Exploring Differences in School Climate Among School Types

Audrey Conway Roberts, Ph.D., Bowling Green State University
R. Joseph Waddington, Ph.D., University of Notre Dame
Shannon Sampson, Ph.D., University of Kentucky

Corresponding author: Audrey Conway Roberts
Email: audrobe@bgsu.edu
ORCID: https://orcid.org/0000-0002-1402-1887

The operationalization of school climate is inconsistent, adding to an unwieldy existing body of climate instruments, bringing into question measurement validity concerns. Further, there is minimal literature regarding if comparisons can be made between different school sectors (i.e., public, private, charter). We seek to understand if a commonly used school climate scale (5Essentials of School Climate and Culture Survey) in a representative sample of teachers can be used to make valid comparisons between sectors, and if items might exhibit bias toward any school type using a Differential Item Functioning (DIF) approach. Although not misfitting, a great deal of significant DIF at different magnitudes exists in the items. We propose caution should be used when interpreting research comparing between and across school sectors without explicitly exploring measurement validity.

Keywords: school climate, teachers, instrument validation and use, DIF analysis

Exploring Differences in School Climate Among School Types

As the United States Department of Education and individual State Departments of Education recognize the importance of accountability measures beyond indicators of academic success or failure, understanding and measuring school climate has received additional attention (Barksdale et al., 2021; Duckworth & Yeager, 2015). Climate measures are widely used in many high-stakes accountability systems throughout the United States (e.g., Illinois, South Carolina, New Mexico, California) to evaluate teachers, principals, and schools (Clifford et al., 2012). The Every Student Succeeds Act (ESSA) requires a School Quality and Student Success indicator to be submitted by each state.

Concurrent with the recent focus on capturing non-academic measures, such as climate, is the growth of school choice programs and policies in most states. This includes the expansion charter schools and publicly funded programs expanding access to private schools (e.g., vouchers, tax credit scholarships). Charter schools have typically been subject to the same data reporting and accountability requirements as traditional public schools, and in recent years, choice-program participating private schools are as well (Berends & Waddington, 2018). As a result, one primary concern that has not been addressed in school climate measurement research is whether real differences are perceived by teachers in different school types or if there could be differential item functioning present in measures used.

In order to make comparisons across different contexts, it is critical to have a shared understanding of what a construct is and what it is not (Sampson et al., 2021). Furthermore, a measure with strong evidence of validity allows users to make meaningful comparisons within schools and between schools, and to identify areas of strength and suggest improvements. As outcomes from climate measures are commonly linked to high-stakes accountability decision-

making through teacher and school level evaluation systems, measures must have evidence of validity within the contexts they are being used. The potential impact of this validation would allow for researchers and policymakers to establish a true shared survey measure with which comparisons could be made.

This study further explores the utility of an established instrument (5Essentials of School Climate and Culture Survey) measuring educational climate from a survey of teachers in a multi-sector state representative sample. Within this scale, Gilman et al (2021) has found DIF present in several items at the student level between younger and older students. We seek to answer if the scale can be used to make valid comparisons between schools, and if the items comprising the instrument might exhibit bias toward any school type using a Differential Item Functioning technique. This paper will examine the following question: In what ways do item responses from teachers on their perceptions of school climate differ between school types?

Literature Review

School climate encompasses a variety of dimensions representing shared experiences between students, teachers, administrators, and parents regarding the "norms, goals, values, interpersonal relationship, teaching and learning practices, and organizational structures" (Cohen et al., 2009, p. 182). School climate is shaped by both "structural and systemic attributes" that are relatively stable, however may be perceived differently by each individual member (Schweig et al., 2019, p. 2). Additionally, there is no stated consensus regarding the essential dimensions of school climate, rather, there are generally accepted subarea indicators that are frequently included pertaining to safety, interpersonal relationships, teaching and learning, the institutional environment, and to a certain extent, the presence of technology and the use of social media (NSCC, 2021). School climate has been found to impact students' academic achievement and self-efficacy, well-being, and social emotional competencies (Berkowitz et al., 2019; Wang et

al., 2016; Yang et al., 2020; Zysberg & Schwabsky, 2021). For teachers, school climate is related to teaching self-efficacy, job satisfaction, and perceptions of leadership practices (Aldridge & Fraser, 2016; Sanchez et al., 2022; Zakariya, 2020).

Elements of School Climate

Many elements of an academic experience are shaped by state, district, and/or school level policies that would impact all members of a school community (e.g., hiring policies, facility management and resources available, safety, professional support staff). However, other parts of a learning environment (e.g., the teacher, the students, and the quality of instruction) are dynamic, varying significantly from classroom to classroom, and from year to year. Students in primary schools may stay with their classmates and rotate to different teachers, while students in secondary schools may rotate with different groups of teachers and students each day. Teachers may have multiple distinct perspectives of each classroom environment, or one perspective of the classroom/s they facilitate. In addition, these perceptions of the classroom have a reciprocal quality, with student behavior adapting within the educational environmental context initiated and maintained by the teacher (O'Brennan et al., 2014).

Broadly, climate research examines the multidimensional structure of the school environment from one or more perspectives (e.g. students, teachers, staff, and/or parents). School and classroom climate measures are comprised of many of the same elements (e.g., learning environment, student engagement, safety, etc.). Schweig et al. (2019) noted "significant overlap" exists between school and classroom climate, although distinctions exist in "key ways" (p. 5). Both school and classroom climate share elements of academic and emotional support, relationships, engagement, and safety. Key differences lie in the relative specificity of various learning environments.

However, the individual classroom and the school, which contains many individual classrooms, are not equivalent and must be seen as distinct measurable concepts (Schweig et al., 2019). Classroom climate attends to the specific experiences and perceptions of students and teachers and how the elements historically nested within the classroom environment have functioned. Classroom climate has been treated as unidimensional and multidimensional and frequently school climate studies refer to climate as both multidimensional, in varying capacities, and as being a general indicator, comprised of various items (Grazia & Molinari, 2021; Wang & Degol, 2016; Wang et al., 2020).

Climate Scale Interpretation

A frequent lack of cohesive operationalization of climate extends into the measurement tools and designs used to evaluate school and classroom climates (Marsh et al., 2012; Wang et al., 2020). "Particularly in educational climate and contextual studies… there is widespread confusion about the appropriate nature of data, design, statistical models, and interpretation. Fundamental design and analytic problems in many published studies seriously undermine substantive interpretations" (Marsh et al., 2012; p. 107) If a scale contains items that function differently for respondents with shared group membership, this can lead to biased results. For instance, an item that is easier to endorse for respondents from one type of school may unfairly advantage them. In a more recent systematic literature review, Grazia and Molinari (2021) conclude that what we know from previous reviews on school climate instruments suggest that although there are numerous existing validated measures, researchers continue to adapt and create new measures, when this should be avoided from a measurement perspective, if possible, to create more consensus.

When making claims across school sectors using scale scores, it is assumed the scales

have consistent meaning across the groups. This is the idea of measurement invariance; that items within a scale are consistently ordered from least- to most- difficult to endorse no matter the respondent. For this reason, a measure must be validated for populations that include the sectors of interest, and the scale must function consistently for each of the sectors. When items within a scale consistently function differently for one group than another, validity of inferences based on the scores derived from the scale is threatened (Kane, 2006; Messick, 1989, 1988). This is considered "differential item functioning", also known as DIF (Dorans & Holland, 1993; Holland & Thayer, 1988; Holland & Wainer, 1993).

Differences Between School Type

One of the first statewide snapshots of school climate took place in Illinois in 2013 as the "*5Essentials*" framework developed by the University of Chicago Consortium on School Research was measured across all public schools, including charter schools (Klugman et al., 2015). Historically, climate studies have been situated in public schools, using aggregated perspectives of students (Mitchell et al., 2010; Thapa et al., 2013). Climate studies that consider different types of schools are less prominent but do appear in the literature. Existing work to date has attempted to measure aspects of school or classroom climate in different school types, but none have measured climate on a common survey across multiple sectors nor unpacked whether measurement differences exist across these sectors.

There may be no differentiation between school types, like in the case of Ryberg and colleagues (2020), with students combined from public and charter schools. Or differences between school types may be prominently discussed, although frequently in pairs of types (i.e. private or public, public or charter, etc.). Broadly, teachers in private schools have reported a stronger sense of community than those in public schools (Bryk et al., 1993; Royal et al., 1996).

Catholic school administrators had more positive climate perceptions than evangelical Protestant schools (Sikkink, 2012). Gerlinger & Wo (2016) suggested private schools might be able to provide more socioemotional resources to students and found less bullying than public school students after controlling for a variety of other factors. Deangelis and Lueken (2019) concluded private and charter schools in Indiana reported fewer discipline problems while using fewer disciplinary practices than traditional public schools.

Several studies have indicated private schools have generally a more positive school climate than public schools (e.g., DeAngelis & Lueken, 2019; Farina, 2019; Gerlinger & Wo, 2016; Krommendyk, 2007). Teacher turnover rates notably differ between public, private, and charter schools (Guthery & Bailes, 2019; Ingersoll & Tran, 2023; Newton et al., 2018). Charter school teachers are frequently younger, have less experience, and are more likely to be teachers of color (Dallavis & Berends, 2023). Wei et al. (2014) found charter school teachers perceived more supportive teaching environments with higher expectations of students and stronger student engagement compared with traditional public school teachers. Duszka (2018) found traditional public and magnet schools had more positive ratings from students than charter schools, after controlling for other factors. However, public school staff reported the most negative perceptions, and more notably, there was no significant effect of school type in a panel regression model.

Differential Item Functioning

Differential Item Functioning (DIF) refers to the phenomenon where different groups of respondents from different sectors (e.g., public vs. private schools, or urban vs. rural schools) respond differently to specific test items, even when their ability levels are the same. Validating the instrument across different groups ensures that the scale measures the same construct in the

same way for all respondents, regardless of their background or environment. This makes comparisons between schools more meaningful and reliable. To make comparisons across contexts, it is critical to have a shared understanding of what a construct is and what it isn't (Sampson et al., 2021). In the absence of DIF, differences in scores between schools is more likely to reflect true differences in school climate, not discrepancies in how the scale functions for each group.

In addition to a common operationalization, to be able to make effective decisions, instruments must be consistent, accurate, and fair so users can trust the resulting evidence as credible (Schweig et al., 2019). Addressing DIF helps ensure that no group is advantaged or disadvantaged because they interpret specific items within the scale differently. However, items with DIF might still provide valuable information if the differences are minor and do not meaningfully impact the interpretation of scores or the validity of comparisons. If DIF doesn't significantly affect the overall model fit or interpretability of the measures, it may simply reveal interesting insights about how the construct functions for different groups.

As outcomes from climate measures are commonly linked to high-stakes outcomes through teacher and school level evaluation systems, measures must have evidence of validity within the contexts they are being used. To be able to make claims across school sectors, a measure must be validated for multiple populations. Historically, climate studies have been situated in public schools (Thapa et al., 2013). It is important to understand if a climate measure can make valid comparisons between different school types, as these comparisons are frequently made in accountability systems. Moreover, as states are focusing valuable resources into options of school choice, comparisons between different sectors requires calibrated and appropriate instruments that can provide more information about school sector effectiveness and outcomes.

Methods

The teachers participating in this study came from a broader study of school effectiveness in Indiana. As one of the primary objectives was to explore the organizations differences between types of schools, the research team stratified schools by school type (e.g., traditional public, charter, private), enrollment (total number of students), location (the geographical indicator describing the type of area where the school is located – urban, suburban, town, or rural, as defined by the National Center for Education Statistics (n.d)), and grade level to create a representative sample. In total, 548 schools agreed to participate: 335 traditional public or magnet, 24 charter, 112 Catholic, and 77 other private (nearly all faith-based). After schools agreed to participate in the study, the research team sampled a maximum of 10 teachers from each school.

The research team prioritized teachers employed full-time in grades 3-8 who specialized in either mathematics or English/language arts, as these are the two subjects assessed annually by the Indiana Department of Education, followed by grades K-2. The survey was sent electronically to a school email address and took an average of 44 minutes to complete; a $25 Amazon gift code was provided for teachers who completed the entire survey. In recruiting teachers for the study, the research team secured their informed consent and reminder emails were sent to teachers who had not completed the survey. All participants' direct identifying information was removed prior to final data cleaning and delivery to preserve confidentiality. In total, 5,399 teachers received the survey, of which 5,031 completed the full survey, 246 refused, 7 were ineligible, and 65 were unavailable for a total teacher response rate of 93.7%. Of the 5,031, 4,974 teachers had complete responses. Similar to national statistics, the overall sample of teachers was skewed female (84.7%) and White (90.7%). The average years spent in

their current school for teachers was 8 (SD = 7.9); 57% of teachers had been in their current school for 5 or fewer years.

A driving force behind survey measures chosen were their predictive validity and relationships to student achievement, specifically the organizational conditions that enable [and promote] student achievement. Stemming from the 5Essentials framework, the School Climate and Culture Survey (Bryk et al., 2010) has been validated in a variety of contexts (elementary, middle, and high schools) and has been used in more than 6,000 schools across the nation (for a more comprehensive report, see Hart et al., 2020; Gilman et al., 2021; Porter et al., 2023). Efforts have been made validate the items across different school contexts, but these are geared more toward student demographic characteristics, and primarily only in public schools (Hart et al., 2020, Porter et al., 2023). The measure consisted of 23 Likert-scale items with four possible options to endorse (not a challenge, a slight challenge, a moderate challenge, a great challenge). The stem read: "To what extent do you consider each of the following factors *a challenge* to student learning in your classroom?" with 23 individual challenges and was in the first section of the larger survey. Alpha reliability for the scale is 0.88. (see Table 1 for descriptions of items).

In previous work using this scale and sample assessing differences across sectors using ANOVA, Roberts (2019) found significant differences between school types between almost all items. Trends indicated a significant difference in average response between traditional public and Catholic, other private and charter, and charter with Catholic and other private, but not between Catholic and other private schools. For other items, there was a significant difference in average response between traditional public, Catholic and other private, and a difference between charter with Catholic and other private, but no difference between traditional public with charter or Catholic with other private schools. For a small group of items, there was a significant

difference in average response between traditional public with Catholic and other private

schools, but not between the other combinations. Additional work has been done with this scale

and found items functioned differently between elementary and middle school teachers (Roberts,

2019), however that work did not consider school type.

Data Analysis

The Rasch model (1960/1980) is a framework with which to examine the properties of an

instrument or other psychometric measure. An instrument should measure one latent trait, which

exists as a continuum, based on how a set of persons respond to items related to the trait, and the

difficulty to endorse each of the items. A person's response to an item is independent of their

response on any other item, conditional on said person's ability. The Rating Scale Model (RSM;

Andrich, 1978) is used for polytomous instrument items, modeled as Wright & Masters (1982):

$$P_{nik} = \frac{e^{(Bn-D_i-F_k)}}{1-e^{(Bn-D_i-F_k)}}$$

where P_{nik} is the probability that a person n at a given threshold of k will respond to an item i.

Resulting person and item measures are expressed in log odds units or logits. Item difficulty

refers to the level of difficulty a particular item is to endorse, with higher scores indicating a

harder item. An array of options will provide a more useful measure of the latent trait. Person

location refers to the level of the latent trait the individual possesses. Infit and mean square

values estimate how well items individually fit model predictions with an acceptable range in

logits from -2.0 to 2.0 (Linacre, 2020). Differential item functioning (DIF) is used to identify

bias and degree to which items on a scale might be biased based on group membership (Zumbo,

2007). If the item location along the continuum varies across groups after matching, the latent

trait is considered differently at a systematic level, making subsequent comparisons between

groups problematic and potentially inappropriate (Cauffman & MacIntosh, 2006).

Results

The Rasch RSM analysis produced a person reliability estimate of 0.86 and an item reliability estimate of 1.00. Item difficulty level estimates indicated an acceptable range of difficulty for participants to endorse (-1.05 to 1.60). Infit values ranged from 0.72 to 1.53; outfit values ranged from 0.72 to 1.57. Each of these values are considered within acceptable bounds (Linacre, 2020, p. 361). Principal Components Analysis (PCAR) confirmed the scale was unidimensional. Prior to tests of differential item functioning (DIF), the average measure and score were calculated for each of the four groups to understand overall differences in responses. Groups with lower values find their climate more of a challenge to student learning, and therefore have more negative perceptions of classroom climate.

The total mean measure was -.65, for Catholic school teachers -.92, for other private school teachers -1.01, for charter teachers -.41, and for traditional public teachers -.59. These values suggest teachers' perceptions of climate are somewhat similar for Catholic and other private teachers as a group and are mostly positive. Charter and public school teachers have favorable ratings of their classroom climate, but they are lower than those in private school settings. An ANOVA test confirms these significant differences overall ($F = 74.23$, $p < 0.001$) and between each set of school pairings ($p < .01$) with the exception of Catholic and other private schools ($p = .65$). This finding is important as even if no DIF occurs between the teachers in the different schools, significant and important differences do exist in the teachers' perceptions of school climate.

DIF Analysis

To determine if differential item functioning occurred (DIF), both global and pairwise DIF were conducted. The global DIF analysis examines how a focal group compares to the entire

sample per item, and the pairwise analysis compares a specific sub-group with another subgroup. Following analyses, 21 of 23 items exhibited a significant DIF at the .05 level. Six items displayed a slight to moderate absolute DIF for one or more school type, and three items a moderate to large absolute DIF for one or more school type (Zwick et al., 1999) (see Table 2). The remaining twelve items with a significant DIF fell outside the established criteria and are considered to have negligible DIF.

Patterns emerge in the DIF effect sizes between the school type pairings. The item *safety* exhibits moderate to large DIF between Traditional Public Schools (TPS) and Catholic schools, as well as slight to moderate DIF between TPS and Other Private, Charter and Catholic, and Charter and Other Private schools. In each case, the item is biased against the focal group. Item *highratio* showed DIF between TPS and each of other school types, with items being biased against TPS. Item *supportstaff* showed a bias against the reference group with moderate to large DIF between TPS and Catholic, TPS and Other Private, Charter and Catholic, and Charter and Other Private schools.

There are several items that have slight DIF between charter schools and other private schools, favoring other private schools, but no other school types. These items are the items *preptime*, *differentability*, *teacherplanning*, and *studenttardy*. Items *parentinterest* and *supportsped* showed DIF between TPS and Catholic and TPS and Other Private schools. For *parentinterest*, there is bias against TPS; for *supportsped*, the bias is toward non-TPS. The items *techaccess* and *facilities* exhibited DIF between TPS and each of the other school types, with items favoring the non-TPS schools.

Two items had DIF effects over 1.0, *testpressure* and *teacherturnover*. For the *testpressure* item, there were significant differences in DIF effect size for TPS and Charter

schools, TPS and Other Private schools, Charter and Other Private Schools, and Catholic and Other Private schools. This is the only instance of significant DIF between Catholic and Other Private Schools. For *teacherturnover* each DIF effect is greater than an absolute value of 1, indicating a substantially large bias between school types, reflective of clear differences in how teacher turnover impacts teachers' perceptions of classroom climate and should be possibly eliminated from future iterations.

Discussion

The analyses above confirm this previous validated school climate instrument can effectively be used to measure teachers' perceptions and is able to distinguish between perceptions of more positive and more negative climates. This measure has been extensively used over time, in both larger and smaller scale applications, so this finding is not surprising. However, the thresholds regarding the probability of choosing a response indicating less or more of an item being a challenge suggests that teachers were more likely to express items as being not or a slight challenge more often than a moderate or great challenge. Although there are no misfitting items, there is a great deal of significant slight, moderate, and large differential item functioning between the school types in a portion of the items. For those that do have a moderate to large magnitude, differences are more concentrated between TPS and other school types, and Charter and other school types, which might be a cause for concern for researchers and policymakers attempting to make straightforward statistical comparisons between different school organizational structures.

From the reviews of climate literature, it is unlikely that practitioners or researchers are conducting an item level validation each time data is analyzed, much less an in-depth examination of possible DIF. Based on the results of Roberts (2019) and the current study, there

is a wide variation in how teachers in different grades and in different school environments perceive different facets of school climate. This finding supports prior research (Krommendyk, 2007; Lubienski et al., 2008) that climate differs between school types, and goes beyond by finding that not only do climate perceptions differ by teachers in different school sectors, but that that teachers are interpreting the items differently according to the sector they teach in. Many of the items that constitute school climate at the institutional and student level are often out of the control of individual teachers, a product of federal, state, or district funding and policies, but undoubtedly impact a teachers' perceptions of their environment. For instance, the item *"To what extent do you consider each of the following factors a challenge to student learning in your classroom? 'pressure to perform well on standardized tests'"* could be perceived differently for teachers in private school settings over those in public or charter schools because the pressures for producing high achievement scores may not exist in the same way. The state of Indiana is unique in that many schools still participate in state standardized testing as to receive voucher students.

The item that was recommended for removal for significant and persistent DIF dealt with teacher turnover, which is often out of the control of individual teachers. Yet, turnover does impact a teacher's perceived and actual working and teaching environment. Turnover rates continue to be a recognized and persistent problem in education across school sectors (Ingersoll & Tran, 2023). Charter schools are more likely to have higher rates of turnover than public schools; however, private schools have been seen as having less turnover in schools, suggesting a larger influence of sector organizational characteristics on school climate (Guthery & Bailes, 2022; Stuit & Smith, 2012; Taie & Lewis, 2023). With 86% of U.S. K-12 public schools reporting challenges in hiring teachers for the 2023-2024 school year, there is significant

fluctuation in the personnel seen by students and teachers regularly in a school contributing to constantly changing school climate (National Center for Education Statistics [NCES], 2023).

There are some limitations to this study. First, while both principals and teachers were surveyed in the larger study, only teachers were given the climate instrument, and asked to speak on their perceptions of challenges in their individual classrooms. This measure does not include the perspectives of administrators, students, or parents, which could provide a more comprehensive picture of an overall school climate (Ramsey et al., 2016). However, research has suggested students and teachers are similar in their perceptions of school climate within the same school (Higgins-D'Alessandro & Guo, 2009). Additionally, the survey only measures teachers' perceptions of climate at one point in time, and climate can be possibly unstable for many teachers from year to year. A longitudinal or repeated measures study could provide information on how climate perceptions may change over time, and in response to policy changes. The sample taken is reflective of the larger study sample, the proportion of school sector and the general teacher population in the state and is relatively homogenous both in gender and race/ethnicity. Underrepresented teachers in the sample may provide valuable evidence hidden by more robustly represented groups.

Conclusion

As a DIF analysis aims to determine if participants respond differently to the same item based on a particular grouping, results do indicate that teachers in different school sectors may conceptualize statements regarding climate in their schools affecting their classrooms systematically differently. One suggested general cause of DIF is multidimensionality in items measuring some concept outside of the primary latent variable (Roussos & Stout, 1996). However, significant DIF may also be explained from the general complexity of items that may

represent one or more latent variables simultaneously (McDonald, 2000). After all, school climate is a concept researchers have argued as both unidimensional and multidimensional (Thapa et al., 2013; Wang et al. 2020).

Teresi (2006) addressed how "relative bias" may occur if individuals rate themselves relative to others in the same environment. For instance, a teacher may perceive their classroom to be a difficult environment, but perhaps not as difficult as a previous year; or compare themselves with a colleague, over- or undercompensating their own ratings. A teacher may be basing decisions on one classroom period with one group of students, or an average of the classroom periods instructed with different students. Each of these realistic scenarios might not be measured with one static instrument, which further demonstrates the need for continuous validation of climate measures, and the potential difficulties comparing results even from year to year across groups, across school sector, or between schools that differ on a variety of characteristics.

The significant presence of DIF with this instrument should not be interpreted as a negative bias against any group of teachers. Rather, it is more likely reflective of the natural differences in learning environments that differ from school to school. DIF may occur as the result of one group having less experience or background knowledge related to the item or items. Using a somewhat novel approach with a DIF analysis allows for a thoughtful exercise in understanding how different school sectors' variation in organizational structures is reflected statistically in teachers' perceptions of school climate. In essence, this established climate measure (and likely others) is/are not a one-size-fits all instrument, and caution should be used when interpreting research that compares across school sectors without exploring its measurement validity in each context. As climate measures continue to be used in federal and

state accountability systems throughout the United States, stakeholders at all levels should participate in ongoing evaluation of the effectiveness of their measures when attempting to make comparisons between different types of schools. It is the recommendation of this research that climate comparisons should not be made within one school sector for climate measures without first demonstrating the measure that can accurately differentiate between groups. Additional research, including cognitive interviews and other qualitative or mixed methods designs should be conducted to determine if this persists in other climate measures or other contexts.

References

Aldridge, J.M., Fraser, B.J. (2016). Teachers' views of their school climate and its relationship with teacher self-efficacy and job satisfaction. *Learning Environments Research, 19*, 291–307. https://doi.org/10.1007/s10984-015-9198-x

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561- 573.

Berends, M. & Waddington, R. J. (2018). School choice in Indianapolis: Effects of charter, magnet, private, and traditional public schools. *Education Finance and Policy*, *13*(2), 227-255. https://doi.org/10.1162/edfp_a_00225

Berkowitz, R., Moore, H., Astor, R. A., & Benbenishty, R. (2019). A research synthesis of the associations between socioeconomic background, inequality, school climate, and academic achievement. *Review of Educational Research*, *87*(2), 425-469. https://doi.org/10.3102/0034654316669821

Bryk, A. S., Lee, V. E., & Holland, P. B. (1993). *Catholic schools and the common good*. Harvard University Press.

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. University of Chicago Press.

Cauffman, E., & MacIntosh, R. (2006). A Rasch differential item functioning analysis of the Massachusetts Youth Screening Instrument. *Educational & Psychological Measurement*, *66*(3), 502-521. https://doi.org/10.1177/0013164405282460

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. https://doi.org/10.1177/016146810911100108

Dallavis, J. W., & Berends, M. (2023). Charter schools after three decades: Reviewing the research on school organizational and instructional conditions. *Education Policy Analysis Archives*, *31*(1). https://doi.org/10.14507/epaa.31.7634

DeAngelis, C. A. & Lueken, M. F. (2019). School sector and climate: An analysis of K–12 safety policies and school climates in Indiana. *Social Science Quarterly*, *101*(1), 376-405. https://doi.org/10.1111/ssqu.12737

Duszka, C. D. (2018). *School climate in the school choice era: A comparative analysis of district-run public schools and charter schools*. [Doctoral Dissertation, Florida International University]. FIU Electronic Theses and Dissertations, 3922. https://doi.org/10.25148/etd.FIDC006985

Farina, K. A. (2019). Promoting a culture of bullying: Understanding the role of school climate and school sector. *Journal of School Choice*, *13*(1), 94-120. https://doi.org/10.1080/15582159.2018.1526615

Gerlinger, J., & Wo, J. C. (2016). Preventing school bullying: should schools prioritize an authoritative school discipline approach over security measures? *Journal of School Violence*, *15*(2), 133–57. https://doi.org/10.1080/15388220.2014.956321

Gilman, L.J., Zhang, B. & Jones, C.J. (2021). Measuring perceptions of school learning environments for younger learners. *Learning Environments Research, 24*, 169–181. https://doi.org/10.1007/s10984-020-09333-y

Grazia, V. & Molinari, L. (2021) School climate multidimensionality and measurement: A systematic literature review. *Research Papers in Education*, *36*(5), 561-587, https://doi.org/10.1080/02671522.2019.1697735

Guthery, S., & Bailes, L. P. (2022). Patterns of teacher attrition by preparation pathway and

initial school type. *Educational Policy*, *36*(2), 223-246.

https://doi.org/10.1177/0895904819874754.

Hart, H., Young, C., Chen, A., Zou, A., & Allensworth, E.M. (2020). *Supporting school improvement: Early findings from reexamination of the 5Essentials Survey*. University of Chicago Consortium on School Research. https://files.eric.ed.gov/fulltext/ED608120.pdf

Ingersoll, R. M. & Tran, H. (2023). Teacher shortages and turnover in rural schools in the US: An organizational analysis. *Educational Administration Quarterly*, *59*(2), 396-431. https://doi.org/10.1177/0013161X231159922

Klugman, J., Gordon, M. F., Sebring, P. B., & Sporte, S. E. (2015). *A first look at the 5Essentials in Illinois schools*. UChicago Consortium on School Research. https://consortium.uchicago.edu/publications/first-look-5essentials-illinois-schools

Krommendyk, M. (2007). *The association between school choice and school climate: Comparing school climate in private, religious, charter, and public schools*. [Doctoral Dissertation, Western Michigan University]. http://scholarworks.wmich.edu/dissertations/885

Linacre, J. M. (2020). Winsteps® Rasch measurement computer program User's Guide. Retrieved from https://www.winsteps.com/manuals.htm

Lubienski, S. T., Lubienski, C., & Crane, C. C. (2008). Achievement differences and school type: The role of school climate, teacher certification, and instruction. *American Journal of Education*, *115*, 97–138. https://doi.org/10.1086/590677

Marsh, H.W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B. O., Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error.

Multivariate Behavioral Research, *44*, 764-802.

https://doi.org/10.1080/00273170903333665

McDonald, R. P. (2000). A basis for multidimensional item response theory. *Applied*

Psychological Measurement, *24*, 99-114. https://doi.org/10.1177/01466210022031552

Mitchell, M. M., Bradshaw, C. P., & Leaf, P. J. (2010). Student and teacher perceptions of

school climate: A multilevel exploration of patterns of discrepancy. *The Journal of*

School Health, 80, 271–279. http://dx.doi.org/10.1111/j.1746-1561.2010.00501.x

National Center for Education Statistics. (n.d.). *Locale classifications*.

https://nces.ed.gov/programs/edge/Geographic/LocaleBoundaries#:~:text=A%20locale%

20classification%20is%20a,or%20proximity%20to%20populated%20areas.

National Center for Education Statistics. (2023, October 17). *Most public schools face challenges*

in hiring teachers and other personnel entering the 2023-24 academic year [Press

release]. https://nces.ed.gov/whatsnew/press_releases/10_17_2023.asp

National School Climate Council (2007). *The school climate challenge: Narrowing the gap*

between school climate research and school climate policy, practice guidelines and

teacher education policy. Retrieved from https://schoolclimate.org/wp-

content/uploads/2021/05/school-climate-challenge-web.pdf

Newton, X., Rivero, R., Fuller, B., & Dauter, L. (2018). Teacher turnover in organizational

context: Staffing stability in Los Angeles charter, magnet, and regular public schools.

Teachers College Record, *120*(2018), 1-36.

https://doi.org/10.1177/016146811812000306

Porter, S.C., Jackson, C.K., Kiguel, S., & Easton, J.Q. (2023). Investing in adolescents: High

school climate and organizational context shape student development and educational

attainment. University of Chicago Consortium on School Research.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests* (Reprint, with Foreword and Afterword by B. D. Wright, University of Chicago Press, 1980). Danmarks Paedogogiske Institut.

Roberts, A. C. (2019). *Validation of a school climate instrument using a Rasch rating scale model.* [Doctoral dissertation, University of Kentucky]. Theses and Dissertations--Education Sciences, 49. https://doi.org/10.13023/etd.2019.088

Roussos, L. & Stout, W. (1996). A multidimensionality-based DIF analysis paradigm. *Applied Psychological Measurement, 20*(4), 355-371. https://doi.org/0146-6216/96/040355-17$2.10

Royal, M., D'Angelis, K., & Rossi, R. (1996). *Teachers' sense of community: How do public and private schools compare?* Office of Educational Research and Improvement, National Center for Educational Statistics. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=97910

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

Sampson, S. O., Bradley, K., & Nelson, A. (2021). Creating a structurally sound instrument by utilizing principles of measurement theory. In U. Luhanga & A. G. Harbaugh (Eds.), *Basic elements of survey research in education: Addressing the problems your advisor never told you about* (Chapter 4). Information Age Publishing.

Sanchez, J. E., Paul, J. M., & Thornton, B. W. (2022). Relationships among teachers' perceptions of principal leadership and teachers' perceptions of school climate in the high school setting, *International Journal of Leadership in Education*, *25*(6), 855-875. https://doi.org/10.1080/13603124.2019.1708471

Schweig, J., Hamilton, L. S., & Baker, G. (2019). *School and classroom climate measures: Considerations for use by state and local education leaders.* RAND Corporation. https://www.rand.org/pubs/research_reports/RR4259.html

Sikkink, D. (2012). Religious school differences in school climate and academic mission: A descriptive overview of school organization and student outcomes. *Journal of School Choice, 6*, 20-39. https://doi.org/10.1080/15582159.2012.651394

Stuit, D. A. & Smith, T. M. (2012). Explaining the gap in charter and traditional public school teacher turnover rates. *Economics of Education Review*, *31*(2), 268-279. https://doi.org/10.1016/j.econedurev.2011.09.007

Taie, S., and Lewis, L. (2023). *Teacher attrition and mobility. Results from the 2021–22 Teacher Follow-up Survey to the National Teacher and Principal Survey (NCES 2024-039)*. National Center for Education Statistics at the Institute of Education Sciences. https://nces.ed.gov/pubs2024/2024039M.pdf

Teresi, J. A. (2006). Different approaches to differential item functioning in health applications: Advantages, disadvantages and some neglected topics. *Medical Care*, *44*(11), S152-S170. http://www.jstor.org/stable/41219515

Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research, 83*, 357–385. https://doi.org/10.3102/0034654313483907

Wang, M. T., & Degol, J. L. (2016). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review, 28*(2), 315–352. https://doi.org/10.1007/s10648-015-9319-1

Wang, M. T., Degol, J. L., Amemiya, J., Parr, A., & Guo, J. (2020). Classroom climate and children's academic and psychological wellbeing: A systematic review and meta-analysis. *Developmental Review, 57*(2020), 100912. https://doi.org/10.1016/j.dr.2020.100912

Wei, X., Patel, D., & Young, V. (2014). Opening the "black box": Organizational differences between charter schools and traditional public schools. *Education Policy Analysis Archives*, *22*(3), 1-31. https://doi.org/10.14507/epaa.v22n3.2014

Wright, B. D. & Masters, G. N. (1982). *Rating scale analysis.* MESA Press.

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*, 49-69. https://doi.org/10.1016/j.jsp.2020.08.002

Zakariya, Y.F. (2020). Effects of school climate and teacher self-efficacy on job satisfaction of mostly STEM teachers: A structural multigroup invariance approach. *International Journal of STEM Education, 7*(10). https://doi.org/10.1186/s40594-020-00209-4

Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, *4(*2), 223–233. https://doi.org/10.1080/15434300701375832

Zwick, R., Thayer, D. T., & Lewis, C. (1999). An empirical Bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement, 36*, 1-28. https://doi.org/10.1111/j.1745-3984.1999.tb00543.x

Zysberg, L. & Schwabsky, N. (2021). School climate, academic self-efficacy and student achievement. *Educational Psychology, 41*(4), 467-482. https://doi.org/10.1080/01443410.2020.1813690

Table 1

Classroom Climate Survey Item Descriptions

Item Name	To what extent do you consider each of the following factors a challenge to student learning in your classroom?
teachmorale	Low morale among fellow teachers/administrators
background	Students who come from a wide range of backgrounds
safety	Threat(s) to your safety or safety of students
noise	The noise level in the school building
supportstaff	Amount of professional support staff
specneeds	Students with special needs (e.g., hearing, vision, speech impairment, physical disabilities, mental or emotional/psychological impairment)
preptime	Amount of time to prepare for class
highratio	High student/teacher ratio
diffability	Students with different academic abilities
uninterest	Uninterested students
disrupt	Disruptive students
parentint	Parents uninterested in their children's learning progress
techaccess	Access to technology
testpressure	Pressure to perform well on standardized tests
rsshelp	Lack of school resources to provide the extra help for students who need it
planning	Lack of teacher planning time built into the school day
facilities	Inadequate physical facilities
studmorale	Low morale among students
turnover	Teacher turnover in this school
absent	Student absenteeism
tardy	Student tardiness
supportsped	Lack of guidance or support for teaching special education students (i.e., students with IEPs)
supportell	Lack of guidance or support for teaching English Language Learners

Table 2

Differential Item Functioning Effect Size

Item	Contrasting Pair		DIF Effect
	Focus	Reference	
safety	TPS	CA	-.65
safety	TPS	OP	-.51
safety	CH	CA	-.55
safety	CH	OP	-.41
supportstaff	TPS	CA	.66
supportstaff	TPS	OP	.93
supportstaff	CH	CA	.48
supportstaff	CH	OP	.74
preptime	CH	OP	.47
highratio	TPS	CH	-.63
highratio	TPS	CA	-.61
highratio	TPS	OP	-.82
diffability	CH	OP	.45
parentint	TPS	CA	-.74
parentint	TPS	OP	-.50
techaccess	TPS	CH	.91
techaccess	TPS	CA	.76
techaccess	TPS	OP	.84
testpressure	TPS	CH	-.42
testpressure	TPS	OP	-1.05
testpressure	CH	OP	-.63
testpressure	CA	OP	-.83
noplan	CH	OP	.49
facilities	TPS	CH	.62
facilities	TPS	CA	.52
facilities	TPS	OP	.93
turnover	TPS	CH	1.40
turnover	CH	CA	-1.29
turnover	CH	OP	-1.40
tardy	CH	OP	-.49
supportsped	TPS	CA	.42
supportsped	TPS	OP	.74

Note: TPS = traditional public school; CH = charter school; CA = Catholic school; OP = other private school. Negative values indicate a bias against the focal group and positive values indicate a bias against the reference group.