can be anticipated. The structural validity of the 9-factor model is tested using CFA in the next section.

TABLE 6 | Goodness of fit indices for three different Rasch models.

Model	Log-likelihood	AIC	BIC
Unidimensional	–30432.45	61164.90	61900.77
3-dimensional	–29902.69	60111.38	60861.97
9-dimensional	–29440.69	59181.37	59917.23

3.3 Confirmatory Factor Analysis

Two different models were evaluated. Model 1 (see **Figure 2**) is guided by the results of the EFA and considered the nine factor solution

TABLE 7 | Infit and outfit statistics for the 9-dimensional Rasch model.

	Item	Outfit	Infit
1	p04_01	0.834	0.837
2	p04_02	0.842	0.846
3	p04_03	0.842	0.882
4	p05_01	0.764	0.791
5	p05_02	0.843	0.887
6	p05_03	0.670	0.705
7	p05_04	0.840	0.802
8	p06_01	0.195	1.255
9	p06_02	0.335	0.976
10	p06_03	0.351	1.017
11	p06_04	0.206	1.319
12	p06_05	0.217	0.702
13	p06_06	0.188	1.252
14	p06_07	0.212	1.377
15	p06_08	0.339	0.963
16	p06_09	0.072	0.550
17	p06_10	0.071	0.604
18	p06_11	0.043	0.468
19	p06_12	0.045	0.449
20	p06_13	0.028	0.335
21	p07	0.984	1.003
22	p08_01	0.894	0.919
23	p08_02	0.834	0.836
24	p08_03	0.655	0.744
25	p08_04	0.683	0.721
26	p08_05	0.820	0.806
27	p09_01	0.652	0.707
28	p09_02	0.618	0.709
29	p09_03	0.760	0.781
30	p09_04	0.824	0.847
31	p10_01	0.504	0.635
32	p10_02	0.614	0.729
33	p10_03	0.401	0.521
34	p11_01	0.164	1.200
35	p11_02	0.132	0.817
36	p11_03	0.049	0.507
37	p11_04	0.430	0.661
38	p11_05	0.109	0.865
39	p12_01	0.257	0.579
40	p12_02	0.212	0.491
41	p12_03	0.309	0.718
42	p12_04	0.230	0.702
43	p13_01	0.294	0.431
44	p13_02	0.394	0.536
45	p14_01	0.832	0.857
46	p14_02	0.894	0.907

TABLE 8 | Factor loadings for EFA on the students QEC questionnaire.

Dimension	Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
Respectful environment	p04_01				0.787					
	p04_02				0.756					
	p04_03						0.324			
	p05_01				0.630					
	p05_02						0.337			
	p05_03				0.624					
	p05_04				0.462					
	p06_01	0.502								
	p06_02									0.381
	p06_03									0.544
	p06_04	0.528								
	p06_05									0.435
	p06_06	0.553								
	p06_07	0.543								
	p06_08									0.407
	p06_09	0.714								
	p06_10	0.608								
	p06_11	0.837								
	p06_12	0.828								
	p06_13	0.893								
Organized environment	p07						0.225			
	p08_01						0.407			
	p08_02				0.451					
	p08_03						0.766			
	p08_04						0.752			
	p08_05						0.626			
	p09_01			0.848						
	p09_02			0.773						
	p09_03			0.678						
	p09_04			0.628						
Safe environment	p10_01							0.837		
	p10_02							0.680		
	p10_03							0.773		
	p11_01		0.857							
	p11_02		0.823							
	p11_03		0.981							
	p11_04		0.972							
	p11_05		0.932							
	p12_01					0.801				
	p12_02					0.866				
	p12_03					0.708				
	p12_04					0.690				
	p13_01								0.906	
	p13_02								0.842	
	p14_01			0.497						
	p14_02			0.446						

● *Male group.* ● *Female group.*

described in the previous section. The second model, Model 2 (see **Figure 3**), represents the hypothesized structure of three main dimensions of the School Climate construct: respectful environment, organized environment and, safe environment. RMSEA values less than 0.05 are considered good, between 0.05 and 0.08 are considered adequate, and between 0.08 and 0.10 are considered mediocre (Schermelleh-Engel et al., 2003). Values of CFI larger than 0.95 and TLI more than 0.95 indicate a good fit. An acceptable fit is indicated by a CFI value larger than 0.90 and a TLI large than 0.90 (McDonald and Ho, 2002). **Table 4** shows the RMSEA, CFI and TLI index. It can be seen that Model 1 not only fitted better than Model 2, but also its model-data fit is satisfactory according to the fit indexes (Hu and Bentler, 1999). **Table 5** shows the correlations among person latent trait estimates on the nine factors. As expected, (F1) *Tolerance of diversity*, (F5) *Experience with bullying* and, (F9) *Experience with discrimination* were significantly associated. All the items composing these factors are related to questions about tolerance and discrimination. Factors composed of items regarding respect were also highly correlated, i.e., (F4) *Respect for others and for school facilities*, (F6) *respect for the student code of conduct*, and (F7) *Experience with violence*. On the other hand, (F2) *Perception to be afraid inside school* and (F3) *Attitude towards cheating* had a negative but very low association, which means that both factors are dissociated.

3.4 Rasch Analysis

In line with the strategy adopted for the CFA analyses, two multidimensional Rasch models were fitted and compared. An

TABLE 9 | The Langer-improved Wald test X^2.

Item	Total X^2	df	p	X^2 (a)	df	p	X^2 (b|a)	df	p
p06_01	897.9	2	0.0001	130.5	1	0.0001	767.4	1	0.0001
p06_02	305.5	2	0.0001	42	1	0.0001	263.5	1	0.0001
p06_03	838.8	2	0.0001	5.7	1	0.0168	833.1	1	0.0001
p06_04	228.9	2	0.0001	73.9	1	0.0001	155	1	0.0001
p06_05	47	2	0.0001	33.5	1	0.0001	13.5	1	0.0002
p06_06	138.6	2	0.0001	69.8	1	0.0001	68.8	1	0.0001
p06_07	2.8	2	0.2493	2.4	1	0.1257	0.4	1	0.5127
p06_08	437.4	2	0.0001	14	1	0.0002	423.4	1	0.0001
p06_09	57.3	2	0.0001	34.8	1	0.0001	22.5	1	0.0001
p06_10	81.7	2	0.0001	38.5	1	0.0001	43.2	1	0.0001
p06_11	104.9	2	0.0001	35.8	1	0.0001	69.1	1	0.0001
p06_12	100.8	2	0.0001	40	1	0.0001	60.8	1	0.0001
p06_13	111.6	2	0.0001	66.7	1	0.0001	44.9	1	0.0001

unidimensional model was also fitted as a reference for comparisons. **Table 6** shows log-likelihood (larger values are preferable), AIC and BIC (smaller values are preferable) fit statistics indicating that the 9-dimensional model has the best fit, followed by the 3-dimensional and the unidimensional model.

According to Linacre (2002) items showing infit and outfit values > 2.0 distorts or degrades the measurement system; 1.5 – 2.0 are unproductive for construction of measurement, but not degrading; 0.5 – 1.5 are productive for measurement; < 0.5 are less productive for measurement, but not degrading. The inspection of item fit statistics (infit and outfit) indicated that not all items fitted well to the model. As it can be seen from **Table 7**, only 45.7% of the items (21 out of 46) have an outfit value between 0.5 and 1.5 considered acceptable. On the other hand 89.1% (41 out of 46) items presented an acceptable infit value. In all cases when the item fit statistics were outside of the acceptable range 0.5–1.5, the values were lower than 0.5 and thus the items can be considered to be less productive for measurement but not capable to degrade the measurement system (Linacre, 2002). However, these results should be interpreted with caution in that items could not be measuring each dimension in a proper way.

The EAP reliabilities for each of the 9 factors (F1-F9) were 0.25, 0.43, 0.79, 0.80, 0.60, 0.80, 0.71, 0.79, and 0.42, respectively. The low values for F1, F2 and F9 are in line with the fact that items composing this factors shown poor fit statistics.

3.5 Differential Item Functioning

As a way to combine both EFA and DIF analyses, given that we are interested in assessing possible differences between male's and female's perception on school climate, we studied the factor structure for both groups separately. **Table 8** shows the highest factor loading for each item. In this table, the blue dots represent the highest factor loading for each item for the male group whereas the red dots represent the highest factor loading for each item for the female group. In general, it can be seen that the factor structure is the same for both groups, except for items regarding discrimination and tolerance of diversity (F9 and F1). Given that these items have a different factor structure for males and females it is of interest to explore if both groups have the same understanding about these questions. This goal is achieved by evaluating DIF.

A DIF analysis on the items *p*06_01 – *p*06_13 was conducted using the Langer-improved Wald method. This method is based on the comparison of the parameter estimates obtained from the fit of an item response theory (IRT) model so what is tested is the difference of IRT item parameter estimates between groups. **Table 9** shows the Langer-improved Wald X^2 to test the equality of item parameters. For this subset of items, it can be seen that there are significant differences ($p \leq 0.001$) in almost all items indicating the presence of DIF. An exception is item *p*06_07 (about feeling discriminated because of political ideas) for which the test turns to be non-significant. For illustrative purposes, a graphical representation of the tested differences is shown in **Figure 4** for items *p*06_01, *p*06_07 and *p*06_11. This figure shows the Item Characteristics Curves (ICC) which represent the probability to choose the *No* option given the level of *School Climate perception* for male and female groups separately. While no apparent differences are seen for Item *p*06_07, the ICC for items *p*06_01 and *p*06_11 show differences in responding for male and females. Interestingly, the ICC for item *p*06_01 (about feel discriminated because of your gender) shows a larger propensity of a "No" answer for the female group with a low level of school climate perception. When the level of school climate perception increases, such larger propensity is seen for the male group, phenomenon known as non-uniform DIF. All the other items, including *p*06_11, shown uniform DIF (i.e., no differences in propensity to answer at different levels of the measured construct).

4 DISCUSSION

We have analysed the factor structure and item measurement bias (DIF) in a Chilean School Climate scale which is part of a large study including other non-cognitive indicators. We adopted an strategy that combined both empirical and theoretical evidence to select the number of factors composing the scale.

Although the found scale can appear to be highly multidimensional, our findings are not only aligned with the

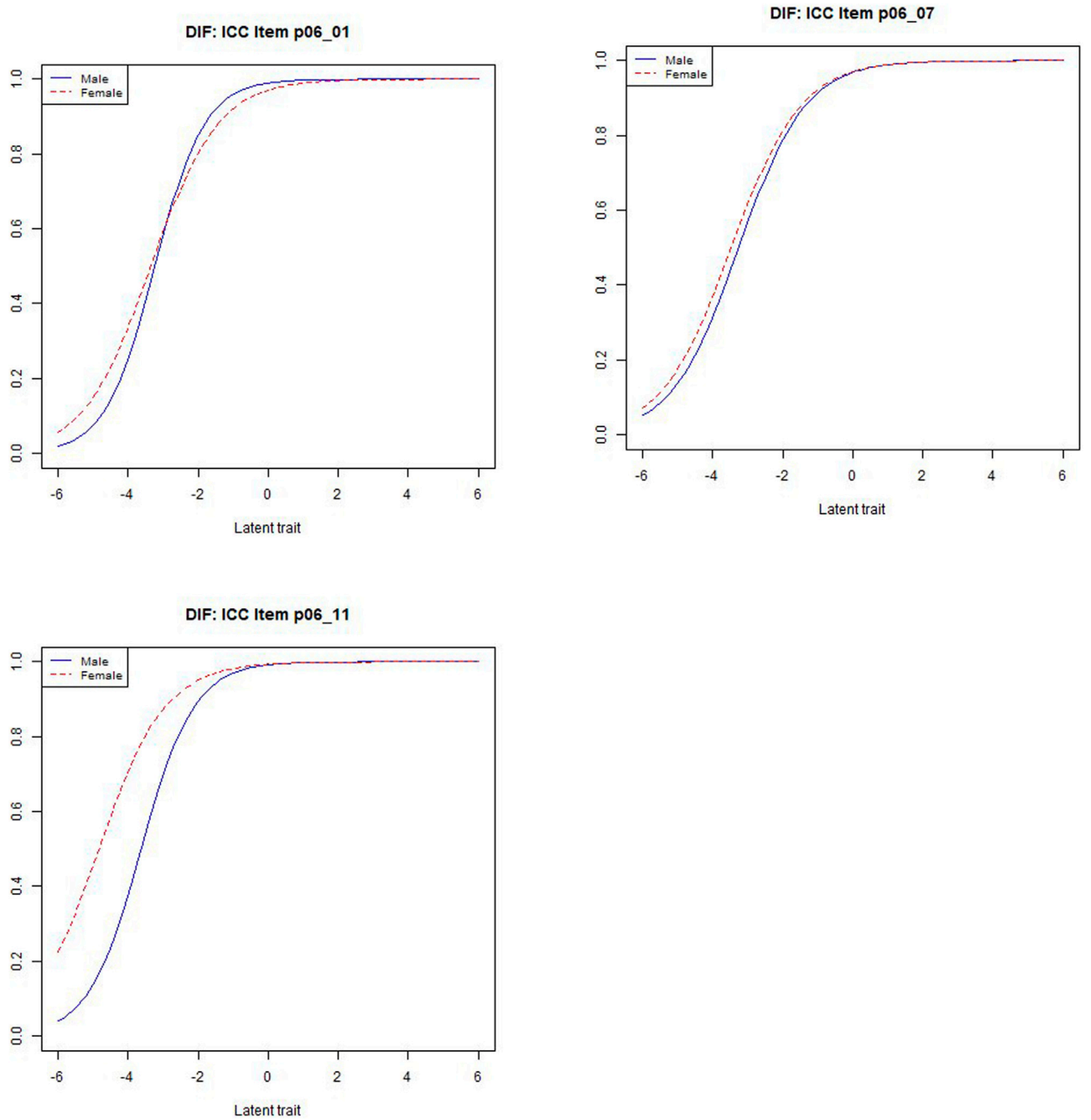

FIGURE 4 | ICCs by gender for three items of the students QEC questionnaire.

literature in that School Climate is widely recognized as a multidimensional factor (Wang and Degol, 2016; Grazia and Molinari, 2020), but also the derived results were shown to be useful to provide clear interpretations of each dimension which is in line with the multi-factor conceptual scheme adopted by the Agency to define the IDPS and to plan its use. Each of the found dimensions have an impact on the school climate in terms of *what*, *how*, *what for* and *with whom* the students learn to coexist and interact. Although the school climate construct can be conceptualized as composed of three main dimensions, its impact transcends other dimensions as the ones found in this study. Moreover, our findings show that the three dimensions, Respectful environment, Organized environment, and Safe environment, can not be treated separately but rather they are interrelated through the items composing the corresponding factors (i.e., items in a particular dimension are more related to items from another dimension than their own dimension; see **Tables 5** and **8**). This fact can be seen as a signal to re-organize these three dimensions, recognizing that more than

three are needed to measure the school climate construct. In adition, the final CFA model fitted very well, confirming the multi-factor theoretical structure defined by the conceptual model.

It should be noted that other studies assessing the factorial structure of school climate scales have found highly multidimensional structures. For instance, based on students' perception, Muñoz et al. (2018) confirmed that coexistence in school is a multidimensional phenomenon that is best explained by a group of eight interrelated dimensions. Zullig et al. (2010) reported factor analysis results confirming an eight-factor solution for a student-reported school climate instrument. Brand et al. (2003) produced a 50-item instrument that assessed 10 distinct dimensions of school climate, consisting of Disciplinary Harshness, Negative Peer Interactions, Positive Peer Interactions, Structure and Clarity of Rules and Expectations, Student Commit-ment to Achievement, Teacher Support, Instructional Innovation, Student Participation in Decision Making, Support for Cultural Pluralism, and Safety Problems. Moreover, these authors reported that the results of CFA across two different years of samples and students suggested that the 10-dimensional structure was robust.

The Rasch analyses complemented the CFA results and supported the multidimensional nature of the scale. However, the results of the item fit analyses showed that not all the items would be measuring each sub-dimension satisfactorily. Interestingly, the items showing misfit are those related to questions about discrimination. These items presented a highly disproportionate number of answers in one category (see **Table 2**) leading to low variability, which could be the reason of poor item fit. Because subscales may be affected by the way questions about discrimination are formulated, our findings are useful to recommend alternatives. For instance, a revised version of the QEC questionnaires could consider different response scales for these items. Collapsing the number of items in order to gain variability and obtain the needed information is another possibility.

Regarding item measurement bias, the analyses indicated the presence of DIF among male and female students in items related to the Discrimination and Tolerance of diversity dimensions. These results are in concordance with those derived from the EFA analyses performed separately for male and females. They are also aligned with previous research that have shown significant differences by gender and grade on school climate perception (Yates, 2001; Yates, 2003).

In this study we have only used the information gathered from students. This choice was mainly motivated by the fact that all the other indicators measured by the QEC questionnaire are also measured on students and, future studies are planned to include these indicators. Moreover, because the School Climate indicator is measured on the three actors (students, teachers and parents), we also plan to extend our analysis to analyze whether the factorial structure found for the School Climate indicator is invariant across actors.

AUTHOR CONTRIBUTIONS

FC contributed to the conception, design, and analysis of data and drafted and revised the manuscript; JG contributed to critically revising the manuscript. FC and JG wrote sections of the paper. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

The FC acknowledges the support by ANID (Agencia Nacional de Investigación y Desarrollo) Chile. Doctorado Nacional Fellowship 2017 - 21171096. The JG acknowledges support of Fondecyt grant 1201129. This research used the databases of the Agency of Quality of Education (ACE) as a source of information. The author thanks the ACE for access to the information. All the results of the study are the responsibility of the author and do not commit the ACE in any way.

REFERENCES

Agencia de Calidad de la Educación (2017a). Informe técnico 2017. indicadores de desarrollo personal y social (idps) medidos a través de cuestionarios. Avilable at: http://archivos.agenciaeducacion.cl/Informe_tecnico_IDPS_2017.pdf. (accessed July 3, 2020).

Agencia de Calidad de la Educación (2017b). Base de datos de la Agencia de Calidad de la Educación 2017. Santiago, Chile.

Agencia de Calidad de la Educación (2015). Informe Técnico Simce 2015. Avilable at: http://archivos.agenciaeducacion.cl/Informe_Tecnico_SIMCE_2015_Final.pdf. (accessed July 3, 2020).

Bartlett, M. S. (1951). THE EFFECT OF STANDARDIZATION ON A $\chi 2$ APPROXIMATION IN FACTOR ANALYSIS. *Biometrika* 38, 337–344. doi:10.1093/biomet/38.3-4.337

Bauer, D. J. (2017). A More General Model for Testing Measurement Invariance and Differential Item Functioning. *Psychol. Methods* 22, 507–526. doi:10.1037/met0000077

Bentler, P. M. (1990). Comparative Fit Indexes in Structural Models. *Psychol. Bull.* 107, 238–246. doi:10.1037/0033-2909.107.2.238

Bentler, P. M., and Bonett, D. G. (1980). Significance Tests and Goodness of Fit in the Analysis of Covariance Structures. *Psychol. Bull.* 88, 588–606. doi:10.1037/0033-2909.88.3.588

Brand, S., Felner, R., Shim, M., Seitsinger, A., and Dumas, T. (2003). Middle School Improvement and Reform: Development and Validation of a School-Level Assessment of Climate, Cultural Pluralism, and School Safety. *J. Educ. Psychol.* 95, 570–588. doi:10.1037/0022-0663.95.3.570

Cai, L. (2012). *Flexmirt: Flexible Multilevel Item Factor Analysis and Test Scoring [computer Software]*. Seattle, WA: Vector Psychometric Group, LLC.

Cai, L. (2017). flexMIRT© Version 3.51: Flexible Multilevel Multidimensional Item Analysis and Test Scoring [Computer Software]. NC: Champel HillVector Pshychometric Group.

Cai, L., Thissen, D., and du Toit, S. H. C. (2011).*Irtpro: Flexible, Multidimensional, Multiple Categorical Irt Modeling [computer Software]*. Lincolnwood, IL: Scientific Software International.

Camilli, G. (2006). "Test Fairness,". *Educational Measurement*. Editor R. Brennan (Westport, CT: American Council on Education), 4, 220–256.

Cattell, R. B. (1966). The Scree Test for the Number of Factors. *Multivariate Behav. Res.* 1, 245–276. doi:10.1207/s15327906mbr0102_10

Cerda, G., Pérez, C., Elipe, P., Casas, J. A., and Del Rey, R. (2019). School Coexistence and its Relationship with Academic Performance Among Primary Education Students. *Revista de Psicodidáctica (English ed.)* 24, 46–52. doi:10.1016/j.psicoe.2018.06.001

Cornejo, R., and Redondo, J. M. (2001). El clima escolar percibido por los alumnos de enseñanza media: Una investigación en algunos liceos de la Región Metropolitana. *Ultima década* 9, 11–52. doi:10.4067/s0718-22362001000200002

Cortez Muñoz, M., Zoro Sutton, B., and Aravena Castillo, F. (2019). Gestionando la contingencia más que la convivencia: El rol de los encargados de convivencia escolar en chile. *Psicoperspectivas* 18, 18–32.

Cortina, J. M. (1993). What Is Coefficient Alpha? an Examination of Theory and Applications. *J. Appl. Psychol.* 78, 98–104. doi:10.1037/0021-9010.78.1.98

Costa, P. A., Mancilla, K. C., Apellidos, F., and Figueroa, Á. (2020). Gestión de la convivencia escolar a nivel intermedio: Diseño y validación de una escala. *Revista Evaluar* 20, 1–19.

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *psychometrika* 16, 297–334. doi:10.1007/bf02310555

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., and Schellinger, K. B. (2011). The Impact of Enhancing Students' Social and Emotional Learning: a Meta-Analysis of School-Based Universal Interventions. *Child. Dev.* 82, 405–432. doi:10.1111/j.1467-8624.2010.01564.x

Fan, W., and Williams, C. (2018). The Mediating Role of Student Motivation in the Linking of Perceived School Climate and Achievement in reading and Mathematics. *Front. Educ.* 3, 50. doi:10.3389/feduc.2018.00050

Fierro-Evans, C., and Carbajal-Padilla, P. (2019). Convivencia escolar: Una revisión del concepto. *Psicoperspectivas* 18, 9–27.

Gálvez-Nieto, J. L., Salvo, S., Pérez-Luco, R., Hederich, C., and Trizano-Hermosilla, Í. (2017). Invarianza factorial del Cuestionario para Evaluar Clima Social del Centro Escolar en estudiantes chilenos y colombianos. *Revista Latinoamericana de Psicología* 49, 119–127. doi:10.1016/j.rlp.2016.09.003

Grazia, V., and Molinari, L. (2020). School Climate Multidimensionality and Measurement: a Systematic Literature Review. *Res. Pap. Education*, 1–27. doi:10.1080/02671522.2019.1697735

Hair, J. F., Black, W. C., Babin, B. J., and Anderson, R. E. (2013). *Multivariate Data Analysis*. (Pearson Education Limited).

Høigaard, R., Kovač, V. B., Øverby, N. C., and Haugen, T. (2015). Academic Self-Efficacy Mediates the Effects of School Psychological Climate on Academic Achievement. *Sch. Psychol. Q.* 30, 64.

Hu, L. t., and Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Struct. Equation Model. A Multidisciplinary J.* 6, 1–55. doi:10.1080/10705519909540118

Immekus, J. C., Snyder, K. E., and Ralston, P. A. (2019). Multidimensional Item Response Theory for Factor Structure Assessment in Educational Psychology Research. *Front. Educ.* 4, 45. doi:10.3389/feduc.2019.00045

Joreskog, K. G. (2007). "Factor Analysis and its Extensions," in *Factor Analysis at 100: Historical Developments and Future Directions*. Editors R. Cudeck and R. C. MacCallum (Lawrence Erlbaum Associates Publishers), 47–77.

Kaiser, H. F. (1974a). An index of Factorial Simplicity. *Psychometrika* 39, 31–36. doi:10.1007/BF02291575

Kaiser, H. F. (1974b). An index of Factorial Simplicity. *Psychometrika* 39, 31–36. doi:10.1007/bf02291575

Kline, T. (2005). "Classical Test Theory: Assumptions, Equations, Limitations, and Item Analyses," in *Psychological Testing: A Practical Approach to Design and Evaluation* (Thousand Oaks, California: SAGE Publications, Inc.), 91–106.

Langer, M. (2008). *A Reexamination of Lord's Wald Test for Differential Item Functioning Using Item Response Theory and Modern Error Estimation*. University of North Carolina, Chapel Hill. doi:10.17615/chn0-dz45 (Unpublished doctoral dissertation).

Ley, N. (2011). *20.536 Violencia Escolar*. Santiago, Chile: Ministerio de Educación.

Linacre, J. (2002). What Do Infit and Outfit, Mean-Square and Standardized Mean? *Rasch Meas. Trans.* 16, 878.

Lombardi, E., Traficante, D., Bettoni, R., Offredi, I., Giorgetti, M., and Vernice, M. (2019). The Impact of School Climate on Well-Being Experience and School Engagement: A Study with High-School Students. *Front. Psychol.* 10, 2482. doi:10.3389/fpsyg.2019.02482

López, V., Bilbao, M., Ascorra, P., Moya Diez, I., and Morales, M. (2014). School Climate Scale: Spanish Adaptation and Validation in Chilean Students. *Universitas Psychologica* 13, 1111–1122. doi:10.11144/javeriana.upsy13-3.ecea

Maxwell, S., Reynolds, K. J., Lee, E., Subasic, E., and Bromhead, D. (2017). The Impact of School Climate and School Identification on Academic Achievement: Multilevel Modeling with Student and Teacher Data. *Front. Psychol.* 8, 2069. doi:10.3389/fpsyg.2017.02069

McDonald, R. P., and Ho, M. H. (2002). Principles and Practice in Reporting Structural Equation Analyses. *Psychol. Methods* 7, 64–82. doi:10.1037/1082-989x.7.1.64

McEvoy, A., and Welker, R. (2000). Antisocial Behavior, Academic Failure, and School Climate. *J. Emotional Behav. Disord.* 8, 130–140. doi:10.1177/106342660000800301

McGrath, K. V., Leighton, E. A., Ene, M., DiStefano, C., and Monrad, D. M. (2020). Using Integrative Data Analysis to Investigate School Climate across Multiple Informants. *Educ. Psychol. Meas.* 80, 617–637. doi:10.1177/0013164419885999

Mineduc (2011). *Política nacional de convivencia escolar*.

Mineduc (2015). *Política nacional de convivencia escolar 2015-2018*.

Mineduc (2019). *Política nacional de convivencia escolar. la convivencia la hacemos todos*.

Ministerio de Educación de Chile (2014). Otros indicadores de calidad educativa. Avilable at: https://bibliotecadigital.mineduc.cl/bitstream/handle/20.500.12365/10447/OIC.pdf?sequence=1&isAllowed=y. (accessedFebruary 24, 2021).

Muñoz, P. E., Casas, J. A., Del Rey, R., Ortega-Ruiz, R., Cerda, G., and Pérez, C. (2018). Validation and Cross-Cultural Robustness of the School-wide Climate Scale (Scs) across Spanish and Chilean Students. *Stud. Educ. Eval.* 56, 182–188. doi:10.1016/j.stueduc.2018.01.002

Muthén, L. K., and Muthén, B. O. (2007). *Mplus User's Guide*. 7th Edn. Los Angeles, CA: Muthén & Muthén, Vol. 7.

Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric Theory*. 3rd Edn. New. York: McGraw-Hill.

Olsson, U., Drasgow, F., and Dorans, N. J. (1982). The Polyserial Correlation Coefficient. *Psychometrika* 47, 337–347. doi:10.1007/BF02294164

Prati, G., Cicognani, E., and Albanesi, C. (2017). Psychometric Properties of a Multidimensional Scale of Sense of Community in the School. *Front. Psychol.* 8, 1466. doi:10.3389/fpsyg.2017.01466

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramelow, D., Currie, D., and Felder-Puig, R. (2015). The Assessment of School Climate. *J. Psychoeducational Assess.* 33, 731–743. doi:10.1177/0734282915584852

Revelle, W. (2020). *Psych: Procedures for Psychological, Psychometric, and Personality Rearch*. Evanston, Illinois: Northwestern University R package version 2.0.12.

Schermelleh-Engel, K., Moosbrugger, H., and Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-Of-Fit Measures. *Methods Psychol. Res. Online* 8, 23–74.

Steiger, J. H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behav. Res.* 25, 173–180. doi:10.1207/s15327906mbr2502_4

Steiger, J., and Lind, J. (1980). *Statistically Based Tests for the Number of Common Factors*. Annual meeting of the Psychometric Society, Iowa City, IA,

Tapia-Fonllem, C., Fraijo-Sing, B., Corral-Verdugo, V., Garza-Terán, G., and Moreno-Barahona, M. (2020). School Environments and Elementary School Children's Well-Being in Northwestern Mexico. *Front. Psychol.* 11, 510. doi:10.3389/fpsyg.2020.00510

Thapa, A., Cohen, J., Guffey, S., and Higgins-D'Alessandro, A. (2013). A Review of School Climate Research. *Rev. Educ. Res.* 83, 357–385. doi:10.3102/0034654313483907

Tubbs, J. E., and Garner, M. (2008). The Impact of School Climate on School Outcomes. *J. Coll. Teach. Learn.* 5, 17. doi:10.19030/tlc.v5i9.1230

Tucker, L. R., and Lewis, C. (1973). A Reliability Coefficient for Maximum Likelihood Factor Analysis. *Psychometrika* 38, 1–10. doi:10.1007/bf02291170

Unidad de CurrículumEvaluación - Ministerio de Educación y. (2013). Fundamentos otros indicadores de calidad educativa. Avilable at: https://s3.amazonaws.com/archivos.agenciaeducacion.cl/documentos-web/Ordenacion/Fundamentos_OIC.pdf. (accessed February 24, 2021).

Wang, M.-T., and Degol, J. L. (2016). School Climate: A Review of the Construct, Measurement, and Impact on Student Outcomes. *Educ. Psychol. Rev.* 28, 315–352. doi:10.1007/s10648-015-9319-1

Wang, M.-T., L. Degol, J., Amemiya, J., Parr, A., and Guo, J. (2020). Classroom Climate and Children's Academic and Psychological Wellbeing: A Systematic Review and Meta-Analysis. *Developmental Rev.* 57, 100912. doi:10.1016/j.dr.2020.100912

Woods, C. M., Cai, L., and Wang, M. (2013). The Langer-Improved Wald Test for DIF Testing with Multiple Groups. *Educ. Psychol. Meas.* 73, 532–547. doi:10.1177/0013164412464875

Wright, B., and Masters, G. (1982). *Rating Scale Analysis*. Chicago: MESA Press.

Xia, Y., and Yang, Y. (2018). Rmsea, Cfi, and Tli in Structural Equation Modeling with Ordered Categorical Data: The story They Tell Depends on the Estimation Methods. *Behav. Res.* 51, 409–428. doi:10.3758/s13428-018-1055-2

Yan, Z. (2020). Developing a Short Form of the Self-Assessment Practices Scale: Psychometric Evidence. *Front. Educ.* 4, 153. doi:10.3389/feduc.2019.00153

Yates, S. M. (2003). Gender Differences in Perceptions of School Climate: A Longtitudinal Study of School Reform. Paper presented at Educational Research, Risks & Dilemma.

Yates, S. M. (2001). Students' Achievement and Perceptions of School Climate during the Transition from Single Sex Education to Co-education. *Int. Education J.* 2, 317–328.

Zullig, K. J., Koopman, T. M., Patton, J. M., and Ubbes, V. A. (2010). School Climate: Historical Review, Instrument Development, and School Assessment. *J. psychoeducational Assess.* 28, 139–152. doi:10.1177/0734282909344205

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.