MALAYSIAN ONLINE JOURNAL OF EDUCATIONAL MANAGEMENT (MOJEM)

OCTOBER 2024, VOLUME 12, ISSUE 4, 24 - 41
E-ISSN NO: 2289 – 4489

LONGITUDINAL STUDY OF SCHOOL CLIMATE INSTRUMENT WITH SECONDARY SCHOOL STUDENTS: VALIDITY AND RELIABILITY ANALYSIS WITH THE RASCH MODEL

Siti Nadya Zynuddin[1*], Bambang Sumintono[2]

[1]
Faculty of Education,
Universiti Malaya,
50603 Kuala Lumpur

[2]
Faculty of Education,
Universitas Islam Internasional,
Depok,
Indonesia

Corresponding Author:
Faculty of Education,
Universiti Malaya,
50603 Kuala Lumpur
E-mail:
nadya.zynuddin@um.edu.my

ABSTRACT

The school climate plays a pivotal role in students' outcomes. Previous literature has highlighted several methodological approaches employed in the school climate domain, including longitudinal studies. However, little is known about the validity and reliability of school climate instruments for longitudinal studies using Rasch analysis. Rasch model is a powerful approach to validate assessment on both item and test levels. Rasch model is coined from the probability of each response and includes item difficulty parameters to characterize the measured items. Moreover, the score represents the item and the person involved with the assessment. Thus, the current study aimed to validate school climate instruments for longitudinal studies with a six-month gap within the context of secondary school students by utilising Rasch analysis. This study evaluated aspects of reliability and validity, such as unidimensionality, rating scale analysis, item fit statistics, item targeting, and differential item functioning. A total of 1,495 secondary school students from public schools in Selangor, Malaysia, completed a 28-item Malay version of the school climate survey at Time-1 and Time-2, with a six-month gap. The results of the Rasch analysis indicated that the instrument had excellent reliability and separation indices, excellent unidimensionality and construct validity, a functional rating scale, good item-person targeting, and good item fit statistics. The current findings provided valid and reliable insights pertinent for policymakers to strategise interventions and initiatives to enhance the quality of school climate and overall education, particularly in the Asian context.

Keywords: **School Climate, Longitudinal, Rasch Analysis, Validity and Reliability**

INTRODUCTION

Globally, the ultimate goal of education is to mould students into holistic individuals. With students spending most of their time in school, the school climate contributes significantly towards students' development. Education stakeholders increasingly recognise school climate as the catalyst for school improvement (Graham, 2022; Hamlin, 2020). For example, in the United States, policies focus on school climate under the United States Education Act and the Every Student Succeeds Act (ESSA) (Darling-Hammond et al., 2016). School climates must be conducive and promising to benefit students, teachers and communities.

The research on school climate has caught scholars' attention for almost half a century (Baumrind, 1966; Field et al., 1977). Generally, seminal scholars define school climate as comprising the varying interior features and the quality of the school in nurturing students' potential (Cohen et al., 2009; Gonder, 1994; Hoy & Miskel, 2013). Past literature has highlighted several domains related to school climate, including students' academic performance (Forsberg et al., 2021; Mateos et al., 2021; Monsillion et al., 2023), well-being (Mateos et al., 2021; Monsillion et al., 2023), engagement and attendance (Molinari & Grazia, 2023; Ryberg et al., 2020), as well as the safety (Lenz et al., 2021; Monsillion et al., 2023; Yang et al., 2020), and incidences of bullying (Nishimura et al., 2020; Rohatgi & Scherer, 2020) and delinquency (Rohatgi & Scherer, 2020; Xu et al., 2023) in schools.

In addition, Trends in International Mathematics and Science Study (TIMSS) 2019 findings reported that a school's sense of belonging positively influences student achievements in science and mathematics (Mullis et al., 2020). One of the critical factors that may influence the result is the school climate. A nation with a centralized educational system, Malaysia continuously enhances its curriculum while embracing change and the dynamics of a volatile, unpredictable, complex, and ambiguous (VUCA) environment. Malaysia's educational changes are consistent with global initiatives to improve school climate. As such, education stakeholders are significant in the local and global contexts of this study.

Methodologically, the school climate domain has been investigated through several approaches, including experiments, objective measurements, multisource, multimethod, mixed-method and qualitative studies, and scale development (Zynuddin et al., 2023). In the same vein, methodologically, numerous school climate studies reported on the validity and reliability of cross-sectional approaches (Aldridge & Blackstock, 2024; Graham, 2022; Mateos et al., 2021). However, little is known about the validity and reliability of school climate instruments over time using a longitudinal approach. While numerous studies on school climate employed the classical test theory for instrument measurement, research on its validity and reliability, specifically using Rasch analysis, remains scarce (Marraccini et al., 2020).

Using the Item Response Theory (IRT) model, the Rasch model is a powerful approach to validate assessment on both item and test levels. Rasch model is coined from the probability of each response and includes item difficulty parameters to characterize the measured items (Embretson et al., 2000). The current study employed the Rasch model to provide a valid and reliable evidence-based assessment tool. The Rasch model suggests that the measure will have adequate construct validity when it appropriately fits the estimates to the model and the item parameters are reasonably acceptable (Hinkin et al., 1997). Thus, contributing to the knowledge of school climate instrumentation, the current study measured a school climate instrument with Rasch analysis using a longitudinal approach.

LITERATURE REVIEW

School Climate in Education

The school climate concept, as articulated by Freiberg (1999), plays a crucial role in human development. A nurturing and supportive environment facilitates learning by fostering positive relationships among students and between students and teachers. School climate encompasses the values, norms, and beliefs contributing to students' sense of physical, emotional, and social security (Cohen et al., 2009). Theoretical frameworks have elucidated various aspects of school climate, which are distinct but overlap. Authoritative discipline theory, for instance, delineates two fundamental dimensions within the structure of school climate: responsiveness and demandingness (Baumrind, 1966; Gregory & Cornell, 2009). Responsiveness pertains to the social and emotional support provided by adults and peers to meet children's needs, distinguishing it from demandingness, while the latter involves the establishment of clear behavioural expectations, rules, fair enforcement, and supportive guidance from adults.

Stockard and Mayberry's theoretical framework of school climate highlighted similar domains of responsiveness and demandingness. However, differing from the authoritative discipline theory, Stockard and Mayberry described the concept of school climate as comprising social action and order (Stockard & Mayberry, 1992). Mainly, social action is similar to responsiveness, which is garnered from the day-to-day relationships among teachers, staff, and students. On the other hand, social order has similarities with demandingness (or structure), attenuating behavioural issues, and corroborating safety.

These theories align with Bronfenbrenner's (1977) bioecological theory of human development, which postulated associations between the social environment and human development. For example, Rudasill et al.'s (2018) study on System Views of School Climate (SVSC) was based on the traditional Bronfenbrenner's five-level ecological model. SVSC highlighted the term nanosystem, explaining that school climate comprises three domains: (1) shared beliefs and values, (2) relationships and social interactions and (3) safety. More recently, Marciniak et al. (2020) further expanded SVSC with the idea of outside settings, which addressed the linkages between school climate and the home, community, and other outside-school settings.

Conversely, through an economic lens, Human Capital Theory posits that accumulating skills and knowledge influences an individual's productivity. In this theory, Schultz (1961) argued that human capital is not financial but comprises knowledge, skill, creativity, and health. Combined, it is worth noting that, guided by theoretical arguments as explained above, there are several subdomains for responsiveness or social support, namely Teacher-Student Relations and Student-Student Relations. The other subdomains that entail demandingness or structure are Fairness of Rules, School Safety and Students' Conduct Problems, and Liking of School.

Empirically, recent findings found that the school climate has linkages with socio-emotional and behavioural outcomes as the malleability of these traits is related to academic performance (Forsberg et al., 2021; Larson et al., 2020; Mateos et al., 2021). These findings were echoed by Heckman et al. (2013), who revealed that the parents, school, and environment shape students as an investment in their later years. Similarly, Field et al. (1977) argued that educational institutions such as schools are where students socialise and build character instead of only engaging cognitively. A plethora of studies in this area involved students' academic outcomes (Marraccini et al., 2020a; Mateos et al., 2021), well-being (Monsillion et al., 2023), and engagement and attendance (Hamlin, 2020; Rohatgi & Scherer, 2020; Xu et al., 2023), as well as the schools' leadership (Larson et al., 2020; Pogodzinski et al., 2022), safety (Lenz et al., 2021; Monsillion et al., 2023; Yang et al., 2020), and incidences of bullying (Nishimura et al., 2020; Rohatgi & Scherer, 2020; Yang et al., 2020) and delinquency (Forsberg et al., 2021; Grazia & Molinari, 2020; Rohatgi & Scherer, 2020; Xu et al., 2023).

Recent reports from an international body have underscored certain countries, such as Australia, Japan, Korea, Singapore, and Switzerland, as having maintained students' academic performance in Mathematics amidst the

COVID-19 pandemic (OECD, 2024). These educational systems share common characteristics, notably ongoing support from teachers and parents, which are recognised as pivotal factors in school climate.

Thus, a supportive school climate potentially mitigates academic challenges, including narrowing the achievement gap due to socioeconomic status (SES) (Graham, 2022; Mateos et al., 2021). A positive school climate will optimise student development despite students' backgrounds, which has been accrued in most educational policies. Interestingly, students' perceptions of school climate within the same context and region vary based on their background. On the one hand, it was found that Asian students perceived a more positive school climate, while in contrast, African American and multi-racial students had more negative perceptions (Bear et al., 2017). Similarly, Graham (2022) also highlighted the contradicting perceptions of the school climate of Black and "Other" students compared to White students who had better perceptions of the school climate.

According to Creswell et al. (2018), it is pertinent to answer theory-guided research questions and hypotheses. Thus, the Rasch model was used in this study as an evidence-based, valid, and reliable assessment tool. Following the Rasch model, a measure will have good construct validity when the estimates are correctly fitted to the model and the item parameters are reasonably acceptable (Hinkin et al., 1997). As the study employed a quantitative approach that emphasized objective testing based on numerical data collection through questionnaire instruments, it is critical to measure the studied variables using statistical descriptions to perform descriptive, inferential, predictive, and statistical testing.

School Climate Instruments

Since the inception of the school climate concept, numerous instruments have been developed to assess it. These instruments have been employed across Western and Eastern contexts, yielding extensive empirical research. Notably, several instruments are publicly available for use: The Authoritative School Climate Survey (Cornell, 2014), The Consortium on Chicago School Research Survey of Chicago Public Schools (Consortium on Chicago School Research, 2007), Delaware Bullying Victimisation Student Scale (Bear et al., 2019), the Delaware School Climate Survey (Bear et al., 2011), Flourishing Children Survey Social Competence Adolescent Scale (Lippman et al., 2012), the Education Department School Climate Survey (EDSCLS) (National Centre on Safe Supportive Learning Environments, 2016), and the MDS3 School Climate Student Survey (Bradshaw et al., 2014) that was developed by the John Hopkins Centre for Youth Violence Prevention, focusing on three domains of school climate, namely safety, engagement and environment.

Recently, several measurement studies across the region have been conducted to understand the customised concept across different regions and contexts. For example, the Socio-Educational Environment Questionnaire developed and adapted in Canada, assessing several dimensions of school climate, was deemed a stable instrument using a second-order factor model (Grazia & Molinari, 2021). Besides that, Eastern scholars Nishimura et al. (2020) developed the Japan School Climate Inventory (JaSC) to measure school climate across subgroups of varying grades and genders among Japanese elementary and junior high school students. The findings indicate a satisfactory validity in measuring school climate consistently across diverse groups of Japanese students.

In the United States, the Education Department School Climate Survey (EDSCLS) is a free-access tool for local and state education agencies to assess three key domains of school climate: engagement, safety, and environment (Ryberg et al., 2020). This study utilising multilevel analysis revealed that, at the school level, the engagement and environment domains exhibit a simpler factor structure, whereas no adequately fitting model could be identified for the safety domain. These results underscore the complexity of achieving valid measurements solely through individual-level structures.

Prior research has predominantly centred on evaluating school climates within secondary education settings. However, Aldridge and Blackstock (2024) conducted pioneering work by constructing and validating a school climate survey tailored to primary school students. The results underscored the survey's importance and demonstrated its robust psychometric properties, offering a valuable tool for researchers, educational institutions, and policymakers interested in gauging primary school students' perceptions of school climate.

Moreover, studies by Pogodzinski et al. (2022) discussed The 5Essentials surveys represent a suite of survey tools crafted by the University of Chicago Consortium on School Research (UChicago Consortium) designed to assess school climate across five core domains: Effective Leadership, Supportive Environment, Collaborative Teachers, Ambitious Instruction, and Involved Families (Bryk, 2010). The results indicate that these school climate indicators did not significantly influence parents' decisions when selecting schools. Instead, parents focused more intensely on individualised factors of the student and their family circumstances. In conclusion, several previous research on school climate instruments (SCIs) have been conducted and developed through a Western lens and within the context of developing countries.

In the same vein, previous studies have also highlighted that research linking school climate with Rasch analysis, to the best of the researcher's knowledge, remains limited, particularly when using Rasch analysis. For example, Johnson et al. (1995) emphasized that school climate instruments should separately measure the affective and cognitive domains. Furthermore, their findings indicated that school climate instruments should differ for primary, junior, and senior high school students. A recent study by Pogodzinski et al. (2022) employed Rasch modelling to analyse item-level responses for the 5Essentials surveys, which were developed by the University of Chicago Consortium on School Research (UChicago Consortium) to measure school climate. Assessments can be successfully validated at the item and test levels using the Rasch model. The Rasch model is constructed from the likelihood of each response and includes item difficulty parameters to characterize the assessed items (Embretson et al., 2000).

Furthermore, the score represents the item and the person involved in the evaluation (Lord, 1980). Unlike classical test theory, the Rasch model makes at least a few fair and strong assumptions. Firstly, the assessment's construct is unidimensional. The second is local independence, which presupposes that each item's reaction is independent of the others. The item is not sample-dependent and the latent trait estimates are not test-dependent when both of the Rasch model's assumptions are satisfied.

METHODS

The current study employed a longitudinal quantitative method. A longitudinal methodology was applied to increase the strength of the empirical findings. The longitudinal research improved causes and consequences and reduced bias amongst respondents (Maxwell & Cole, 2007; Zapf et al., 1996).

Research Design
This study employed a longitudinal quantitative approach. The data collection was conducted at two different times, six months apart, with Time-1 in August 2021 and Time-2 in February 2022. The quantitative method used was required to validate the scale. The study used a School Climate Instrument (SCI) among regular Malaysian national secondary school students. The response options for the statements about the school climate were self-reported on a Likert rating scale. The data were collected in Malaysia using an online platform, namely Google Form.

Participants
All responses from 1,714 respondents for Time-1 and Time-2 were analysed using WINSTEPS version 4.8 for data cleaning (Widhiarso & Sumintono, 2016). In the analysis, the term "person misfit" referred to respondents who provided inconsistent responses, which would influence the overall analysis of the findings. If the respondent was a misfit for both waves of the data collection, the respondent was excluded. Consequently, several respondents were excluded, resulting in 1,495 respondents proceeding to the analysis of findings.

The instrument also obtained information on the respondents' background, such as gender, religion, ethnicity, and whether or not they live with both parents, one of them, or a guardian, as these are all factors considered for the respondent profile. The demographics of the 1,495 respondents are presented in Table 1.

Based on the statistics, female respondents outnumber the males. There were 1,036 female students among the 1,495, accounting for 69.3% of the total, with the 459 male students accounting for the remaining 30.7%. Furthermore, in terms of religion, more than three-quarters of the respondents were Muslims (75.6%), followed by Buddhists (11.4%), Hindus (9.1%), Christians (3.4%), and others (0.5%). For ethnicity, most of the respondents were Malay (74.1%), followed by Chinese (13.1%), Indian (10.7%), and others (2.2%). Most respondents lived at home with their mother and father (88%). A handful of them lived with a single caretaker, particularly 9.0% who lived with only their mother, 1.4% who lived with only their father, and 1.6% who lived with a guardian.

Table 1: *Demographic Data of Respondents (N=1,495)*

Demography	Variables	Frequency	Percentage (%)
Gender	Male	459	30.7
	Female	1,036	69.3
Religion	Muslim	1,130	75.6
	Buddhist	171	11.4
	Hindu	136	9.1
	Christian	51	3.4
	Others	7	0.5
Ethnicity	Malay	1,107	74.1
	Chinese	195	13.0
	Indian	160	10.7
	Others	33	2.2
Living with	Both mother and father	1,315	88.0
	Mother	135	9.0
	Father	21	1.4
	Guardian	24	1.6

Instrument

The School Climate Instrument (SCI) used in this study was based on the Delaware School Climate Survey-Student (DSCS-S), comprising five domains: teacher-student relations, student-student relations, fairness of rules, liking of school, and school safety (Bear et al., 2019). The DSCS-S consisted of 28 items measured using a four-point Likert rating scale (1 = 'strongly disagree' to 4 = 'strongly agree'). This measurement was reported stable across grades (elementary, middle, and high schools), racial-ethnic groups, and gender (Bear et al., 2011). It was also reported to have a good internal reliability coefficient, ranging from 0.70 to 0.88 across the five subscales and 0.94 for the total scale. The survey was translated from English to Malay using the back-translation method (Brislin, 1970).

Procedure

Before conducting the study, ethical approval was obtained to conduct this research among the school students. Furthermore, additional permission was obtained from the Educational Planning and Research Division (EPRD), Ministry of Education, and State Educational Office (JPN) to carry out research involving school students. The researcher also obtained an extension from the said bodies since the current study had to collect data several times over a longer time frame.

The participants' identities were kept highly confidential. On the cover page of the questionnaire, it was explicitly noted that the survey was strictly anonymous and voluntary, thereby ensuring the confidentiality and privacy of the respondents. By participating in the survey, the respondents agreed to the terms and conditions outlined in the questionnaire.

Measurement Model

The data collected in this study was ordinal and intended for counting (frequency of responses) rather than measuring latent traits, which, in this case, was school climate. The appropriate analysis for this data is through the Rasch Rating Scale Model (RSM). The procedure began by counting the raw data and calculating the odds and probability for each participant and item. These probabilities were then transformed into equal-interval data through a nonlinear function called logarithms (Boone & Staver, 2020; Sumintono & Widhiarso, 2014). The result of this transformation was measurement data on a consistent equal-interval scale, known as logits (an abbreviation for logarithm odds units). Following this, a conjoint measurement process calibrated these logits to ascertain the relationship between the difficulty level of the items and the participant's abilities. The process was done through WINSTEPS version 4.8, a software specifically designed for Rasch analysis.

The RSM approach is particularly suitable for measuring latent or hidden traits in assessing human opinions, perceptions, and attitudes (Bond & Fox, 2015; Engelhard, 2013). This type of analysis results can explain item difficulty levels with accurate and precise measurements (item calibration), detect item fit as compared to an ideal model (the Rasch model), as well as measure the respondents' perceptions (Lee et al., 2021; Rusland et al., 2020; Ratnaningsih et al., 2024).

Numerous results indicating instrument quality can be presented based on the RSM analysis approach. This includes reliability and separation, both for person and item, instrument unidimensionality and its construct validity, the functionality of the rating scale, item-person targeting, and item fit statistics analysis (whether all scale items have a good fit) (Akhtar & Sumintono, 2023; Ratnaningsih et al., 2024). For instance, to investigate the quality of the items, according to Boone et al. (2014), item fit is measured on three indices, which are *Outfit Mean Square* (y): $0.5 < y < 1.5$, *Outfit Z standard* (z): $-2.0 < z < +2.0$, and *Point Measure Correlation* (x) : $0.4 < x < 0.85$.

RESULTS

Reliability and Dimensionality

The WINSTEPS software, which applied the rating scale model approach, was utilised to analyse the two data sets. Table 2 presents the instrument's psychometric attributes, reliability, and dimensionality. As shown in the table, the mean measure (logit) was 0.00 logits with a minuscule standard error, showing no item outliers in the SCI, and there was good item precision in both data sets. Both standard deviation indices were above 0.50, indicating a good spread of item difficulty for Time-1 and Time-2 (0.87 and 0.79). This was also supported by both raw variance indices being above the minimum threshold of 20%, and in both data sets, the Eigenvalues were less than 3.00, indicating the instrument's good dimensionality.

Table 2. *Summary of Reliability and Dimensionality of the SCI*

	Time-1 Aug 2021	Time-2 Feb 2022
Item logits		
Mean	0.00	0.00
Standard Deviation, SD	0.87	0.79
Standard Error, SE	0.17	0.15
Outfit Mean Square		
Mean	1.00	0.97
SE	0.08	0.11
Separation	15.21	12.33
Reliability	1.00	0.99
Raw variance	39.2%	39.1%
Eigen value first contrast	2.4	2.3

The average outfit mean-square was near the expected value of 1, showing a uniform fit to the model with a minuscule standard error (Tucker-Drob et al., 2016). The separation index (more than 3.00), and reliability (more than 0.90) of the item and person statistics suggest excellent reliability (Fisher, 2007).

Rating Scale Analysis

Table 3 presents the distribution and performance of each response category (scores 1 to 4) of the rating scale for the two respective data set samples. The average measure increased monotonically for both samples, from -0.78 logits in category '1' to +3.47 logits in category '4' in the Time-1 data, whereas from -0.25 logits to +4.10 logits in Time-2 data. Both showed higher category ratings corresponding with higher measures of school climate perception and fulfilling monotonic assumptions. This indicated properly functioning rating categories in the SCI, with each successive rating scale category representing a higher latent trait level, as expected in a well-structured rating scale (Liu & Lim, 2020; Ratnaningsih et al., 2024).

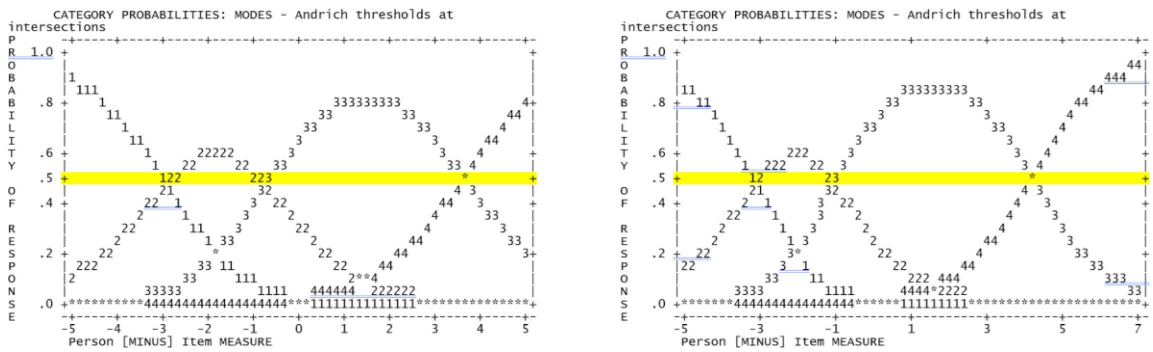
Table 3. *Rating Scale Category Statistics of the SCI*

Category	**Time-1 Aug 2021**			**Time-2 Feb 2022**		
	Average Measure	Outfit MNSQ	Step	Average Measure	Outfit MNSQ	Step
1	-0.78	1.89	NONE	-0.25	2.60	NONE
2	0.06	0.94	-2.90	-0.01	0.90	-3.12
3	+1.66	0.93	-0.73	+1.73	0.89	-1.04
4	+3.47	0.94	+3.63	+4.10	0.87	+4.16

Outfit Mean Square (MnSq) values of the two data sets were close to the ideal value of 1.00 across all categories, except for rating 1, which was above 1.50. This suggests that all the responses in the category were predictable except for rating 1. However, this is still considered an acceptable result for the beginning of a scale. These MnSq values support the notion that the rating scale categories operated effectively and that the respondents' selections were coherent with the Rasch model's expectations.

The threshold (step) values, representing the transition points between adjacent categories, did not indicate any disordered thresholds for both data sets (Time-1 and Time-2). Figure 1 clearly shows that, in both samples, each category had its peak, with all of them even above 50%, which means that the scale can differentiate between varying levels of school climate perception. These findings align with Fisher's (2007) recommendations for an optimally functioning rating scale. The consistent increase in average measures across the categories further supports the scale's capacity to capture the burnout trait's gradations effectively.

Figure 1: *Category Probability Curves of the SCI from the Two Data Sets*



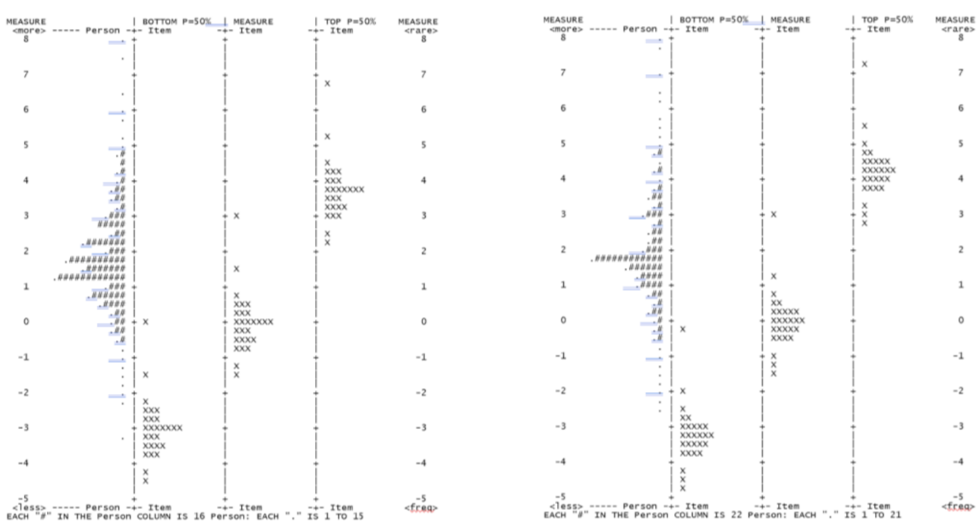
a. Time-1 *b. Time-2*

Item Targeting

One of the advantages of using the Rasch measurement model is its ability to detect outlier and misfit answers from respondents. As mentioned previously, there were no item outliers in the data. However, there were person outliers, two in Time-1, and four in Time-2, who provided maximum scores for all items. The rest of the personal data from both sets provided acceptable self-reporting responses of their perceptions of school climate.

In addition, item-person targeting is important to know the quality of the scale better. Respondents' answers on the scale were analysed to see how well the ranges of personal abilities matched the item difficulty levels. This is usually called the floor and ceiling effect, where the respondents' answers are consistently skewed towards lower or higher scores, indicating potential issues in targeting.

Figure 2: *Wright Map, or Item-Person Map of Both SCI Samples*



a. Time-1 Aug 2021 b. Time-2 Feb 2022

Evidence of targeting is shown in Figure 2, which is the Wright map or item-person map from both data sets. The left column of each Wright map shows the range of school climate scores of the respondents, from lower scores at the bottom to higher scores at the top; whereas item difficulties are in the three columns to the right (represented by XX). The SCI demonstrated a low to moderate item difficulty for both samples. Items located to the left of the centre (bottom p=50%), representing lower difficulty levels, had a probability exceeding 50% in the bottom category of the scale. This is while the one column of items in the far right (top p=50%), representing higher difficulty levels, had a probability below 50%, at the top category of the scale.

In Figure 2, nearly the entire range of person columns had low scores, meaning there was no floor effect. On the other hand, for the higher scores, only 2% and 3% of the respondents (30 and 42 students in Time-1 and Time-2 data respectively) could not be measured, which was considered acceptable (Liu & Lim, 2020).

Item Fit Statistics Indices

Besides providing psychometric attribute analysis at the instrument level, the unique aspect of the Rasch measurement model is its ability to analyse at the item level. This capacity is usually called individual-centred statistics (Engelhard et al., 2018), which provides information about item quality. Table 4 shows the result of the item analysis. The logit values, indicative of item difficulty or item location, spanned from -1.61 to +3.05 logits (SD = 0.87) for Time-1 data and -1.50 to +2.88 logits (SD = 0.79) for Time-2 data, signifying that all items in the two data sets fell within an acceptable range without any outliers, thus ensuring a cohesive scale. The Time-2 data set had a smaller range than Time-1, but the difference between them was less than 0.20 logits, showing that the SCI is a stable instrument. Furthermore, standard error measurements across items were notably minuscule at 0.15 to 0.17 logits, as shown in Table 4, reflecting high measurement precision.

Table 4: *Item Fit Statistics Indices of the SCI From Two Data Sets*

Item	Time-1 Aug 2021				Time-2 Feb 2022			
	Logit	Infit MNSQ	Outfit MNSQ	PT.Meas.Corr	Logit	Infit MNSQ	Outfit MNSQ	PT.Meas.Corr
A1	1.59	1.12	1.22	0.44	1.29	1.15	1.29	0.43
A2	-1.61	0.95	0.95	0.55	-1.50	0.98	0.98	0.59
A3	-0.69	1.11	1.09	0.56	-0.57	1.19	1.15	0.59
A4	-0.08	0.85	0.86	0.54	0.04	0.91	0.85	0.62
A5	0.00	0.87	0.85	0.56	0.15	0.97	0.94	0.58
A6	-0.79	0.94	0.92	0.62	-0.47	1.00	0.92	0.64
A7	-0.54	1.06	1.04	0.49	-0.31	1.11	1.09	0.54
AR8	0.31	***2.09***	***2.11***	0.43	0.24	***2.04***	***2.09***	0.44
A9	0.54	0.83	0.82	0.59	0.41	0.89	0.84	0.58
A10	0.11	0.86	0.82	0.65	-0.17	0.80	0.73	0.65
A11	0.62	1.09	1.08	0.62	0.70	0.96	0.92	0.65

A12	0.08	0.85	0.82	0.61	0.12	0.71	0.64	0.68
AR13	3.05	***1.92***	***2.27***	0.21	2.88	***1.98***	***2.68***	0.22
A14	-0.20	0.85	0.82	0.59	-0.12	0.81	0.72	0.61
A15	0.07	0.91	0.88	0.63	-0.12	0.78	0.69	0.65
A16	-0.35	0.96	0.91	0.62	0.07	0.91	0.82	0.66
A17	-0.43	0.72	0.69	0.60	-0.19	0.65	0.58	0.67
A18	0.26	0.80	0.78	0.61	0.24	0.74	0.66	0.66
A19	0.68	0.66	0.64	0.63	0.50	0.57	0.50	0.69
A20	0.01	0.78	0.73	0.65	-0.33	0.70	0.61	0.66
A21	0.55	1.04	1.03	0.53	0.26	0.86	0.82	0.60
A22	0.34	0.81	0.80	0.62	0.36	0.73	0.67	0.65
A23	-0.05	0.74	0.70	0.63	0.11	0.68	0.59	0.66
AR24	-0.86	***1.51***	***1.70***	0.48	-1.04	***2.15***	***2.48***	0.40
A25	-0.45	0.64	0.60	0.59	-0.28	0.56	0.47	0.63
A26	-1.36	1.30	1.30	0.47	-1.25	1.24	1.27	0.49
A27	-0.50	0.79	0.75	0.64	-0.61	0.70	0.61	0.67
A28	-0.31	0.74	0.71	0.65	-0.41	0.71	0.63	0.69

Item with code R (such as AR8) refers to a reversed score or unfavourable statement; numbers in italics and bold are misfit indices.

Focusing on the item fit statistics, which typically encompass mean squares (MnSqs), t-statistics (ZStds), and point measure correlations, since both data sets had more than 400 respondents, ZStds indices were not used because they were not sensitive to bigger samples (Boone et al., 2014). For the MnSq values, both infit and outfit values should lie between 0.50 and 1.50 to be acceptable (Bond & Fox, 2015). Table 5 shows that several items were close to the ideal value of 1.00, except for three unfavorable ones (AR8, AR13, and AR24). This indicates that most of the items in both samples from the SCI fit well with the model. The point measure correlations further validated the item quality. All items in both samples showed positive correlations (larger than 0.20). Such positive values corroborated the alignment of each item with the underlying latent variable, school climate. A consistent point measure correlations suggested that the items measured the intended construct effectively and were congruent with the overall direction of the scale.

From the analysis of the two-waved data samples, it can be extrapolated that most of the SCI items, except for the three unfavourable items, were not only precise but also had good internal consistency, as indicated by their fit statistics and point measure correlations. The items' ability to span a range of difficulties is crucial for an instrument intended to measure a construct as complex and nuanced as school climate. Each item has been shown to contribute meaningfully to the measurement of the variable, with a high degree of precision and alignment with the latent trait.

Differential Item Functioning (DIF)

As one of the beneficial aspects of applying individual-centred statistics, the Rasch model can be analysed for Differential Item Functioning (DIF). This analysis is to understand the different responses to each item based on each population subgroup and if there was any bias, for instance. Since the DIF analysis is sensitive to the number of data (Boone et al., 2014; Bond & Fox, 2015), in this study, only the gender variable (male and female subgroups) could be examined in the DIF context. This analysis was essential for ensuring test fairness and validity across different genders. Potential biases could be detected by identifying items that functioned differently for distinct genders, thus ensuring that test scores reflected true differences in the measured construct rather than mere biases against any group.

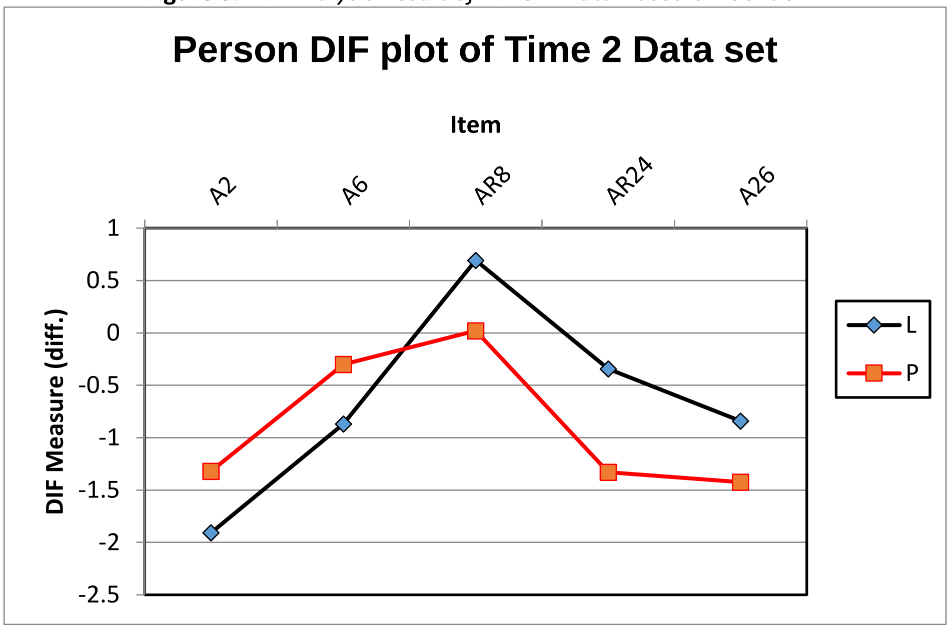
Table 5. *Differential Item Functioning (DIF) of the SCI based on Gender*

Data Set	Item	DIF Contrast	Rasch-Welch t	M-H Prob.
Time-1	A16	-0.68	-5.72	0.000
	A26	0.74	6.11	0.000
Time-2	A2	-0.59	-4.54	0.000
	A6	-0.57	-4.20	0.000
	AR8	0.67	5.24	0.029
	AR24	0.99	7.37	0.000
	A26	0.58	4.37	0.000

There were two items that had DIF for the gender variable in the Time-1 data (items A16 and A26) and five items in the Time-2 data (items A2, A6, AR8, AR24 and A26) out of the total 28 items in the SCI. As shown in Table 5, each of these items had values that were beyond the acceptable ranges for DIF contrast (between -0.5 to +0.5), Rasch-Welch t (between -2.0 to +2.0), and Mantel-Haenszel probability (larger than 0.05). However, the one item that had DIF in both data sets was A26, which meant six items in the SCI needed to be rechecked, especially the unfavourable ones (AR8 and AR24, two out of three items in the Time-2 data). This ensured a better-adapted Malaysian version of the instrument for future studies.

Among the items found to have DIF in the Time-2 data, as shown in Figure 3, there were several interesting findings. There were two items that the female students found more difficult to agree with, which were A2 ("the teacher treats students from various races with full respect") and A6 ("teachers always take care of their students"), than the male students. On the other hand, there were three items that the male students found more difficult to agree with (AR8, AR24, and A26). For instance, for A26 ("most students strive earnestly to achieve good grades"), the male students tended to disagree with this more than their female counterparts.

Figure 3: *DIF Analysis Result of Time-2 Data Based on Gender*



DISCUSSION

The objective of the present study was to validate the SCI used in a longitudinal study with secondary school students in Malaysia using the Rasch measurement model. As stated earlier, the SCI has been tested by many researchers (Bear et al., 2011, 2019; Yang et al., 2018, 2020) but has been mostly limited to Western societies and had mixed results. Past studies also show that the SCI has been mostly used in cross-sectional single studies and rarely in longitudinal studies (Marraccini et al., 2020). Thus, this paper presented a report of an SCI psychometric-attribute study conducted in the Malaysian context using the Rasch model as comprehensive evidence that showcased this measurement model possessing robust validation method. Moreover, the Rasch measurement model is an excellent validation tool that provides comprehensive information about the psychometric attributes of the instrument (Bond & Fox, 2015; William J Boone et al., 2017).

The findings of this present study indicated that the SCI scale is a valid tool to measure school climate within a sample of Malaysian high school students. The scale's unidimensionality, determined through principal component analysis of residuals, revealed that a primary dimension was captured over a minimum level of 20% of the variances in in-person responses (both Times-1 and 2 are close to 40%), with minimal noise levels that did not affect measurement quality (Liu & Lim, 2020). All reliability indices, such as person and item reliabilities and the Cronbach alpha, were excellent, indicating that both data sets were large enough for the SCI to have good internal consistency (Boone et al., 2017). The item separation indices (15.21 and 12.33 for Times-1 and 2) indicated good item difficulty even though the scale has 28 items. This means the scale can differentiate differing school climate perceptions above the threshold criteria (Fisher, 2005). This implies that the SCI's 28 items, which had good personal reliability, had a clear distinction of person ability levels.

In the rating scale analysis of the SCI, monotonic assumptions were fulfilled with an increase in measures in both data sets, with the MnSq indices confirming this, except for the beginning of the rating scale (Van Zile-Tamsen, 2017). However, for step calibration, both samples had better rating functionality (Embretson & Reise, 2000; Fayers & Machin, 2007). The visible rating scale in both Time-1 and Time-2 samples worked well for all categories; this indicates that the Rasch analysis was useful in identifying the effectiveness of the instrument's rating scale (Bond et al., 2015; Boone et al., 2014).

As the Rasch model uniquely emphasises individual-centred statistics, a comprehensive investigation of the item level could be conducted. Interestingly, nearly all items in both data sets of the SCI had good item psychometric attributes, except for three items (AR8, AR13 and AR24). These three were unfavourable items (negatively worded), which could have potentially confused respondents (DiStefano & Motl, 2009; Sliter & Zickar, 2014), thus requiring further revision. Regarding polarisation, the point measure correlations of all items (including the three unfavourable) were positive, showing all items measured in one construct.

For the item-person targeting, the findings showed that the SCI did not have a floor effect. However, about the ceiling effect, 3% of the sample could not be measured as precisely as those were outliers. This shows that the scale effectively measured the personal abilities in both waves of this longitudinal study. The mean person measure, which was 1.81 logits and 1.95 logits for Times-1 and 2, indicated that the SCI for both samples in this study showed better levels of school climate. The standard deviations of the person logits (1.39 for Time-1 and 1.59 for Time-2) were lower than the item logits (0.87 and 0.79), suggesting a wider variation of person abilities compared to item difficulty but still having good targeting (Liu & Lim, 2020).

Four demographic data were collected in this study, namely gender, age, educational level, and year of study. Among these, only gender could be used for differential item functioning (DIF) analysis because the number of persons in the subgroups was not varied for the other three (Adams et al., 2021; 2022). The DIF analysis showed that there was DIF in both data sets, with two items in Time-1 and six in Time-2. In total, six items in the SCI were detected to have DIF. Thus, further study is needed to make sure item bias is avoided. The other items of the SCI were considered fair, with negligible effects on measurement quality. The current study supports the Item Response Theory (IRT) model, which emphasizes that the Rasch model is a powerful approach for validating assessments at both the item and test levels to answer theory-guided research questions and hypotheses.

CONCLUSION

This longitudinal study utilising the SCI has demonstrated its robust psychometric properties. The Rasch model analysis confirmed its good reliability and validity, well-structured and cohesive rating scale, and sound dimensionality. Moreover, the SCI significantly enhanced the measuring precision and alignment with the underlying trait it aimed to assess. The longitudinal Rasch analysis supported the conclusion that the 28-item SCI was suitable for use as a research tool to assess school climate, particularly in developing countries and from an Eastern perspective. This study uniquely contributes to the understanding of the Rasch model's application in Eastern educational contexts, offering robust insights that can inform future research on culturally relevant assessment practices and their impact on learning outcomes. Moreover, the current study provides empirical evidence of valid and reliable school climate instruments tailored to developing and Eastern contexts. Future research efforts could further enrich this field by exploring different educational levels across various types of schools and educational institutions.

Limitations and Recommendations for Future Studies

The present research has enriched the existing body of knowledge by applying advanced research methodologies, including a longitudinal study and a focus on an Eastern cultural context (Cole & Maxwell, 2003; Podsakoff et al., 2012; Selig & Preacher, 2008). Nevertheless, it is not exempt from limitations. Firstly, the reliance on self-reported surveys introduces potential biases, a concern frequently acknowledged in the literature (Conway & Lance, 2010; Podsakoff et al., 2003, 2012). While the longitudinal design employed in this study represents progress in mitigating such biases, varying results might emerge if data were collected from multiple sources. Therefore, adopting a multimethod or multisource approach presents a promising methodological avenue for future exploration.

Secondly, the sample comprised 16-year-old secondary school students, potentially introducing age-related biases in the generalisability of findings. Furthermore, the study was confined to daily secondary schools in Selangor, Malaysia. This limitation arises from the diversity within Malaysia's secondary education system, which includes technical schools, vocational colleges, boarding schools, special education schools, sports schools, art schools, and religious schools (Ministry of Education, 2019). It should be noted that this study specifically targeted daily national schools due to their diverse student backgrounds, talents, and capabilities. Consequently, the findings may differ if a broader range of school types had been included in the study.

Another limitation of the current study involves the online nature of data collection across two waves. Recent literature suggests that online surveys may elicit non-serious responses, potentially compromising the validity of findings (Brough, 2018). Moreover, collecting data in multiple waves increases the risk of data attrition. To mitigate these issues, the researchers sought to broaden participation by engaging more schools willing to commit to the study, thereby enhancing the robustness and generalisability of the findings.

REFERENCES

Adams, D., Chuah, K.M., Sumintono, B. and Mohamed, A. (2022). Students' readiness for e-learning during the COVID-19 pandemic in a South-East Asian university: A Rasch analysis. *Asian Education and Development Studies,* 11(2), pp. 324-329. https://doi.org/10.1108/AEDS-05-2020-0100

Adams, D., Chuah, K.M., Sumintono, B. and Mohamed, A. (2021). Bricks to clicks: Students' engagement in e-learning during the COVID-19 pandemic. *Asia Pacific Journal of Educators and Education,* 36 (2), pp. 99-117 http://apjee.usm.my/APJEE_36_2_2021/apjee36022021_6.pdf

Akhtar, H., & Sumintono, B. (2023). A Rasch analysis of the international personality item pool big five markers questionnaire: Is longer better? *Primenjena Psihologija*, *16*(1), 3-28. https://doi.org/10.19090/pp.v16i1.2401

Andrich, D., & Marais, I. (2019). *A course in Rasch measurement theory, measuring in the educational, social and health sciences*. Springer.

Aldridge, J. M., & Blackstock, M. J. (2024). Assessing students' perceptions of school climate in primary schools. *Learning Environments Research*, *0123456789*. https://doi.org/10.1007/s10984-024-09492-2

Baumrind, D. (1966). Effects of authoritative parental control on child behavior. *Child Development*, *37*(4), 887–907.

Bear, G. G., Gaskins, C., Blank, J., & Chen, F. F. (2011). *Delaware School Climate Survey Student Version 6-12*. 2018–2019.

Bear, G. G., Gaskins, C., Blank, J., & Chen, F. F. (2019). *Technical manual for the Delaware school survey: Scales of school climate; bullying victimization; student engagement; positive, punitive, and social emotional learning techniques; and social emotional competencies*. Centre for Disabilities Studies. https://doi.org/10.1016/j.jsp.2011.01.001

Bear, G. G., Yang, C., Mantz, L. S., & Harris, A. B. (2017). School-wide practices associated with school climate in elementary, middle, and high schools. *Teaching and Teacher Education*, *63*, 372–383. https://doi.org/10.1016/j.tate.2017.01.012

Bond, T. G., & Fox, C. M. (2015). Applying the Rasch Model. In *Applying the Rasch model: Fundamental Measurement in the Human Sciences*.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer.

Boone, William J, Noltemeyer, A., Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis : A primer for school psychology researchers and practitioners. *Cogent Education*, *25*. https://doi.org/10.1080/2331186X.2017.1416898

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*(9), 593–604. https://doi.org/10.1111/josh.12186

Brough, P. (2018). Advanced Research Methods for Applied Psychology. In *Advanced Research Methods for Applied Psychology*. https://doi.org/10.4324/9781315517971

Bryk, A. S. (2010). Schools for. *Phi Delta Kappa International*, *April*, 23–30.

Cohen, J., Mccabe, E. M., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record*, *111*(1), 180–213.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 558–577. https://doi.org/10.1037/0021-843X.112.4.558

Consortium on Chicago School Research. (2007). *Consortium Survey Measures*.

Conway, J. M., & Lance, C. E. (2010). What reviewers should expect from authors regarding common method bias in organizational research. *Journal of Business and Psychology*, *25*(3), 325–334. https://doi.org/10.1007/s10869-010-9181-6

Cornell, D. (2014). *Overview of the Authoritative School Climate Survey*. https://safesupportivelearning.ed.gov/survey/authoritative-school-climate-survey

Creswell, W. John & Creswell, J. D. (2018). Research design: Qualitative, quantitative adn mixed methods approaches. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9).

Darling-Hammond, L., Bae, S., Cook-Harvey, C. M., Lam, L., Mercer, C., Podolsky, A., Stosich, E. L., Cook-harvey, L., Mercer, C., Podolsky, A., Stosich, E. L., Cook-Harvey, C. M., Lam, L., Mercer, C., Podolsky, A., & Stosich, E. L. (2016). Pathways to new accountability through the Every Student Succeeds Act. In *Palo Alto: Learning Policy Institute. This report retrieved from https://learningpolicyinstitute. org/wp-content/uploads/2016/04/Pathways_New-Accountability_Through_Every_Student_Succeeds_Act_04202016. pdf* (Issue April). http://learningpolicyinstitute.org/our-work/publications-resources/

Embretson, S. E., Reise, S., & Reise, S. P. (2000). *Item Response Theory for Psychologists (Multivariate Applications Book Series.)*.

Field, A. J., Bowles, S., & Gintis, H. (1977). Schooling in capitalist America: Educational Reform and the contradictions of economic life. In *The Journal of Human Resources* (Vol. 12, Issue 2). https://doi.org/10.2307/145392

Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, *21(1)*, 1095.

Forsberg, C., Chiriac, E. H., & Thornberg, R. (2021). Exploring pupils' perspectives on school climate. *Educational Research*, *63*(4), 379–395. https://doi.org/10.1080/00131881.2021.1956988

Freiberg, H. J. (1999). *School climate: Measuring, improving and sustaining healthy learning environments*. Falmer Press, Taylor & Francis Group. http://library1.nida.ac.th/termpaper6/sd/2554/19755.pdf

Gonder, P. O. (1994). Improving school climate & culture. In *American association of school administrators* (Issue LMI). https://files.eric.ed.gov/fulltext/ED371485.pdf

Graham, J. (2022). Explaining the racial school climate gap: Evidence from Georgia. *AERA Open*, *8*(1), 1–19. https://doi.org/10.1177/23328584221131529

Grazia, V., & Molinari, L. (2020). School climate multidimensionality and measurement: A systematic literature review. *Research Papers in Education*, *00*(00), 1–27. https://doi.org/10.1080/02671522.2019.1697735

Grazia, V., & Molinari, L. (2021). School climate research: Italian adaptation and validation of a multidimensional school climate questionnaire. *Journal of Psychoeducational Assessment*, *39*(3), 286–300. https://doi.org/10.1177/0734282920967141

Gregory, A., & Cornell, D. (2009). "Tolerating" adolescent needs: Moving beyond zero tolerance policies in high school. *Theory into Practice*, *48*(2), 106–113. https://doi.org/10.1080/00405840902776327

Hamlin, D. (2020). Can a positive school climate promote student attendance? evidence from New York City. *American Educational Research Journal*, *XX*(X), 1–28. https://doi.org/10.3102/0002831220924037

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, *103*(6), 2052–2086. https://doi.org/10.1257/aer.103.6.2052

Hinkin, T. R., Tracey, J. B., & Enz, C. A. (1997). Scale construction: Developing reliable and valid measurement instruments. *Journal of Hospitality & Tourism Research*, *21*(1), 100–120. https://doi.org/10.1177/109634809702100108

Hoy, W. K., & Miskel, C. G. (2013). Educational administration: Theory, research, and practice. In *Journal of Educational Administration: Vol. 9th Editio* (Issue 1). https://doi.org/10.1108/09578230810849862

Johnson, W. L., Johnson, A. M., Ryan, J. M., Edens, K. M., & Dixon, P. N. (1995). A Rasch analysis of factor derived from the Charles F. Kettering Ltd. School climate profile. *Educational and Psychological Measurement*, *55*(2), 456–467. https://doi.org/https://doi.org/10.1177/0013164495055003013

Larson, K. E., Nyuyen, A. J., Oroszo Solis, M. G., Humphreys, A., Bradshaw, C. P., & Johnson, S. L. (2020). A systematic literature review of school climate in low and middle income countries. *International Journal of Educational Research*, 20.

Lenz, A. S., Rocha, L., & Aras, Y. (2021). Measuring school climate: A systematic review of initial development and validation studies. *International Journal for the Advancement of Counselling*, *43*(1), 48–62. https://doi.org/10.1007/s10447-020-09415-9

Lippman, L., Guzman, L., & Moore, K. A. (2012). *Measuring flourishing among youth*. https://cms.childtrends.org/wp-content/uploads/2013/05/FlourishingChildren.pdf

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge Taylor & Francis Group.

Marciniak, J., Hirschi, A., Johnston, C. S., & Haenggli, M. (2020). Measuring career preparedness among adolescents: Development and validation of the career resources questionnaire—adolescent version. *Journal of Career Assessment*, 1–17. https://doi.org/10.1177/1069072720943838

Marraccini, M. E., Fang, Y., Levine, S. P., Chin, A. J., & Pittleman, C. (2020a). Measuring student perceptions of school climate: A systematic review and ecological content analysis. *School Mental Health*, *12*(2), 195–221. https://doi.org/10.1007/s12310-019-09348-8

Marraccini, M. E., Fang, Y., Levine, S. P., Chin, A. J., & Pittleman, C. (2020b). Measuring student perceptions of school climate: A systematic review and ecological content analysis. *School Mental Health*, *12*(2), 195–221. https://doi.org/10.1007/s12310-019-09348-8

Mateos, N. E., Fernández-Zabala, A., Palacios, E. G., & Díaz-De-cerio, I. I. D. L. F. (2021). School climate and perceived academic performance: Direct or resilience-mediated relationship? *Sustainability (Switzerland)*, *13*(1), 1–14. https://doi.org/10.3390/su13010068

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, *12*(1), 23–44. https://doi.org/10.1037/1082-989X.12.1.23

Ministry of Education. (2019). *Dasar pendidikan kebangsaan [National education policy]*. https://www.moe.gov.my/images/KPM/UKK/2018/BUKU DPK JILID KEEMPAT VERSI AKHIR.pdf

Molinari, L., & Grazia, V. (2023). Students' school climate perceptions: Do engagement and burnout matter? *Learning Environments Research*, *26*(1), 1–18. https://doi.org/10.1007/s10984-021-09384-9

Monsillion, J., Zebdi, R., & Romo-Desprez, L. (2023). School mindfulness-based interventions for youth, and considerations for anxiety, depression, and a positive school climate—A systematic literature review. *Children*, *10*(5). https://doi.org/10.3390/children10050861

Mullis, I. V. ., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*.

National Center on Safe Supportive Learning Environments. (2016). *Education Department School Climate Survey (EDSCLS)*. https://safesupportivelearning.ed.gov/edscls

Nishimura, T., Wakuta, M., Tsuchiya, K. J., Osuka, Y., Tamai, H., Takei, N., & Katayama, T. (2020). Measuring school climate among Japanese students—development of the Japan school climate inventory (JASC). *International Journal of Environmental Research and Public Health*, *17*(12), 1–11. https://doi.org/10.3390/ijerph17124426

OECD. (2024). *OECD Contributions to the 2030 Agenda and Beyond*.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*, 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Pogodzinski, B., Cook, W., Lenhoff, S. W., & Singer, J. (2022). School climate and student mobility. *Leadership and*

Policy in Schools, *21*(4), 984–1004. https://doi.org/10.1080/15700763.2021.1901121

Richard W. Brislin. (1970). Back translation for cross-cultural research. In *Journal of cross cultural psychology* (Vol. 1, Issue 3, pp. 185–216).

Rohatgi, A., & Scherer, R. (2020). Identifying profiles of students' school climate perceptions using PISA 2015 data. *Large-Scale Assessments in Education*, *8*(1). https://doi.org/10.1186/s40536-020-00083-0

Ryberg, R., Her, S., Temkin, D., Madill, R., Kelley, C., Thompson, J., & Gabriel, A. (2020). Measuring school climate: Validating the education department school climate survey in a sample of urban middle and high school students. *AERA Open*, *6*(3), 1–21. https://doi.org/10.1177/2332858420948024

Schultz, T. W. (1961). Invest in human capital. *The Amercian Economic Review Vol. 51*, *51*(No. 1), 1–20.

Selig, J. P., & Preacher, K. J. (2008). *Monte Carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects [Computer Software]*. http://quantpsy.org/

Stockard, J., & Mayberry, M. (1992). *Effective educational environments*.

Tucker-Drob, E. M., Briley, D. A., Engelhardt, L. E., Mann, F. D., & Harden, K. P. (2016). Genetically-mediated associations between measures of childhood character and academic achievement. *Journal of Personality and Social Psychology*, *111*(5), 790–815. https://doi.org/10.1037/pspp0000098

Xu, M., MacDonnell, M., Wang, A., & Elias, M. J. (2023). Exploring social-emotional learning, school climate, and social network analysis. *Journal of Community Psychology*, *51*(1), 84–102. https://doi.org/10.1002/jcop.22881

Yang, C., Bear, G. G., & May, H. (2018). Multilevel associations between school-wide social-emotional learning approach and student engagement across elementary, middle, and high schools. *School Psychology Review*, *47*(1), 45–61. https://doi.org/10.17105/SPR-2017-0003.V47-1

Yang, C., Chan, M. K., & Ma, T. L. (2020). School-wide social emotional learning (SEL) and bullying victimization: Moderating role of school climate in elementary, middle, and high schools. *Journal of School Psychology*, *82*(March), 49–69. https://doi.org/10.1016/j.jsp.2020.08.002

Zapf, D., Dormann, C., & Frese, M. (1996). Longitudinal studies in organizational stress research: A review of the literature with reference to methodological issues. *Journal of Occupational Health Psychology*, *1*(2), 145–169. https://doi.org/10.1037/1076-8998.1.2.145

Zynuddin, S. N., Kenayathulla, H. B., & Sumintono, B. (2023). The relationship between school climate and students' non-cognitive skills: A systematic literature review. *Heliyon*, *9*(4), e14773. https://doi.org/10.1016/j.heliyon.2023.e14773