

School Climate in American Secondary Schools: A Psychometric Examination of PISA 2009 School Climate Scale

Letao Sun¹ & Kenneth Royal^{2,*}

¹Department of Educational Policy Studies & Evaluation, University of Kentucky, Lexington, KY, USA

²Department of Clinical Sciences, North Carolina State University, Raleigh, NC, USA

*Correspondence: College of Veterinary Medicine, NC State University, 1060 William Moore Dr., Raleigh, NC, 27607, USA. E-mail: kdroyal2@ncsu.edu

Received: July 9, 2017

Accepted: July 26, 2017

Online Published: August 20, 2017

doi:10.5430/jct.v6n2p6

URL: <https://doi.org/10.5430/jct.v6n2p6>

Abstract

The purpose of this study was to examine the quality of the Program for International Student Assessment (PISA) 2009 school climate survey instrument and evaluate perceptions of secondary school principals' located in the United States about school climate using an Item Response Theory (IRT) methodological approach. In particular, this study sought to determine if the instrument's items are of sufficient psychometric quality to effectively measure schools' climate status in the United States. Collectively, results indicate the School Climate Assessment (SCA) scale is of sufficient psychometric quality to effectively measure schools' climate status in the United States. However, there are areas for which the instrument can be improved. Recommendations for improvement are provided.

Keywords: *school climate; PISA; secondary education; international assessment; educational quality; psychometrics*

1. Introduction

1.1 Background

School climate is a key concern in discussions about school effectiveness because of its profound impact on students' behavior and learning outcomes (Adelman & Taylor, 2011; Hattie, 2003). School climate has been systematically studied in organizational research to examine school effectiveness (Cohen, McCabe, Michelli, & Pickeral, 2009; Creemers & Reezigt, 1999; Kreft, 1993). Although there is no commonly accepted definition for school climate, most scholars and researchers agree that school climate relates to the subjective experiences of members (e.g., students, parents and school personnel) within the school and reflects the values, goals, interpersonal relationships, teaching and learning practices, safety, as well as external environment (Cohen, 2006; Freiberg, 1999; Loukas, 2007).

A series of studies have found a positive school climate is associated with academic achievement and positive youth development (Berkowitz & Bier, 2006; Cohen, McCabe, Michelli, & Pickeral, 2009; Greenberg et al., 2003; Griffith, 1999). For example, Griffith (2002) found that both student and school level perceptions of school climate are positively correlated with students' GPA. Similarly, Gareau et al. (2009) found a positive relationship between school climate factors and student achievement outcomes across all organizational levels. Cohen (2001) concluded that a safe, caring, and responsive school climate can foster effective risk prevention and health promotion efforts. Research also found that positive school climate can create a positive climate for learning through promoting cooperative learning, building group cohesiveness, and gaining respect and mutual trust among students (Finnan, Schnepel, & Anderson, 2003; Ghaith, 2003). Therefore, it is essential to accurately measure school climate in order to improve the quality of education in school.

1.2 Purpose

The purpose of this study was to examine the quality of the Program for International Student Assessment (PISA) 2009 school climate survey instrument and evaluate perceptions of secondary school principals' located in the United States about school climate using an Item Response Theory (IRT) methodological approach. In particular, this study

sought to determine if the items are of sufficient psychometric quality to effectively measure schools' climate status in the United States.

2. Method

2.1 Data Sources

The primary database used in this research was derived from the Program for International Student Assessment (PISA) conducted in 2009. PISA is an internationally standardized assessment that measures students' capabilities in mathematics, reading, and science literacy. According to Organization for Economic Co-operation and Development (OECD), PISA focuses on young people's ability to use their knowledge and skills to meet real-life challenges, rather than merely determining if students have mastered a specific school curriculum (Sun, Bradley & Toland, 2013). Beginning in 2000, PISA was administered every three years to groups of 15-year-old students in principal industrialized countries. One subject (or literacy area) was the focus at each administration. In the 2009 PISA, three questionnaires were designed for students, parents and schools respectively. Each form contained a number of scales to assess student, parent, and school effects on reading achievement. The focus of this study is on the measures of school climate from the perspective of school principals.

2.2 Instrumentation

The School Climate Assessment (SCA) scale represents a set of school-related variables that can influence student performance and explain differences in teacher effectiveness. The index of factors affecting school climate was derived from school principals' reports on the extent to which the learning of students was hindered by 13 items (Table 1). A four-point rating scale was used with the following categories: 1 = Not at all; 2 = Very little; 3 = To some extent; and 4 = A lot. As all items were inverted for scaling, higher values indicate positive teacher and student behaviors.

Table 1. Items Appearing on the SCA.

1.	Teachers' low expectations of students
2.	Student absenteeism
3.	Poor student-teacher relations
4.	Disruption of classes by students
5.	Teachers not meeting individual students' needs
6.	Teacher absenteeism
7.	Students skipping classes
8.	Students lacking respect for teachers
9.	Staff resisting change
10.	Student use of alcohol or illegal drugs
11.	Teachers being too strict with students
12.	Students intimidating or bullying other students
13.	Students not being encouraged to achieve their full potential

2.3 Sample

The present study utilized 2009 PISA data from participants located in the United States. A total of 165 schools were represented, but 3 (1.8%) schools contained missing data and subsequently were removed from the analysis. The final sample consisted of 162 schools.

2.4 Procedures

Parallel Factor Analysis (PFA) was conducted to examine the underlying assumptions prior to the IRT analysis. The PFA was conducted using SPSS statistical software. Excellent model fit when constraining nearly all items to load on one factor supported further data analysis using IRT. The Rasch Rating Scale Model (RRSM) (Andrich, 1978) was utilized to evaluate the psychometric properties of the SCA. The RRSM is an appropriate measurement model for measuring ordinal survey response data and has been used extensively in the research literature (Royal, 2010; Royal and Gonzalez, 2016; Wolfe et al, 2004). Winstep measurement software (Linacre, 2016) was used to perform the RRSM analysis using joint maximum likelihood estimation (JMLE) procedures.

3. Results

A fundamental requirement for the RRSM is that data are sufficiently unidimensional. Thus, a Parallel Factor Analysis (PFA) to assess the dimensionality of the data was performed. For this instrument, 87.8% of the variability in the final scale score was due to school climate variability, and the remaining 12.2% was due to random measurement error. Three factors with eigenvalues greater than 1 were detected and retained according to Kaiser-Guttman's rule. However, according to Gorsuch (1983), only the ratio of the first (5.41) to second eigenvalue (1.48) was greater than 3, suggesting data were sufficiently unidimensional. In addition, the sequence plot of PFA suggests only one factor exists above the crossing point between the raw data and 95th percentile, provide additional evidence of unidimensional data.

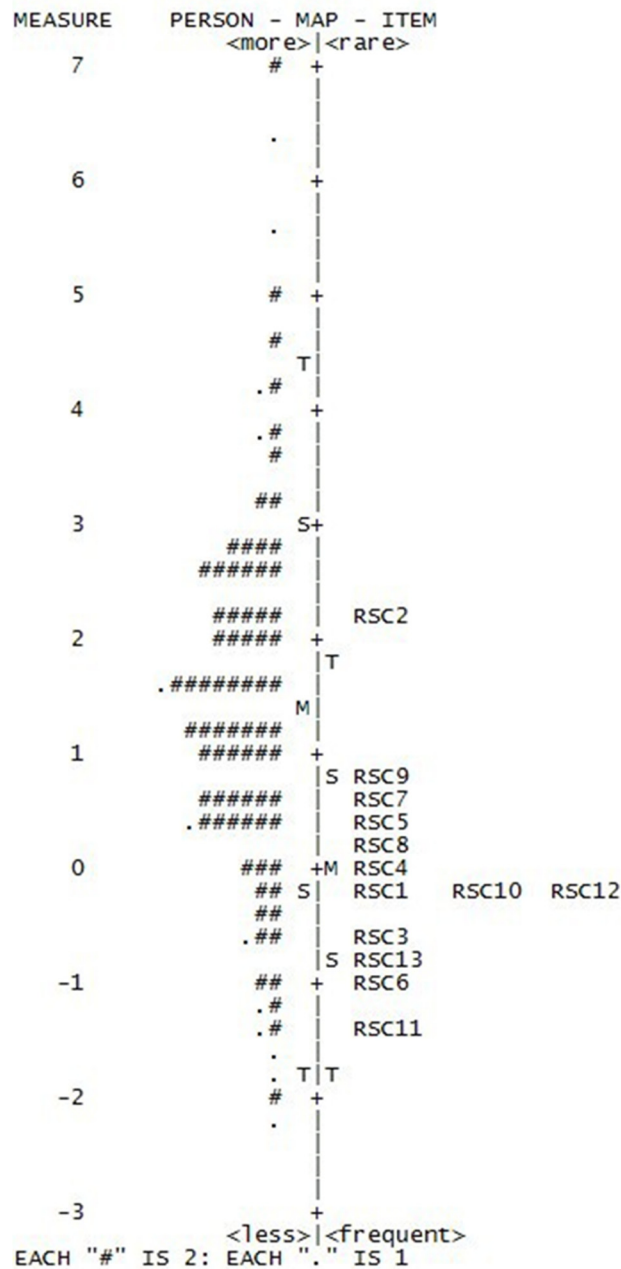


Figure 1. Wright Map

Next, the quality of the rating scale was evaluated. In short, all participants should interpret a well-functioning rating scale the same way, and make use of all categories. According to Linacre (2002), each rating scale category should have a minimum of 10 observations and structure calibration measures should advance in a step-wise manner from negative to positive. Here, the minimum number of observations was 55 for response option 4 ('A lot'), and categories advanced in a step-wise manner as expected. Although respondents made full uses of the scale, the response pattern largely was unimodal in nature. Specifically, responses were positively-skewed with most respondents selecting response option 1 ('Very little'). Collectively, there is sufficient psychometric evidence to conclude the rating scale functioned well.

An investigation of overall data-to-model fit revealed 2 of the 162 (1.2%) schools provided extreme scores and were subsequently removed from the analysis. The infit mean square values of .99 ($SD = .57$) and outfit mean square values of .99 ($SD = .59$) closely approximated 1.00, indicating excellent data-to-model fit. In addition, standardized means (ZSTD) were $-.1$ ($SD = 1.4$) for both infit and outfit measures. Values were between the acceptable range of -2 and 2 (Smith, 1996, 2000), thus further evidencing good data-to-model fit.

The overall model-data fit results at the item level also provided evidence of excellent fit with infit mean square values of .99 ($SD = .23$) and outfit mean square values of .99 ($SD = .25$). Additionally, infit and outfit ZSTD measures were $-.2$ ($SD = 2.0$) providing further evidence of good data-to-model fit. Further, inspection of fit statistics indicating all items were within the acceptable range of $.6$ to 1.4 (Wright & Linacre, 1994), and point-measure correlations ranged from $.45$ to $.72$ indicating excellent discriminatory abilities (Linacre, 2017).

With respect to score reliability, the standardized coefficient alpha reliability statistic was $.879$, indicating moderate-high levels of reliability (Royal & Hecker, 2015). Separation measures indicate the number of statistically distinguishable levels within the data. Here, person separation measures were 2.37 (real) and 2.64 (model), and item separation measures were 5.36 (real) and 5.62 (model). These values indicate the instrument is capable of spreading results into several statistically distinguishable levels.

Next, the item hierarchy was examined relative to the person measures by way of the Wright map. This visual inspection helps discern the extent to which items were appropriately targeted to the sample, and if any items presented egregiously predictable or unpredictable responses. The Wright map is presented in Figure 1.

4. Discussion

4.1 Psychometric Properties of the SCA

Using Messick's (1989) unified framework for interpreting construct validity evidence, results indicate there is adequate psychometric evidence with respect to the substantive, content, generalizability and structural aspects of validity. Specifically, data were sufficiently unidimensional, which speaks to the substantive aspect of validity. Items fit the Rasch Rating Scale Model's expectations and were evidenced to discriminate well which speaks to the content aspect of validity. The rating scale was evidenced to function well which speaks to the structural aspect of validity. Finally, scores were highly reliable and separation measures indicated the instrument was capable to stratifying measures into a number of statistically distinguishable levels. The authors present no evidence to speak to the external or consequential aspects of validity (Royal & Puffer, 2014). Collectively, this is ample evidence to conclude the SCA is a psychometric sound instrument capable of measuring school climate from the perspective of school principals.

4.2 Implications

This study used an IRT to investigate the psychometric properties of the school climate assessment appearing on the 2009 PISA. This study was necessary because the international scope of the PISA and the comparative nature of the assessment often results in many consumers of the data questioning its validity. Because results evidence the SCA is an appropriate measure of school climate, consumers of PISA results (particularly in the United States) may have additional assurance that the results collected as far back as 2009 to present day are evidenced to be valid and trustworthy for this particular scale. Additionally, this study will help researchers and school administrators understand how schools differ in their school climate and if school climate can be used as a predictor of student performance on international assessments.

4.3 Limitations and Future Research

As with any study, this work also possesses some limitations. First, this study like most self-reported survey instruments may contain some element of self-report bias. Self-report bias occurs when a person has a tendency to

give affirmative answers regardless of the content of the question, is unable to understand/judge the question accordingly, or presents his or herself in a favorable view, rather than expressing true behavior (Dodd-McCue, & Tartaglia, 2010). As a result, self-report bias may inflate correlations of constructs across time and reduce unexplained variance available for latent variables (Marsh, 1993). Because one cannot confirm that reported behavior is comparable to actual behavior, the association between perceived school climate and actual school climate is truly unknown.

PISA employed a multi-stage probability sampling design, which considers multiple levels of sampling units in order to gain a representative sample proportional to the size of the desired study population. Since the primary focus of this study was to examine the psychometric properties of the school climate survey instrument and compare climate scores for each school, survey weights and complexity were not considered in the analysis. Future studies on school climate are advised to apply psychometric techniques in a manner consistent with the complex survey design scenario in order to estimate latent scores for each school.

Future studies should investigate whether PISA school climate items function similarly across different types of schools (e.g., public vs. private) within the United States and across countries. If following the IRT-based analytic approach described in this paper, other researchers may wish to assess differential item functioning (DIF) which can speak to the systematic aspect of validity.

5. Conclusion

The current study applied an IRT model to assess the quality of School Climate Assessment survey using the 2009 PISA U.S. school sample. Beyond the limitations of traditional statistical approaches to examine an instrument's quality (e.g., Cronbach's alpha, factor analysis, etc.), this study provide a more detailed and comprehensive assessment of the survey instrument and how it interacts with sample participants. Collectively, the findings from this study conclude the SCA instrument is of sufficient psychometric quality to effectively measure schools' climate status in the United States. However, an area for instrument improvement includes adding a few discerning items to help discriminate school climate for schools that are above average.

References

- Adelman, H. S., & Taylor, L. (2011). Expanding school improvement policy to better address barriers to learning and integrate public health concerns. *Policy Futures in Education*, 9(3), 431-446. <https://doi.org/10.2304/pfie.2011.9.3.431>.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573. <http://dx.doi.org/10.1007/BF02293814>
- Berkowitz, M. W., & Bier, M. C. (2005). What works in character education: A research-driven guide for educators. St. Louis, MO: Character Education Partnership. Retrieved from <http://www.specialneedsbsa.org/OrgHeaders/2523/What%20Works%20in%20Character%20Education.pdf>
- Cohen, J. (2001). *Caring Classrooms/Intelligent Schools: The Social Emotional Education of Young Children. Series on Social Emotional Learning*. Teachers College Press.
- Cohen, J. (2006). Social, emotional, ethical and academic education: Creating a climate for learning, participation in democracy and well-being. *Harvard Educational Review*, 76(2), 201-237. <https://doi.org/10.17763/haer.76.2.j44854x1524644vn>.
- Cohen, J., McCabe, L., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *The Teachers College Record*, 111(1), 180-213.
- Creemers, B. P., & Reezigt, G. J. (1999). The role of school and classroom climate in elementary school learning environments. In H. J. Freiberg (Ed.), *School climate: Measuring, improving and sustaining healthy learning environments* (pp 30-47). Philadelphia, PA: Falmer Press.
- Dodd-McCue, D., & Tartaglia, A. (2010). Self-report response bias: Learning how to live with its diagnosis in chaplaincy research. *Chaplaincy Today*, 26(1), 2-8. <https://doi.org/10.1080/10999183.2010.10767394>.
- Finnan, C., Schnepel, K. C., & Anderson, L. W. (2003). Powerful learning environments: The critical link between school and classroom cultures. *Journal of Education for Students Places at Risk*, 8(4), 391-418. https://doi.org/10.1207/s15327671espr0804_2.

- Freiberg, H. J. (Ed.). (1999). *School climate: Measuring, improving and sustaining healthy learning environments*. Psychology Press.
- Gareau, S., Monrad, D. M., DiStefano, C., May, R. J., Price, K. J., Ishikawa, T., & Mindrila, D. L. (2009). The relationship between school climate and school performance. Retrieved from <http://www.ed.sc.edu/scepc/Documents/EOC%20Climate/SCEPC%202010%20SCEPUR%20submission.pdf>
- Ghaith, G. (2003). Effects of the learning together model of cooperative learning on English as a foreign language reading achievement, academic self-esteem, and feelings of school alienation. *Bilingual Research Journal*, 27(3), 451-474. <https://doi.org/10.1080/15235882.2003.10162603>.
- Gorsuch, R. L. (1983). Three methods for analyzing limited time-series (N of 1) data. *Behavioral Assessment*, 5, 141-154.
- Greenberg, M. T., Weissberg, R. P., O'Brien, M. U., Zins, J. E., Fredericks, L., Resnik, H., & Elias, M. J. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist*, 58(6-7), 466-474. <https://doi.org/10.1037/0003-066x.58.6-7.466>.
- Griffith, J. (1999). The school leadership/school climate relation: Identification of school configurations associated with change in principals. *Educational Administration Quarterly*, 35(2), 267-291. <https://doi.org/10.1177/00131619921968545>.
- Griffith, J. (2002). A multilevel analysis of the relation of school learning and social environments to minority achievement in public elementary schools. *The Elementary School Journal*, 102(5), 349-366. <https://doi.org/10.1086/499708>.
- Hattie, J. (2003). *Teachers make a difference: What is the research evidence?* Camberwell: Australian Council for Educational Research.
- Kreft, I. G. (1993). Using multilevel analysis to assess school effectiveness: A study of Dutch secondary schools. *Sociology of Education*, 66(2), 104-129. <https://doi.org/10.1037/e672122011-001>.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Linacre, J. M. (2016). *WINSTEPS®* (Version 3.81). Computer Software. Beaverton, OR: Winsteps.com.
- Linacre, J. M. (2017). *Correlations: Point-biserial, point-measure, residual*. Retrieved from <http://www.winsteps.com/winman/correlations.htm>.
- Loukas, A. (2007). What is school climate? *Leadership Compass*, 5(1), 1-3.
- Marsh, H. W. (1993). Stability of individual differences in multiwave panel studies: Comparison of simplex models and one-factor models. *Journal of Educational Measurement*, 30, 157-183. <https://doi.org/10.1111/j.1745-3984.1993.tb01072.x>.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). New York: Macmillan.
- Royal, K. D. (2010). Making meaningful measurement in survey research: A demonstration of the utility of the Rasch model. *IR Applications*, 28, 1-16.
- Royal, K. D., & Hecker, K. G. (2015). Understanding reliability: A review for veterinary educators. *Journal of Veterinary Medical Education*. <http://dx.doi.org/10.3138/jvme.0315-030R>
- Royal, K. D., & Gonzalez, L. M. (2016). An evaluation of the psychometric properties of an advising survey for medical and professional program students. *Journal of Educational and Developmental Psychology*, 6(1), 195-203. <https://doi.org/10.5539/jedp.v6n1p195>.
- Royal, K. D., & Puffer, J. C. (2014). The consequential validity of ABFM examinations. *Journal of the American Board of Family Medicine*, 27, 430-431. <https://doi.org/10.3122/jabfm.2014.03.140089>.
- Smith, R. M. (1996). Polytomous mean-square fit statistics. *Rasch Measurement Transactions*, 10, 516-517.
- Smith, R. M. (2000). Fit analysis in latent trait measurement models. *Journal of Applied Measurement*, 1, 199-218.
- Sun, L., Bradley, K. D., & Toland, M. D. (2013). School computer use and student academic performance in secondary schools. *Computers in Education Journal*, 4(3), 84-94.
- Wolfe, W. W., Ray, L. M., & Harris, D. C. (2004). A Rasch analysis of three measures of teacher perception generated from the School and Staffing Survey. *Educational and Psychological Measurement*, 64(5), 842-860.

<https://doi.org/10.1177/0013164404263882>.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.