The Tripod School Climate Index: An Invariant Measure of School Safety and Relationships

Sarah Fierberg Phillips and Jacob F. S. Rowley

Recently revised standards for social work practice in schools encourage data-informed school climate interventions that implicitly require invariant measures of school climate. Invariant measures have the same meaning, scale, and origin across different groups of respondents. Although noninvariant measures bias statistical analyses and can lead users to erroneous conclusions, most school climate measures have not been tested for invariance. This study examines the invariance of the Tripod School Climate Index. Exploratory, confirmatory, and multiple-group confirmatory factor analyses were conducted on data collected from 66,531 students across 222 schools. Results indicate that the index is an excellent fit for the data and invariant by student grade level, demographic background, prior achievement, and dropout risk. Results imply that student responses can be validly aggregated to create school-level scores. The index will not bias studies of school climate interventions or bivariate analyses comparing perceptions of school climate across subgroups of students attending the same school. Given the centrality of school climate interventions to social work practice in schools and the consequences of noninvariance, the development of an index with these properties is an important contribution to the field.

KEY WORDS: *measurement invariance; school climate; school social work*

Recently revised standards for social work practice in schools encourage data-informed school climate interventions (National Association of Social Workers [NASW], 2012). Although not explicitly stated, these standards require invariant measures of school climate. Invariant measures assess school climate in the same way across different groups of students (Davidov, Meuleman, Cieciuch, Schmidt, & Billiet, 2014). They are critical because noninvariant measures bias statistical analyses and can lead researchers and practitioners to erroneous conclusions when used to assess school climate or evaluate climate-related interventions.

In light of the statistical analyses they are likely to perform, social work researchers and practitioners require measures with configural, metric, and scalar invariance. *Configural invariance* indicates that school climate means the same thing to different subgroups of students or schools (Davidov et al., 2014; Vandenberg & Lance, 2000). Researchers and practitioners cannot aggregate individual perceptions of school climate to the school level or compare perceptions across schools using a school climate index that lacks configural invariance.

Metric noninvariance biases regression coefficients and is an indication that a measure is scaled differently

across groups of respondents (Davidov et al., 2014; Steenkamp & Baumgartner, 1998). It can lead researchers to incorrectly estimate the effectiveness of a school climate intervention or the extent to which an intervention works better or worse for particular subgroups of students (Chen, 2008). A study examining the relationship between teacher effectiveness and student achievement illustrates the problems caused by metric noninvariance. In the study, no significant relationship between teacher effectiveness and student achievement was observed in preliminary analyses. However, the coefficient on teacher effectiveness tripled in size and became statistically significant when metric noninvariance was controlled (Kelcey, 2014).

Whereas metric noninvariance can compromise efforts to evaluate school climate interventions, *scalar noninvariance* can lead to incorrect conclusions about subgroup differences in perceptions of school climate. This is particularly relevant to school social workers because NASW's (2012) guiding principles of equal educational opportunity and tiered interventions require researchers and practitioners to carefully monitor the experience of at-risk students. An indication that the origin of a measure differs across groups, scalar noninvariance biases observed

construct scores (Davidov et al., 2014; Steenkamp & Baumgartner, 1998). A scalar noninvariant school climate index cannot be used to accurately compare perceptions of school climate across different subgroups of students. A study of the Chinese version of the Beck Depression Inventory II (Chinese Behavioral Science Corporation, 2000) illustrates this issue. Finding that scalar noninvariance accounts for significant gender differences in depression rates among Taiwanese adolescents (Wu & Huang, 2014), this study has challenged prior research suggesting that the prevalence of depression is higher among Taiwanese girls than boys.

Despite evidence that noninvariant measures compromise statistical analyses and can lead to erroneous conclusions, few existing measures of school climate have been tested for invariance. Of the 206 studies cited in the most recent review of the school climate literature (Thapa, Cohen, Guffey, & Higgins-D'Alessandro, 2013), just two examined the invariance of a school climate measure among different groups of students in the United States (Brand, Felner, Shim, Seitsinger, & Dumas, 2003; Schneider & Duran, 2010). Tests of measurement invariance have been documented for only four (Anderson-Butcher, Amorose, Iachini, & Ball, 2012; Furlong, O'Brennan, & You, 2011; Glaser, Horn, Arthur, Hawkins, & Catalano, 2005; Zullig, Collins, Ghani, Patton, & Ubbes, 2015) of the 14 student perception measures of school climate approved by the U.S. Department of Education's Office of Safe and Healthy Students (National Center on Safe Supportive Learning Environments, 2014). Although 62,300 studies were located in Google Scholar using the search terms "school climate" and "student*" in January 2015, just eight relevant studies not previously identified were located when key words related to measurement invariance were added to the search (Bandyopadhyay, Cornell, & Konold, 2009; Bear, Gaskins, Blank, & Chen, 2011; Bradshaw, Waasdorp, Debnam, & Johnson, 2014; Cornell, Klein, Konold, & Huang, 2012; Diemer et al., 2014; Karcher & Sass, 2010; Rebelez & Furlong, 2013; Wang & Dishion, 2012).

Existing invariance studies also suffer from critical analytic gaps. Of the 14 studies located for this review, fewer than 60% tested for scalar invariance (Bear et al., 2011; Bradshaw et al., 2014; Diemer et al., 2014; Furlong et al., 2011; Glaser et al., 2005; Karcher & Sass, 2010; Rebelez & Furlong, 2013; Wang & Dishion, 2012). Recall that scalar noninvariance biases observed construct scores. Because

extant invariance studies routinely omit tests of scalar invariance, researchers and practitioners using most measures of school climate cannot determine if mean differences across groups accurately reflect real differences in perceptions of school climate or simply measurement error.

At the same time, existing studies have not examined invariance across key groups of students. Studies have most commonly examined invariance across student sex or gender (Anderson-Butcher et al., 2012; Bandyopadhyay et al., 2009; Bear et al., 2011; Bradshaw et al., 2014; Brand et al., 2003; Glaser et al., 2005; Karcher & Sass, 2010; Rebelez & Furlong, 2013; Zullig et al., 2015), race or ethnicity (Bandyopadhyay et al., 2009; Bear et al., 2011; Bradshaw et al., 2014; Brand et al., 2003; Furlong et al., 2011; Glaser et al., 2005; Karcher & Sass, 2010; Schneider & Duran, 2010; Zullig et al., 2015), and grade level (Bear et al., 2011; Bradshaw et al., 2014; Brand et al., 2003; Glaser et al., 2005). Just two studies have examined invariance by socioeconomic status (SES) (Brand et al., 2003) or immigration generation (Diemer et al., 2014). School social workers' commitment to equal educational opportunity (NASW, 2012) requires scholars to pay more attention to invariance across commonly recognized dimensions of social inequality, like SES, that have been associated with disparities in educational outcomes. Practitioners also require school climate measures with the same meaning, scale, and origin among highly and weakly engaged students. They cannot effectively refer students who fail to respond to school climate interventions for interventions aimed at "improving early academic and social-emotional engagement" (NASW, 2012, p. 4) if their school climate measures fail to accurately capture the perceptions of disengaged students.

In sum, although measurement noninvariance threatens validity and can lead to erroneous conclusions when used in comparative research and effectiveness studies, few school climate measures have been tested for invariance. Only a handful of studies have examined scalar invariance, and invariance tests have not been conducted for most measures across key subgroups of students. The present study aims to address these gaps in the literature. It asks whether the Tripod School Climate Index has the same meaning, scale, and origin across student grade level, race or ethnicity, gender, SES, prior achievement, and dropout risk. Given the centrality of school climate interventions to social work practice in schools and the consequences of noninvariance, the

development of an invariant measure of school climate is an important contribution to the field.

METHOD

Analyses use de-identified secondary data collected from all students taking the long form of the Tripod secondary student survey during fall 2012 ($N_{\text{students}} = 27,420$; $N_{\text{schools}} = 101$) or spring 2013 ($N_{\text{students}} = 66,531$; $N_{\text{schools}} = 222$). Developed by Harvard University professor Ronald Ferguson, Tripod student surveys were first administered in 2001. In 2008, the Bill & Melinda Gates Foundation selected Tripod’s teaching effectiveness measure for its Measures of Effective Teaching project. Currently, millions of students participate in Tripod surveys each year as part of ongoing teacher evaluation and school improvement processes.

Table 1 provides demographic information for the spring 2013 sample, which is used for both confirmatory factor analysis (CFA) and tests of measurement invariance. Because large differences in group size can bias tests of measurement invariance (Brown, 2006; Chen, 2007), random equal-size subsamples

were selected for each grouping variable examined. Subsample sizes are also presented in Table 1.

Data Collection

Cambridge Education, Tripod’s implementation partner, collected the primary data used for these analyses in collaboration with associated school districts. Harvard University’s institutional review board approved all procedures and forms. Passive consent was secured from parents by mail. Surveys were administered online or on paper in English or Spanish. Students were instructed to be as honest as possible and told that they could opt out entirely or skip any questions they did not wish to answer.

Measures

School Climate. The proposed school climate index includes seven items representing two commonly accepted dimensions of school climate—relationships and safety. The two safety-related items are “This school feels like a safe place” and “At this school, I must be ready to fight to defend myself” (reverse coded). The five relationship-related items capture the quality of reciprocal relationships between students as well as between students and teachers. They are “Teachers in the hallway treat me with respect even if they don’t know me,” “I treat the adults at this school with respect, even if I don’t know them,” “The way adults treat me makes me angry” (reverse coded), “I would quiet down in the hallway if someone asked me to,” and “I trust other students at this school even if I don’t know them.” For each question, students were asked to select one of five Likert-type response options ranging from 1 = never to 5 = always.

Grouping Variables. Tests of measurement invariance were conducted across student grade level (sixth through eighth or ninth through 12th), race or ethnicity (black/African American, white, or Hispanic/Latino), SES (high, moderate, or low), gender (male or female), prior achievement (prior-term grade point average [GPA] in the A range, B range, or C to F range), and dropout risk (no risk or some to high risk). All grouping variables were student reported. Grade level, race or ethnicity, gender, and prior achievement were calculated directly from corresponding survey items. SES is the simple average of standardized items capturing parent education, family structure, family resources, and cultural capital (National Center for Education Statistics, 2012) restandardized around the student mean ($\alpha = .61$). Moderate SES was defined as within one standard

Table 1: Demographic Characteristics of the Spring 2013 Sample of the Tripod Secondary Survey V17		
Characteristic	Students (n)	Random Subsample Selected (n)
Grade		
6–8	26,956	26,000
9–12	32,614	26,000
Race/ethnicity		
Black/African American	10,065	10,000
White	20,974	10,000
Hispanic/Latino	10,142	10,000
Asian	2,008	0
Socioeconomic status		
High	10,813	10,000
Moderate	37,967	10,000
Low	10,775	10,000
Gender		
Male	27,909	27,000
Female	30,175	27,000
Prior-term grade point average		
A range	19,891	13,000
B range	24,411	13,000
C range or below	13,558	13,000
Dropout risk		
No risk	41,916	14,000
Some to high risk	14,764	14,000

deviation of the mean. Following the work of Balfanz and colleagues (Balfanz & Boccanfuso, 2007; Balfanz & Legters, 2001, 2004), dropout risk was constructed from three items tapping students' grades, behavior, and attendance. A student was considered to be at no risk of dropping out if he or she reported a prior-term GPA in the A or B range and responded "somewhat," "mostly untrue," or "totally untrue" to the statements "I don't really care whether I arrive on time to this class" and "My behavior is a problem for the teacher in this class."

Analyses

Exploratory Factor Analyses (EFAs) and CFAs. EFA was conducted using data collected in fall 2012. Although theory suggests a two-factor model of safety and relationships, this model cannot be identified with just two available safety-related items (Brown, 2006). Therefore, EFA tests the fit of a single-factor model. Modification indices were used to make theoretically sound improvements to the EFA model (Brown, 2006), and CFA tested the fit of the final model using spring 2013 data.

Given the ordinal nature of the data, parameter estimates were calculated using weighted least squares, and robust standard errors corrected for the clustering of students in schools. A nonsignificant chi-square value, root mean square error of approximation (RMSEA) at or below .06, standardized root mean square residual (SRMR) less than .08, and comparative fit index (CFI) close to or greater than .95 indicate excellent fit (Hu & Bentler, 1999), whereas a CFI at or above .90, an RMSEA at or below .08, and an SRMR at or below .10 indicate acceptable fit (Vandenberg & Lance, 2000). Although widely used, these thresholds were established using maximum likelihood estimation methods most appropriate for continuous data. Consequently, CFA results were reestimated and reevaluated using maximum likelihood estimation methods to ensure that method effects do not unduly influence conclusions about model fit.

Multiple Group CFAs. Spring 2013 data were used to examine the invariance of the CFA model following established multiple-group confirmatory factor analytic procedures (Brown, 2006). A model is considered to be configurally invariant if its fit statistics are within commonly accepted thresholds for all subgroups (van de Schoot, Lugtig, & Hox, 2012; Vandenberg & Lance, 2000). Once configural invariance was confirmed, metric and scalar invari-

ance were tested sequentially by constraining the factor loadings and item intercepts to be equal across groups. In each case, the constrained model was compared with the previous model in which the same parameters were allowed to vary freely across groups (van de Schoot et al., 2012; Vandenberg & Lance, 2000).

Although chi-square difference tests (Steiger, Shapiro, & Browne, 1985) have traditionally been used to determine whether the constrained model is significantly different than the unconstrained model, the sensitivity of the chi-square test statistic to sample size and nonnormality has led several scholars to recommend the use of alternate fit indices (Chen, 2007; Cheung & Rensvold, 2002; Meade, Johnson, & Braddy, 2008). Chen's (2007) recommendations were used in this study. When sample size was greater than 300 and equal across groups, a decrease in CFI greater than or equal to .01 supplemented by an increase in RMSEA greater than or equal to .015 or an increase in SRMR greater than or equal to .03 indicated metric noninvariance, whereas a decrease in CFI greater than or equal to .01 supplemented by an increase in RMSEA greater than or equal to .015 or an increase in SRMR greater than or equal to .01 indicated scalar noninvariance.

Analyses used maximum likelihood estimation methods with robust standard errors. Although this is not the ideal approach to analyzing ordinal data, Chen's (2007) thresholds have not been confirmed using weighted least squares. In addition, research has suggested that ordinal data with five or more response options can be analyzed accurately with maximum likelihood estimations methods and robust standard errors (Rhemtulla, Brosseau-Liard, & Savalei, 2012). Maximum likelihood estimation methods also do not lead to inflated Type I error rates when used to test for measurement invariance with ordinal data (Koh & Zumbo, 2008).

RESULTS

EFAs and CFAs

Analyses indicated that the initial EFA model was a poor fit for the data on all available indices: $\chi^2(14) = 3,596.90$, $p \leq .001$; RMSEA = .097, 90% confidence interval (CI) [.094, .099]; CFI = 0.830. Using modification indices, a revised model added three theoretically reasonable pathways correlating the residual variance of four items. This model was an excellent fit for the data on all indices except chi square, which is not surprising given the very large

sample size: $\chi^2(11) = 698.53, p \leq .001$; RMSEA = .048, 90% CI [.045, .051]; CFI = 0.945. Standardized factor loadings for all items were significant and above 0.30 for all but two items, “I trust other students at this school even if I don’t know them” and “This school feels like a safe place”: $\lambda = 0.177, SE = 0.018, p \leq .001$, and $\lambda = -0.244, SE = 0.016, p \leq .001$, respectively. However, these items were retained due to their theoretical importance and the lack of available substitutes in the data.

CFA results generally confirmed the acceptability of the final EFA model: $\chi^2(11) = 1,692.49, p \leq .001$; RMSEA = .048, 90% CI [.046, .050]; CFI = 0.932. Although the chi-square statistic was much larger in the spring 2013 data, this is not surprising given the corresponding increase in sample size. In contrast, the RMSEA and CFI were quite similar across the final EFA and initial CFA models. Although the CFI was somewhat lower in the initial CFA model and slightly below the recommended 0.95 cutoff,

it was well above 0.90. Furthermore, the same model reestimated using maximum-likelihood estimation methods was an excellent fit for the data on all indices except chi square: $\chi^2(11) = 2,007.843, p \leq .001$; RMSEA = .052, 90% CI [.050, .054]; CFI = 0.979; SRMR = 0.029. This model is presented in Figure 1.

Multiple Group CFAs

Table 2 presents the final CFA model fit statistics associated with tests of configural invariance. Whereas the chi-square test statistic was significant in all cases, alternate fit indices generally indicated acceptable to excellent model fit. The SRMR was well within recommended thresholds in all cases, and the CFI was above 0.95 for all but three subgroups (low SES, some to high dropout risk, and prior GPA in the C to F range). However, the CFI for these subgroups was closer to 0.95 than 0.90, indicating better than acceptable fit. Likewise, for all but four subgroups

Figure 1: Confirmatory Factor Analysis Results

Table 2: Fit Statistics between Groups for Final CFA Model

Group	<i>n</i>	χ^2	<i>df</i>	<i>p</i>	CFI	RMSEA	90% CI	SRMR
Full sample	66,531	2,007.84	11	≤ .001	0.979	.052	.050, .054	0.029
Grades 6–8	26,000	1,082.20	11	≤ .002	0.968	.061	.058, .064	0.027
Grades 9–12	26,000	1,434.24	11	≤ .003	0.966	.071	.067, .074	0.033
White	10,000	466.23	11	≤ .004	0.965	.064	.059, .069	0.028
Black/African American	10,000	363.47	11	≤ .005	0.953	.057	.052, .062	0.030
Hispanic/Latino	10,000	413.62	11	≤ .006	0.959	.060	.056, .066	0.031
High SES	10,000	473.00	11	≤ .007	0.983	.065	.060, .070	0.031
Moderate SES	10,000	405.94	11	≤ .008	0.954	.060	.055, .065	0.028
Low SES	10,000	403.19	11	≤ .009	0.948	.060	.055, .065	0.031
Male	27,000	883.28	11	≤ .010	0.961	.054	.051, .057	0.028
Female	27,000	1,510.56	11	≤ .011	0.975	.071	.068, .074	0.033
Prior GPA: A range	13,000	505.28	11	≤ .012	0.964	.059	.054, .063	0.029
Prior GPA: B range	13,000	518.33	11	≤ .013	0.953	.060	.055, .064	0.031
Prior GPA: C–F range	13,000	535.70	11	≤ .014	0.941	.061	.056, .065	0.031
No dropout risk	14,000	755.70	11	≤ .020	0.952	.070	.065, .074	0.035
Some to high dropout risk	14,000	637.03	11	≤ .021	0.931	.063	.059, .068	0.034

Notes: CFA = confirmatory factor analysis; χ^2 = chi-square statistic; *df* = degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation; CI = confidence interval; SRMR = standardized root mean square residual; SES = socioeconomic status; GPA = grade point average.

(high school, high SES, female, and no dropout risk), the RMSEA rounded to .06. In addition, for all groups, the upper limit of its 90% CI was never above .074. These results indicate that the proposed school climate measure is configurally invariant; school climate generally means the same thing across subgroups.

Tests of metric and scalar invariance are presented in Table 3. In all cases, Model 1 is the unconstrained, configural model presented in Figure 1. This model provides the same information as the tests of configural invariance described previously. Model 2 constrains the factor loadings to be equal across groups, and Model 3 constrains the factor loadings, item intercepts, and latent means to be equal across groups. Where applicable, Model 4 tests for partial scalar invariance and is the same as Model 3 except two to three item intercepts (depending on the group being examined) are allowed to vary across groups.

Confirming the findings reported in Table 2, the configural invariant models in Table 3 (Model 1) are an acceptable to excellent fit for the data from every group. The difference in CFI, RMSEA, and SRMR between the metric (Model 2) and configural invariant (Model 1) models are also within [Chen's \(2007\)](#) thresholds in all instances, meaning that the items used to measure school climate are assessed on the same scale across all subgroups. In contrast, tests of scalar invariance yield mixed results. The difference in CFI, RMSEA, and SRMR values for the scalar (Model 3) and metric invariant (Model 2) models

indicate full scalar invariance for grade level and GPA. This means that the origin of the proposed school climate measure is the same for students in different grades and with different prior-term GPAs. However, results for race or ethnicity, SES, gender, and dropout risk do not support full scalar invariance. For these groups, modification indices from the scalar invariant model (Model 3) suggest mean differences in the intercepts of two to three of the following items, depending on the groups being examined: “I trust other students at this school even if I don’t know them,” “At this school, I must be ready to fight to defend myself,” and “Teachers in the hallways treat me with respect, even if they don’t know me.” Freeing the appropriate intercepts in Model 4 leads to acceptable changes in model fit when compared with Model 2. In other words, partial scalar invariance is confirmed across race or ethnicity, SES, gender, and dropout risk.

Although complete scalar invariance is ideal, it is generally accepted that partial invariance is sufficient to produce unbiased mean comparisons ([Steenkamp & Baumgartner, 1998](#); [Vandenberg & Lance, 2000](#)). Consequently, the results of this study indicate that it is possible to use the Tripod School Climate Index to compare means and parameter estimates across all student subgroups examined in this study. However, it is useful to examine the practical significance of having two to three (out of seven) noninvariant item intercepts for some groups. To do so, Table 4 presents the observed school climate mean for each

Table 3: Fit Statistics for Multiple Group Confirmatory Factor Analysis by Grade Level, Race or Ethnicity, SES, Gender, GPA, and Dropout Risk

Group	Model	<i>n</i>	χ^2	<i>df</i>	RMSEA	90% CI	CFI	SRMR	Δ RMSEA	Δ CFI	Δ SRMR
Grade level	Model 1	26,000	2,520.86	22	.066	.064, .068	0.987	0.030			
	Model 2	26,000	1,695.38	28	.048	.046, .050	0.991	0.032	-.018	0.004	0.002
	Model 3	26,000	1,166.75	34	.036	.034, .038	0.994	0.039	-.012	0.003	0.007
Race/ ethnicity	Model 1	30,000	1,238.26	33	.060	.058, .063	0.960	0.030			
	Model 2	30,000	1,741.98	45	.061	.059, .064	0.943	0.051	.001	-0.017	0.021
	Model 3	30,000	3,473.05	57	.077	.075, .080	0.886	0.070	.016	-0.057	0.019
SES	Model 4	30,000	2,063.73	53	.058	.056, .060	0.943	0.054	-.003	0.000	0.003
	Model 1	30,000	1,286.71	33	.062	.059, .065	0.963	0.030			
	Model 2	30,000	1,922.44	45	.065	.062, .067	0.944	0.050	.003	-0.019	0.020
Gender	Model 3	30,000	5,543.88	57	.098	.096, .100	0.836	0.068	.033	-0.108	0.018
	Model 4	30,000	1,859.63	51	.060	.057, .062	0.946	0.051	-.005	0.002	0.001
	Model 1	27,000	2,326.35	22	.062	.060, .064	0.965	0.030			
GPA	Model 2	27,000	2,532.94	28	.058	.056, .059	0.962	0.036	-.004	-0.003	0.006
	Model 3	27,000	3,379.92	34	.060	.059, .062	0.950	0.048	.002	-0.012	0.012
	Model 4	27,000	2,723.61	32	.056	.054, .058	0.960	0.039	-.002	-0.002	0.003
Dropout risk	Model 1	13,000	1,558.28	33	.060	.057, .062	0.954	0.030			
	Model 2	13,000	1,718.84	45	.053	.051, .056	0.949	0.037	-.007	-0.005	0.007
	Model 3	13,000	2,471.84	57	.057	.055, .059	0.927	0.046	.004	-0.022	0.009
Dropout risk	Model 1	14,000	1,378.02	22	.066	.063, .069	0.942	0.034			
	Model 2	14,000	1,786.96	28	.067	.064, .070	0.925	0.051	.001	-0.017	0.017
	Model 3	14,000	2,399.02	34	.070	.068, .073	0.899	0.065	.003	-0.026	0.014
Dropout risk	Model 4	14,000	1,887.37	32	.064	.062, .067	0.921	0.056	-.003	-0.004	0.005

Notes: Model 1 = configural invariant; Model 2 = metric invariant; Model 3 = scalar invariant; Model 4 = partial scalar invariant; SES = socioeconomic status; GPA = grade point average; χ^2 = chi-square statistic; *df* = degrees of freedom; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; SRMR = standardized root mean square residual; Δ RMSEA = RMSEA constrained – RMSEA unconstrained; Δ CFI = CFI constrained – CFI unconstrained; Δ SRMR = SRMR constrained – SRMR unconstrained.

Table 4: The Impact of Differential Item Functioning on Observed School Climate Means

Group	Observed Mean
White students	3.25
Black/African American students	3.32
Hispanic/Latino students	3.29
High SES students	3.07
Moderate SES students	3.12
Low SES students	3.16
Female students	3.20
Male students	3.27
No dropout risk	3.03
Some to high dropout risk	3.11

Notes: Bolded group is the reference group. Unstandardized results presented. The observed school climate average is the simple average of the mean noninvariant and invariant observed item scores by group, constraining the latent mean of the school climate construct to be equal across groups. SES = socioeconomic status.

partially invariant subgroup. The observed school climate mean is calculated as the simple average of the mean observed score for all seven items, constraining the mean of the latent school climate construct to be equal across groups (Brown, 2006). Table 4 indicates that partial scalar invariance has a very small impact on observed mean differences in

subgroups' perceptions of school climate: $\bar{\chi} = 0.07$, $SD = 0.018$, range = 0.05–0.09. In other words, noninvariant item intercepts are responsible for approximately seven hundredths of a point difference (out of a possible five points) in mean school climate scores across racial or ethnic, gender, SES, or dropout risk subgroups. This finding offers additional evidence suggesting that the proposed school climate measure can be used to accurately compare perceptions of school climate across all subgroups of students examined in this study.

DISCUSSION

This study examined the factor structure and invariance of the Tripod School Climate Index. Findings indicate that the index is invariant or partially invariant by student grade level, race or ethnicity, gender, SES, prior achievement, and dropout risk. As one of the few school climate indices with documented evidence of invariance, these findings have important implications for social work researchers and practitioners working in schools.

Because tests of configural invariance indicate that all subgroups of students conceptualize school climate

similarly, findings imply that the responses of students in the same school can be averaged to create school-level scores. Whereas school climate scores are often calculated in this manner, tests of configural invariance are so rare that researchers and practitioners can rarely be certain that school-level scores aggregate the same latent construct across groups.

Findings related to metric invariance are significant because they imply that the Tripod index will produce unbiased regression coefficients when used to test the effectiveness of a school climate intervention or compare the effectiveness of an intervention across subgroups of students. This has particular relevance for social work researchers trying to develop evidence-based school climate interventions for use in diverse contexts. To endorse a particular intervention model, researchers must be confident that their regression coefficients are estimated without bias.

Finally, findings related to scalar invariance suggest that researchers and practitioners can use the Tripod index to accurately compare perceptions of school climate among various subgroups of students. The ability to identify group-level differences in perceptions of school climate and carefully monitor the experience of black and Latino students, students from low SES families, boys, and academically and social-emotionally disengaged students may help researchers and practitioners meet NASW (2012) standards related to equal opportunity and tiered interventions. The literature clearly documents a significant positive relationship between student perceptions of school climate and a host of prosocial outcomes (Thapa et al., 2013).

Despite these important contributions, this study is not without its limitations. The decision to address widely divergent subgroup sample sizes by selecting random equal-size samples probably led to more outliers being included in some subgroups than others. This could account for the partial scalar invariance found for race or ethnicity, SES, and dropout risk, with the initial sample size of some subgroups being quite small. A second limitation is the decision to use maximum likelihood estimation methods for invariance testing. Not the ideal approach for analyzing ordinal data, this decision may have resulted in slightly inflated model fit statistics. However, it likely influenced the fit statistics of subsequent models in the same way. Consequently, any bias probably had little effect on changes in the CFI, RMSEA, or SRMR, which were used to demonstrate invariance. Finally, even though this study examined a

broader array of subgroups than any prior study examining the invariance of a school climate index, its analysis is not exhaustive.

Notwithstanding these limitations, this study suggests a number of directions for future research. The bias caused by noninvariant measures could result in widespread adoption of ineffective school climate interventions or interventions targeting the wrong students. Consequently, more studies examining the invariance of existing school climate measures are needed. Studies including tests of scalar invariance are particularly necessary given the field's commitment to equal educational opportunities and tiered interventions, which require attention to particular subgroups of students. There also is a need for studies examining invariance across different types of schools and over time, and more studies focused on the subgroups of students most likely to require more intensive school social work interventions. Given the centrality of school climate interventions to social work practice in schools, researchers and practitioners need to know that their conclusions will not be affected by the measures they select to assess and monitor school climate. **SWR**

REFERENCES

- Anderson-Butcher, D., Amorose, A., Iachini, A., & Ball, A. (2012). The development of the Perceived School Experiences Scale. *Research on Social Work Practice, 22*(2), 186–194.
- Balfanz, R., & Boccanfuso, C. (2007, June). *Falling off the path to graduation: Middle grade indicators in Boston*. Baltimore: Center for Social Organization of Schools, Johns Hopkins University.
- Balfanz, R., & Legters, N. (2001, January). *How many central city high schools have a severe dropout problem, where are they located, and who attends them? Initial estimates using the Common Core of Data*. Paper presented at the meeting of the Civil Rights Project, Cambridge, MA.
- Balfanz, R., & Legters, N. (2004). *Locating the dropout crisis: Which high schools produce the nation's dropouts? Where are they located? Who attends them?* Baltimore: Johns Hopkins University.
- Bandyopadhyay, S., Cornell, D. G., & Konold, T. R. (2009). Validity of three school climate scales to assess bullying, aggressive attitudes, and help seeking. *School Psychology Review, 38*(3), 338–355.
- Bear, G. G., Gaskins, C., Blank, J., & Chen, F. F. (2011). Delaware School Climate Survey—Student: Its factor structure, concurrent validity, and reliability. *Journal of School Psychology, 49*, 157–174.
- Bradshaw, C. P., Waasdorp, T. E., Debnam, K. J., & Johnson, S. L. (2014). Measuring school climate in high schools: A focus on safety, engagement, and the environment. *Journal of School Health, 84*, 593–604.
- Brand, S., Felner, R., Shim, M., Seitsinger, A., & Dumas, T. (2003). Middle school improvement and reform: Development and validation of a school-level assessment of climate, cultural pluralism, and school safety. *Journal of Educational Psychology, 95*, 570–588.

- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464–504.
- Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, 95, 1005–1018.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9, 233–255.
- Chinese Behavioral Science Corporation. (2000). *Manual of Beck Depression Inventory—II*. Taipei, Taiwan: Author.
- Cornell, D., Klein, J., Konold, T., & Huang, F. (2012). Effects of validity screening items on adolescent survey data. *Psychological Assessment*, 24(1), 21–35.
- Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, 40, 55–75.
- Diemer, M. A., Li, C.-H., Gupta, T., Uygun, N., Sirin, S., & Rogers-Sirin, L. (2014). Pieces of the immigrant paradox puzzle: Measurement, level, and predictive differences in precursors to academic achievement. *Learning and Individual Differences*, 33, 47–54.
- Furlong, M. J., O'Brennan, L. M., & You, S. (2011). Psychometric properties of the Add Health School Connectedness Scale for 18 sociocultural groups. *Psychology in the Schools*, 48, 986–997.
- Glaser, R. R., Horn, M. L. V., Arthur, M. W., Hawkins, J. D., & Catalano, R. F. (2005). Measurement properties of the Communities That Care Youth Survey across demographic groups. *Journal of Quantitative Criminology*, 21(1), 73–102.
- Hu, L., & 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, 1–55.
- Karcher, M. J., & Sass, D. (2010). A multicultural assessment of adolescent connectedness: Testing measurement invariance across gender and ethnicity. *Journal of Counseling Psychology*, 57, 274–289.
- Kelcey, B. (2014, March 6). *Noninvariant measurement in rater-mediated assessments of teaching quality*. Paper presented at the Society for Research on Educational Effectiveness Spring 2014 Conference, Washington, DC.
- Koh, K. H., & Zumbo, B. D. (2008). Multi-group confirmatory factor analysis for testing measurement invariance in mixed item format data. *Journal of Modern Applied Statistical Methods*, 7, 471–477.
- 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, 568–592.
- National Association of Social Workers. (2012). *NASW standards for school social work services*. Washington, DC: Author.
- National Center for Education Statistics. (2012, November). *Improving the measurement of socioeconomic status for the National Assessment of Educational Progress: A theoretical foundation*. Washington, DC: Author.
- National Center on Safe Supportive Learning Environments. (2014, August). *Summary table of Office of Safe and Healthy Students approved school climate surveys*. Washington, DC: American Institutes for Research.
- Rebelez, J. L., & Furlong, M. J. (2013). Psychometric support for an abbreviated version of the California School Climate and Safety Survey. *International Journal of School & Educational Psychology*, 1(3), 154–165.
- Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17, 354–373.
- Schneider, S. H., & Duran, L. (2010). School climate in middle schools: A cultural perspective. *Journal of Research in Character Education*, 8(2), 25–37.
- Steenkamp, J.-B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1), 78–107.
- Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, 50(3), 253–263.
- Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research*, 83, 357–385.
- van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9, 486–492.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70.
- Wang, M. T., & Dishion, T. J. (2012). The trajectories of adolescents' perceptions of school climate, deviant peer affiliation, and behavioral problems during the middle school years. *Journal of Research on Adolescence*, 22, 40–53.
- Wu, P., & Huang, T. (2014). Gender-related invariance of the Beck Depression Inventory II for Taiwanese adolescent samples. *Assessment*, 21, 218–226.
- Zullig, K. J., Collins, R., Ghani, N., Patton, J. M., & Ubbes, V. A. (2015). Preliminary development of a revised version of the School Climate Measure. *Psychological Assessment*, 27, 1072–1081.

Sarah Fierberg Phillips, PhD, MSW, is vice president of research and **Jacob F. S. Rowley, MEd**, is research associate, Tripod Education Partners, Cambridge, MA. Address correspondence to Sarah Fierberg Phillips, Tripod Education Partners, 101 Main Street, 14th floor, Cambridge, MA 02142; e-mail: sarah@tripoded.com.

Original manuscript received March 4, 2015
 Final revision received June 4, 2015
 Accepted June 16, 2015
 Advance Access Publication December 27, 2015

MACRO PERSPECTIVES ON YOUTHS AGING OUT OF FOSTER CARE

MARY E. COLLINS

For the past 20 years, there has been increased attention to the circumstances under which young adults leave the foster care system when they achieve age-defined adulthood. In the United States, this is referred to as “aging out” of foster care. Typically, young people did not benefit from a reunification with their family or a permanent family situation with kin or through adoption.

Macro Perspectives on Youths Aging Out of Foster Care offers an extensive look at this problem through a macro orientation. Attempting to balance the primary focus on microsystems, and consistent with a social work perspective, this book aims to provide a greater emphasis on the larger macrosystems of society, policy, organization, and community. Successful or unsuccessful outcomes of the transition from foster care are heavily dependent on the processes and structures that make up the external environment. Youths exiting foster care may be especially influenced by the circumstances of the larger social context because they often lack the mediating advantages of a strong familial connection. After long stays in the foster care system, they may have limited support, skills, and resources required for a healthy, productive, independent adulthood.

This reorientation of focus to macrosystems affecting the individual transition experience informs questions such as these: What are the barriers to developing and implementing effective approaches? How can we bring more social attention to these youths? How might communities better support youths? To what extent should policy and program supports be designed specifically for this population, as opposed to a more expansive population of vulnerable youths (such as youths receiving child welfare services in their homes, or youths involved in more than one service system), or more general universal supports for all youths?

Macro Perspectives on Youths Aging Out of Foster Care draws on research, theory, and practice to address these issues. It is a useful resource for practitioners in child welfare and youth services, researchers, and policymakers.

ISBN: 978-0-87101-488-7. 2015. Item #4887. 224 pages. \$49.99

1-800-227-3590
www.naswpress.org

CODE:MPFC15