



School Climate Health: Multidimensional Profiles of School Environments in TIMSS 2023

*Charlotte E.A. Aldrich, Katherine A. Reynolds, Dihao Leng,
& Matthias von Davier*



Copyright © 2025

International Association for the Evaluation of Educational Achievement (IEA)

School climate health: Multidimensional profiles of school environments in TIMSS 2023

Charlotte E.A. Aldrich, Katherine A. Reynolds, Dihao Leng, Matthias von Davier

Suggested citation:

Aldrich, C. E. A., Reynolds, K. A., Leng, D., & von Davier, M. (2025). *School climate health: Multidimensional profiles of school environments in TIMSS 2023*. Boston College, TIMSS & PIRLS International Study Center.

<https://doi.org/10.6017/lse.tpisc.kd3138>

For more information about PIRLS contact:

TIMSS & PIRLS International Study Center

Lynch School of Education and Human Development

Boston College

Chestnut Hill, MA 02467

United States

tel: +1-617-552-1600

e-mail: timssandpirls@bc.edu

timssandpirls.bc.edu

Boston College is an equal opportunity, affirmative action employer.

School Climate Health: Multidimensional Profiles of School Environments in TIMSS 2023

Introduction

Students' school experiences, their academic achievement, and social-emotional health can be affected by the complex social system of their school's environment and the roles, relationships, and influence of relevant stakeholders. While greater attention is being given to measures of students' emotional development in schools, the dynamics of the school's social systems are not comprehensively covered by existing student-level measures.¹ In addition, there is room for improvement regarding how to meaningfully aggregate intrapersonal attributes reported at the individual level to systems-level information and how this might inform evaluations of school quality.

This TIMSS Insights report describes a new framework – *school climate health* – for defining similarities of school social systems internationally, operationalized through the items in the TIMSS 2023 context questionnaires. School climate health integrates reports from different groups of stakeholders and posits that school climate and school well-being are dynamic constructs embedded in social systems shaped by the values and roles of those stakeholders. The contextual data available through TIMSS 2023 can be used to derive patterns of well-being and school climate, which allow exploring systems-level questions about the quality of national education systems. Data from the TIMSS 2023 Context Questionnaires provide collections of important indicators believed to be beneficial to school environments as agreed upon by the countries participating in the study. This report builds upon explorations of different facets of well-being in other IEA studies, including the first [PIRLS 2021 Insights report](#), which examines indicators of this construct at a student level.

The report begins with an overview of the concept of *school climate health* and how aspects of this construct are represented in the TIMSS 2023 Context Questionnaires. The report then presents the results of a multilevel latent class analysis (MLCA) that explores patterns among indicators from the context questionnaires to define clusters representing different profiles of school climate health in Grade 4. Clusters are further investigated by relating them to average socioeconomic status within schools. Averages of school-level mathematics and science achievement by cluster are also presented for each country. The report concludes with key takeaways and suggestions for future research related to school climate health in TIMSS.

Conceptualizing School Climate Health

A school's environment reflects a combination of interrelated, multilevel factors that influence the actions of those involved, rather than simply its physical or administrative structures. School climate health addresses this complexity by combining relevant scales related to perceptions of the school environment using methodologies that can capture complex structures of interdependencies. Thus, school climate health is construed as the **quality of the relationships among school participants, the environment that those relationships create, and the satisfaction with the schooling experience of all those involved**. The resulting profiles of school climate health simultaneously capture the state of key environmental factors supportive of student well-being and learning such as student relationships, teacher performance, and school safety.

Existing Literature

The definition of school climate health resulted from an extensive literature review of existing frameworks of well-being (e.g. subjective,^{2,3} psychological,⁴ and positive psychology⁵) and school climate,^{6,7} with special attention to approaches aimed at multidimensional measurement and cross-cultural comparison.⁸ Some frameworks suggest the experiences of multiple stakeholders must be considered to create school environments that support student well-being.⁹ Literature on non-academic outcomes and large-scale educational evaluations both share a call to embrace systems thinking or complex systems theory as a lens to capture the complexity of education at large.^{10,11} Through this lens, non-academic outcomes like well-being¹² are understood as emerging from a system of social actors that work to either sustain or inhibit a positive school climate.¹³ It is essential to understand systems as a combination of interrelated parts that operate holistically within the boundaries of their context.^{14,15}

Every school can be considered as a social network that operates based on how the person-level mechanisms are supported—a complex social system knowable by the roles and mechanisms that are executed within it.¹⁶ A system in a state of well-being supports both individuals and the collective towards efficiently pursuing the shared goals of the system. Schools with healthy school climates could simultaneously support academic achievement and personal and social development.

Happy Schools Framework and Sustainable Development Goals

Emerging international initiatives, such as UNESCO's Happy Schools Framework, define frameworks for non-academic facets of schooling but have not been empirically tested. The Happy Schools project endeavors to implement sustainable school systems that support student academic success as well as well-being, grounded in a whole-school approach to “prioritize happiness and well-being” for “better quality” schooling.¹⁷ This three-part framework comprises several facets that will “promote happiness” by way of well-being and holistically supporting student development. The framework seeks to extend measures of quality within schools, beyond the more academically orientated objectives articulated in objectives from the United Nations Sustainable Development Goals (SDGs) [SDG 4.1 and SDG 4.7 (Exhibit 1.1)].

Exhibit 1.1: SDG Objectives and Indicators Enhanced by Happy Schools Framework

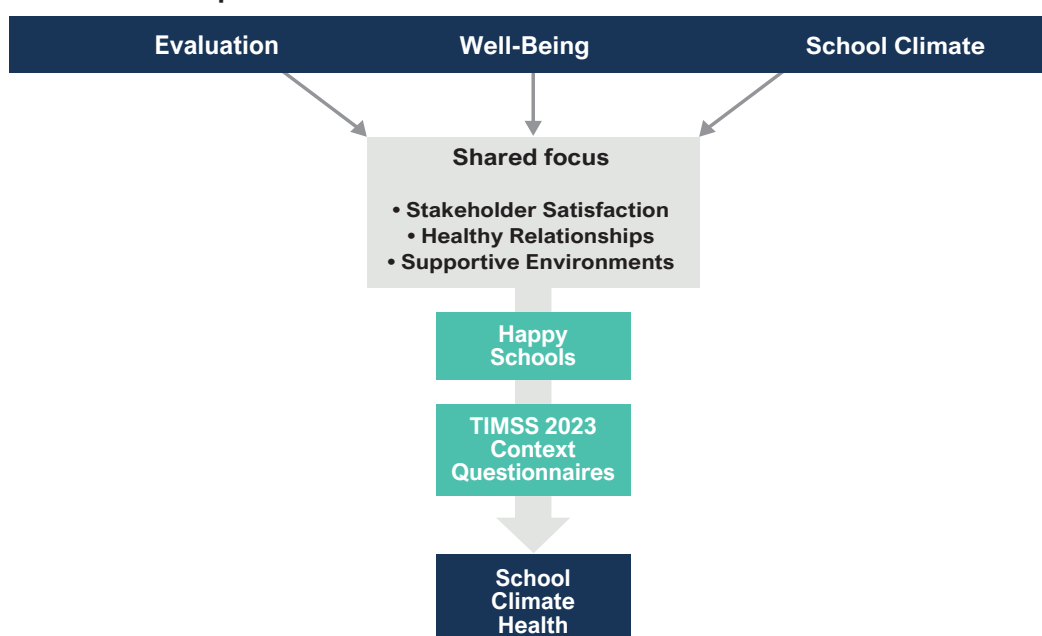
Objective	Overview	Indicator
4.1	By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes.	4.1.1 Proportion of children and young people (a) in Grade 2 or 3; (b) at the end of primary education; and (c) at the end of lower secondary education achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex
4.7	By 2030, ensure that all learners acquire the knowledge and skills needed to promote sustainable development, including, among others, through education for sustainable development and sustainable lifestyles, human rights, gender equality, promotion of a culture of peace and non-violence, global citizenship and appreciation of cultural diversity and of culture's contribution to sustainable development.	4.7.1 Extent to which (i) global citizenship education and (ii) education for sustainable development, including gender equality and human rights, are mainstreamed at all levels in: (a) national education policies; (b) curricula; (c) teacher education; and (d) student assessment

Source: 2030 Agenda for Sustainable Development (United Nations, 2015)

Operationalizing School Climate Health

School climate health exists at the intersection of major themes from literature on well-being and school climate, the Happy Schools Framework, and the constructs measured in the TIMSS 2023 Grade 4 Context Questionnaires. The TIMSS context questionnaires gather information on factors of educational environments that are thought to relate to mathematics and science achievement internationally. At the fourth grade, questionnaires are distributed to sampled students, their teacher(s), school principal, and parents or guardians.

Together, results from the literature review, criteria identified in the Happy Schools Framework, and content of the TIMSS 2023 context questionnaires come together to shape the framework of *school climate health* as depicted in Exhibit 1.2.

Exhibit 1.2: Components of the School Climate Health Framework

Criteria from the Happy Schools Framework were used to identify TIMSS 2023 context questionnaire items and scales that would capture school climate health. This process involved selecting items and scales that could act as indicators for the three areas of shared focus shown in Exhibit 1.2: stakeholder satisfaction, healthy relationships, and supportive environments. This process included reviewing the content of items and scales in relation to the three areas of shared focus that emerged in the literature and the definitions provided in the Happy Schools Framework. Fourteen of the 22 Happy Schools criteria are addressed in 22 context scales or item groups. Exhibit 1.3 shows the scales/item sets arranged thematically by area of shared focus, together with their mapped Happy Schools Framework criteria. Four variables were identified in the School Questionnaire, six from the Teacher Questionnaire, eleven from the Student Questionnaire, and one from the Home Questionnaire. TIMSS questionnaires are developed to measure attributes, attitudes, and descriptive information related to mathematics and science achievement. While a selection of items can be aligned with definitions of well-being or school climate, these items do not necessarily represent the constructs in their entirety.

Exhibit 1.3: Selected TIMSS 2023 Grade 4 Context Questionnaire Items and Scales with Mapped Happy Schools Framework Criteria

	Healthy Relationships 6 Scales/Item Sets (47 items)	Supportive Environments 8 Scales/Item Sets (76 items)	Stakeholder Satisfaction 8 Scales/Item Sets (56 items)
TIMSS 2023 Context Scale or Item Set	Student Indicators <i>Disruptive Lessons, Mathematics</i> <i>Disruptive Lessons, Science</i> <i>Instructional Clarity, Mathematics</i> <i>Instructional Clarity, Science</i> Principal Indicators <i>Principal's Report, Disorder</i> Teacher Indicators <i>Engaging Teacher Practices</i>	Student Indicators <i>Student Sense of Belonging</i> <i>Student Bullying</i> <i>Students Arrive Hungry/Tired</i> Principal Indicators <i>Principal's Report, Limited Resources</i> <i>Principal's Report, Academic Emphasis</i> Teacher Indicators <i>Teacher's Report, Academic Emphasis</i> <i>Teacher's Report, School Discipline</i> <i>Teacher's Report, Students Not Ready for Instruction</i>	Student Indicators <i>Students Like Mathematics</i> <i>Students Like Science</i> <i>Students Confident, Mathematics</i> <i>Students Confident, Science</i> Parent Indicators <i>Parents Satisfied with School</i> Principal Indicators <i>Principal's Report, Teacher Absenteeism</i> Teacher Indicators <i>Teacher Job Satisfaction</i> <i>Teacher's Report, Non-Instruction Responsibilities</i>
UNESCO Happy Schools Criterion	<i>Friendships and Relationships in the School Community</i> <i>Learning as a Team Between Students and Teachers</i> <i>Teamwork and Collaborative Spirit</i>	<i>Learner Freedom, Creativity, and Engagement</i> <i>Positive Collaborative Values and Practice</i> <i>School Vision and Leadership</i> <i>Secure Environment Free from Bullying</i> <i>Warm and Friendly Learning Environment</i>	<i>Democratic School Management</i> <i>Positive Teacher Attitudes and Attributes</i> <i>Sense of Achievement and Accomplishment</i> <i>Teacher Skills and Competences</i> <i>Teacher Working Conditions and Well-Being</i>

International Profiles in School Climate Health

A school climate health profile was established for each school with students participating in TIMSS 2023. The profile was based on the 22 indicator variables in Exhibit 1.3, which were aggregated to the school level for analyses using a multilevel latent class analysis (MLCA).¹⁸ Fifty-eight countries were included in the analysis.^{a,b} MLCA is an established procedure for exploring classroom- and school-level climate groupings based on self-reported questionnaire data.^{19,20} In the multilevel structure of the TIMSS 2023 data, Level 1 represents schools and Level 2 represents countries. Individual-level response data from teachers, students, and parents were aggregated to the school level for the analyses.^c

Multilevel Latent Class Analysis

MLCA allows for multidimensional profiles of schools and countries based on shared patterns of responses to the TIMSS context questionnaire scales and items relating to school climate health. At the school level (Level 1), variation in student, teacher, parent, and principal responses form distinct school climate health profiles. Such profiles illuminate how the different members of a school social system may report different experiences of school climate. At the country level (Level 2), countries are grouped based on similarities in the distribution of their schools across these profiles.

Latent class analysis and model selection requires different iterations of estimated models to be compared for best fit. While there are metrics for comparison of model fit, the process of modeling also involves judgments of utility. For instance, more complex models—i.e., models with more clusters—may have better model fit indices, but more parsimonious models could hold greater meaning in interpretation. This balance guided the modeling approach to support meaningful interpretation of the relationship between responses to discrete scales and the ultimate evaluation of school climate health. Estimation with group-level (country) classes follows a similar approach where countries are assigned to classes based on the extent to which a group of countries has similar proportions of each cluster, according to the schools assigned to the extracted clusters.

The following section focuses on qualitatively describing the clusters identified in the MCLA, along with the percentages of schools assigned to each cluster internationally. Detailed results concerning the country classes are not presented in this report; however, further details about how these may be interpreted are available elsewhere.²¹ Finally, cluster membership is related to socioeconomic status, as well as average achievement in mathematics and science, for Grade 4 students within each country. Further information about analysis procedures, statistical models, and relevant fit statistics is available in Appendix C: Technical Details.

a See the [TIMSS 2023 International Results](#) for additional information about country participation in TIMSS 2023.

b Four countries (Australia, England, the Netherlands, and the United States) did not administer the Home Questionnaire in TIMSS 2023 and do not have any data for *Parents Satisfied with School*.

c TIMSS randomly samples schools within countries and classrooms within schools to obtain a satisfactory representative sample of the population of students in each country. Aggregating to the school level allows for this analysis to incorporate data from all respondent types but does necessitate some caution. For instance, it is possible that only one classroom is sampled within a school such that the teacher is the only teacher voice for the school.

Characterizing the School Climate Health Clusters

Four interpretable clusters were identified based on the results of the MCLA. Brief descriptions of these clusters are presented below, along with the percentage of schools internationally assigned to each cluster based on the model results. These brief descriptions are followed by more detailed explanations of what distinguishes each individual cluster in terms of responses to the school climate health indicator variables.

There are several things to keep in mind when interpreting the cluster definitions presented in the following section. All 22 variables included in the analysis were recoded as ordinal variables with three categories: High (2), Medium (1), and Low (0). These categories should only be interpreted relative to each other, as they were defined to create relatively equal response regions per variable. They are not quantitative measures but reflect a relative ordering of the sets of responses. See Appendix C for further details about this procedure.

Additionally, cluster descriptions follow the analysis in that they are reflective of reports of *school-level* attributes and the basis for cluster definitions is based on variable means relative to the other clusters. Cluster definitions include references to having the highest or lowest mean school scores compared to the other clusters, and for some variables may compare the proportion of respondents in schools in those clusters that give “High” or “Low” endorsements.

Cluster 1: Experiencing Difficulties

Respondents in these schools tend to report (on average) challenges in maintaining healthy relationships between students and teachers (disruptive classrooms), lack of supportive environments (low academic emphasis and more frequent bullying), and low stakeholder satisfaction (low confidence in mathematics and science; low teacher and parent satisfaction).

30% of schools internationally were classified as *Experiencing Difficulties*.



Cluster 2: System Misalignment

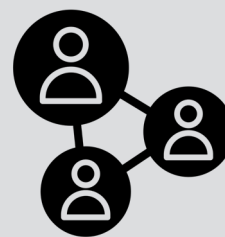
Different respondent types in these schools give different reports; while principal reports are largely positive, other respondents give mixed reports. On average, there are some healthy relationships based on reports of order within the school, but these are not evident in reports related to classroom practices. Principals and teachers tend to report that schools are well-resourced learning environments, yet students report lack of belonging. Positive reports of stakeholder satisfaction come only from principals, whereas students, teachers, and parents report on average low confidence and low satisfaction.

26% of schools internationally were classified as having *System Misalignment*.



Cluster 3: Healthy Social System

Respondents in these schools give largely positive reports. On average, reports indicate that healthy relationships exist between students and teachers, schools are sustaining high sense of student well-being with few limitations to resources and teaching, and stakeholders across respondent types (student, principal, parent, teacher) report high levels of satisfaction.



22% of schools internationally were classified as having a *Healthy Social System*.

Cluster 4: Generational Incongruence

The degree of positivity reported in these schools varies consistently between students and the adults involved in their education. Relationships in the school as reported by teachers and principals, on average, are positive based on teacher practices and the state of discipline; students report operating in classrooms that are disruptive with less clarity from their teachers. Reports on the environment in the school are similarly split: principals and teachers are largely positive while students feel much less belonging and experience bullying. Student reports are much less positive about their confidence or affinity for mathematics and science whereas all the adults report high satisfaction.



21% of schools internationally were classified as having *Generational Incongruence*.

Exhibits 2.1 and 2.2 show the percentages of schools in each recoded response category by cluster, internationally. For most indicators, the “High” category corresponds to the most “positive” outcome—for example, a high score on *Disruptive Lessons, Mathematics* corresponds to students reporting little disorder in the classroom. The exceptions to this pattern are *Principal’s Report, Teacher Absenteeism* and *Engaging Teaching Practices*, where the “Low” category corresponds to less frequent teacher absences or more frequent use of the engaging practices.

Exhibit 2.1: Percent of Schools in Recoded Response Categories by School Climate Health Cluster (Student Indicators)

	Cluster 1: Experiencing Difficulties	Cluster 2: System Misalignment	Cluster 3: Healthy Social System	Cluster 4: Generational Incongruence	Overall
Disruptive Lessons, Mathematics					
Low	68	8	9	28	30
Medium	27	32	33	42	33
High	5	60	57	30	37
Disruptive Lessons, Science					
Low	69	8	10	25	30
Medium	26	32	35	42	33
High	5	60	55	33	36
Instructional Clarity, Mathematics					
Low	50	57	1	18	34
Medium	38	35	13	43	33
High	12	8	86	39	33
Instructional Clarity, Science					
Low	48	62	1	17	35
Medium	39	31	16	41	32
High	13	7	83	41	33
Students Like Mathematics					
Low	38	64	1	29	35
Medium	42	30	11	43	32
High	20	6	88	28	33
Students Like Science					
Low	45	66	1	19	35
Medium	40	29	15	42	32
High	15	5	84	39	33
Students Confident, Mathematics					
Low	50	33	8	33	32
Medium	33	37	26	37	33
High	17	31	66	31	34
Students Confident, Science					
Low	55	33	7	27	32
Medium	32	38	26	38	33
High	13	30	68	35	34
Student Sense of Belonging					
Low	48	59	4	17	34
Medium	37	32	23	38	33
High	15	9	74	45	33
Student Bullying					
Low	56	8	24	30	31
Medium	32	28	37	37	33
High	12	64	39	32	36
Students Arrive Hungry/Tired					
Low	55	18	8	37	31
Medium	33	36	28	38	34
High	12	45	64	25	35
Parents Satisfied with School					
Low	33	74	15	11	35
Medium	40	22	36	33	33
High	27	4	49	55	32

Exhibit 2.2: Percent of Schools in Recoded Response Categories by School Climate Health Cluster (School and Teacher Indicators)

	Cluster 1: Experiencing Difficulties	Cluster 2: System Misalignment	Cluster 3: Healthy Social System	Cluster 4: Generational Incongruence	Overall
Principal's Report, Teacher Absenteeism*					
Low	40	66	47	75	56
Medium	17	15	17	13	16
High	43	19	36	12	28
Principal's Report, Limited Resources					
Low	39	19	53	23	33
Medium	36	34	32	35	34
High	25	48	15	42	32
Principal's Report, Academic Emphasis					
Low	59	36	33	13	37
Medium	29	35	35	29	32
High	13	29	32	57	31
Principal's Report, Disorder					
Low	54	28	37	12	35
Medium	31	35	35	29	32
High	15	37	28	60	33
Teacher's Report, Academic Emphasis					
Low	57	36	23	9	33
Medium	32	38	37	29	34
High	12	27	40	62	33
Teacher's Report, School Discipline					
Low	59	56	19	7	38
Medium	30	31	34	26	30
High	11	13	47	67	32
Teacher Job Satisfaction					
Low	44	57	15	15	35
Medium	38	33	37	37	36
High	18	10	48	48	29
Teacher's Report, Students Not Ready for Instruction					
Low	54	17	42	14	33
Medium	32	33	36	32	33
High	14	50	23	55	34
Teacher's Report, Non-Instruction Responsibilities					
Low	42	43	24	32	36
Medium	33	33	33	34	33
High	25	24	43	34	31
Engaging Teacher Practices*					
Low	27	12	39	57	32
Medium	29	23	29	26	27
High	45	65	32	17	41

* Indicates variable where the "Low" category represents the most positive response option.

Not all variables contributed equally to distinguishing among the four clusters. In more fully describing each cluster, the most discriminating variables are used for interpretation. Among the 22 indicators, the four most discriminating variables are school averages for four scales measuring student perceptions of classroom functioning and attitudes towards mathematics and science. *Instructional Clarity, Mathematics* and *Science* reflect school averages of student reports of their perceptions of the teacher's clarity in mathematics classes and science classes, respectively. *Students Like Mathematics* and *Science* reflect, on average, how much students in the school report liking mathematics and science, respectively. Following these four indicators, students' reports of disorderly behavior in mathematics and science lessons are the most discriminating (*Disruptive Lessons, Mathematics* and *Science*, respectively), along with students' reports on school belonging (*Student Sense of Belonging*). The only parent-reported variable (*Parents Satisfied with School*) is also relatively well-discriminating, along with *Teacher's Report, School Discipline*. Details about the discrimination of all individual variables are available in Appendix C.

Exhibits 2.3 and 2.4 complement the percentages in Exhibits 2.1 and 2.2 by showing the mean scores for each variable (based on the recoding scheme described above) in both tabular and graphical format. Clusters with the highest and lowest means on a given indicator are highlighted in Exhibit 2.3. Exhibit 2.4 provides a graphical representation of how well each variable discriminates among the four clusters. The y-axis represents the cluster mean, which is now rescaled to be between 0 and 1.²² The visual spread corresponds with variables with varying discriminating ability, with greater separation indicating greater discrimination and less separation indicating lower discrimination. These highly discriminating indicators are good candidates for developing definitions based on characteristics that are distinct to each group, and these are the variables elaborated in the following detailed cluster definitions. It should be noted that there is variation in responses within all four clusters.

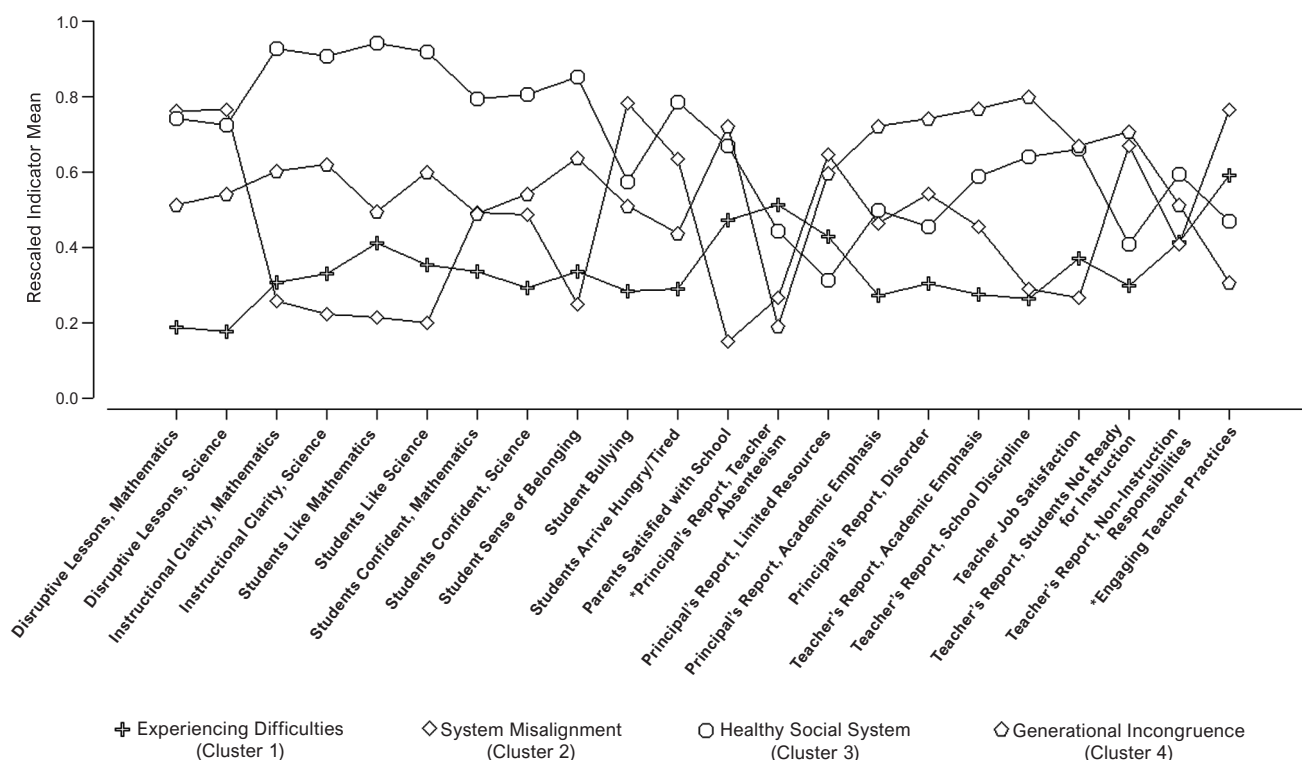
Exhibits 2.1–2.4 are helpful references for interpreting the detailed cluster descriptions presented below.

Exhibit 2.3: Indicator Means by School Climate Health Cluster

	Cluster 1: Experiencing Difficulties	Cluster 2: System Misalignment	Cluster 3: Healthy Social System	Cluster 4: Generational Incongruence
Healthy Relationships Indicators				
Disruptive Lessons, Mathematics	0.38	1.52	1.48	1.03
Disruptive Lessons, Science	0.35	1.53	1.44	1.08
Instructional Clarity, Mathematics	0.61	0.51	1.85	1.20
Instructional Clarity, Science	0.66	0.44	1.81	1.24
Principal's Report, Disorder	0.61	1.08	0.91	1.48
Engaging Teacher Practices*	1.18	1.53	0.94	0.61
Supportive Environments Indicators				
Student Sense of Belonging	0.67	0.50	1.70	1.27
Student Bullying	0.56	1.56	1.14	1.02
Students Arrive Hungry/Tired	0.58	1.27	1.57	0.87
Principal's Report, Limited Resources	0.86	1.29	0.62	1.19
Principal's Report, Academic Emphasis	0.54	0.93	1.00	1.44
Teacher's Report, Academic Emphasis	0.55	0.91	1.17	1.53
Teacher's Report, School Discipline	0.53	0.58	1.28	1.59
Teacher's Report, Students Not Ready for Instruction	0.60	1.34	0.81	1.41
Stakeholder Satisfaction Indicators				
Students Like Mathematics	0.82	0.43	1.88	0.99
Students Like Science	0.70	0.40	1.83	1.20
Students Confident, Mathematics	0.67	0.98	1.58	0.98
Students Confident, Science	0.58	0.97	1.61	1.08
Parents Satisfied with School	0.94	0.30	1.34	1.44
Principal's Report, Teacher Absenteeism*	1.03	0.53	0.89	0.38
Teacher Job Satisfaction	0.74	0.53	1.32	1.33
Teacher's Report, Non-Instruction Responsibilities	0.83	0.82	1.19	1.02

* Indicates variable where a lower mean is associated with more positive responses.

Exhibit 2.4 Graphical MLCA Results – Rescaled Indicator Means by Cluster



* Indicates variable where a lower rescaled mean is associated with more positive responses.

Cluster 1: Experiencing Difficulties



Based on the self-reports of students, parents, principals, and teachers, schools **Experiencing Difficulties (Cluster 1)** struggle to maintain school climate health. This finding across all respondent types differentiates Cluster 1 from the remaining clusters. Cluster 1 has the lowest overall mean scores for the majority of student indicators. Students reported greater frequency of *Student Bullying* in these schools, on average, compared to schools in all other clusters. They also reported more frequently experiencing disruption in their mathematics and science lessons (*Disruptive Lessons, Mathematics* and *Science*, respectively) and arriving at school feeling tired or hungry (*Students Arrive Hungry/Tired*). Students' reports are also the lowest, on average, for *Students Confident, Mathematics* and *Science*.

Even when they do not show the lowest mean score compared to other clusters, Cluster 1 schools' scores remain relatively low on the remaining student level variables: *Student Sense of Belonging*, *Instructional Clarity, Mathematics* and *Science*, *Students Like Mathematics* and *Science*, and their parents' level of satisfaction (*Parents Satisfied with School*). For *Student Sense of Belonging*, *Instructional Clarity, Mathematics* and *Instructional Clarity, Science*, roughly half of schools (48%, 50%, and 48%, respectively) have means recoded as "Low" scores. Although less than half of schools in this cluster have "Low" recoded means for *Students Like Mathematics* and *Students Like Science*, there are also less than one quarter or schools with "High" recoded means. Parent responses (*Parents Satisfied with School*) at the school level for Cluster 1 are more evenly distributed across the recoded response categories, but still, on average, more than twice as many schools reported "Low" mean responses compared to Clusters 3 and 4 (33%, 15%, and 11%, respectively).

On average, teachers also reported more negatively in schools in Cluster 1 compared to other clusters. Teachers in these schools reported that their school is not safe and orderly more frequently than teachers in any other cluster (*Teacher's Report, School Discipline*). They also reported the lowest school levels of academic emphasis (*Teacher's Report, Academic Emphasis*). While *Teacher Job Satisfaction* in Cluster 1 is not the lowest of all four clusters on average, nearly half of schools have teachers who reported “Low” satisfaction (44%), compared to much lower proportions in Clusters 3 and 4 (both 15%). Over one-third of schools in Cluster 1 have teachers who responded in the “High” category for *Engaging Teacher Practices* (45%)^d and “Low” category for *Teacher's Report, Non-Instruction Responsibilities* (42%). Principals in Cluster 1 schools showed the lowest average for *Principal's Report, Academic Emphasis*, the most problems with attendance among teachers (*Principal's Report, Teacher Absenteeism*), and the most disciplinary problems (*Principals' Report, Disorder*).

Schools that are **experiencing difficulties** are navigating apparent dysfunction in maintaining a healthy and productive school climate. These schools are struggling to sustain any of the three features of a healthy school social system (healthy relationships, supportive environments, stakeholder satisfaction). Students reported not feeling a strong connection to their school and experiencing bullying and teachers and principals reported that there is not a strong emphasis on academics and that there are problems with discipline and disorder. The most positive reports came from parents, who indicated that, on average, they were somewhat satisfied with the schools.

Cluster 2: System Misalignment

Reports on the school climate health indicators show contradictions in schools experiencing **System Misalignment (Cluster 2)**. Schools in this cluster have students who reported experiencing the least bullying (i.e., the highest average for *Student Bullying*), as well as the fewest classroom disruptions (i.e., the highest averages for *Disruptive Lessons, Mathematics* and *Science*). Principals, on average, reported to the lowest level of limitations on instruction due to resource shortages (*Principal's Report, Limited Resources*).

However, Cluster 2 schools have the lowest mean scores of any cluster on other student indicators, including *Student Sense of Belonging, Students Like Mathematics* and *Science*, and *Instructional Clarity, Mathematics* and *Science*. Parents of students in Cluster 2 schools also reported the least satisfaction with their child's school (*Parents Satisfied with School*) across all clusters on average. Teachers reported the lowest *Teacher Job Satisfaction* across all clusters. Teachers in these schools, on average, reported employing engaging teaching practices at lower levels compared to other clusters (*Engaging Teacher Practices*).

Different stakeholders report differing perceptions of schools experiencing **system misalignment**, leading to mixed findings regarding the school climate health areas of healthy relationships, supportive environments, and stakeholder satisfaction. Principals and teachers reported that schools in these clusters are well-resourced learning environments, characterized by sufficient resources and students' readiness for instruction. While students reported little bullying or classroom disruption in these schools, the extent of an overall positive evaluation is undercut by reports of low sense of belonging and instructional clarity. Dissatisfaction is also evident in reports of teachers burdened with non-instructional tasks, students who do not like mathematics and science, and parents who expressed relatively negative perceptions of their child's school.

^d As noted in Exhibits 2.2 and 2.3, the “High” category for Engaging Teacher Practices reflects the least desirable response option, meaning teachers report employing the practices less often.



Cluster 3: Healthy Social System

Schools experiencing a **Healthy Social System (Cluster 3)** have the highest levels of school climate health among the four clusters, with respondents who, on average, reported that the school system is functioning well. Schools in Cluster 3 have the most positive average reports for nine of the 22 indicator variables. Cluster 3 schools have the highest means for the following student indicators: *Instructional Clarity, Mathematics and Science; Students Like Mathematics and Science; Students Confident, Mathematics and Science; Student Sense of Belonging; and Students Arrive Hungry/Tired*. On average in Cluster 3 schools, teachers also report being the least burdened by non-instructional tasks (*Teacher's Report, Non-Instruction Responsibilities*).

Even for indicators where Cluster 3 schools do not have the most positive mean score, many schools in this cluster have high percentages of schools with average responses in the “High” recoded category. More than half of Cluster 3 schools had mean responses for *Disruptive Lessons, Mathematics and Science* in the “High” category (57% and 55%, respectively). Nearly half of these schools also had mean responses in the “High” category for *Parents Satisfied with School* (49%). Nearly half of schools in this cluster have teachers reporting in the “High” category for *Teacher Job Satisfaction* (48%).

There is some indication that even in these schools there is room for improvement. More than one-third of schools have principals that reported in the “Low” category for *Principal's Report, Disorder* (37%), indicating a greater prevalence of disciplinary problems. The two academic emphasis indicators (*Principal's Report, Academic Emphasis* and *Teacher's Report, Academic Emphasis*) have means that are noticeably lower than the highest means for these indicators in Cluster 4 schools (Generational Incongruence).

Reports across stakeholders suggest that schools with **healthy social systems** show positive school climate health in the three areas of healthy relationships, supportive environments, and stakeholder satisfaction. Relationships in the classroom, both among students and between students and their teachers, are positive and teachers employ engaging teaching practices. The school environment is positive with high sense of belonging among students and students ready to learn. All groups of stakeholders reported high (if not the highest) levels of satisfaction in these schools on average.



Cluster 4: Generational Incongruence

Indicators of school climate health in schools with **Generational Incongruence (Cluster 4)** are judged differently by students compared to the schools' adult stakeholders. In comparison with Cluster 3 (Healthy Social System), where most schools report in the “High” response category for most student indicators, student responses in Cluster 4 schools are less positive (although often still more positive than Clusters 1 or 2). However, many of the teacher, parent, and principal indicators in Cluster 4 have the highest means of any cluster. Schools in Cluster 4 have the most positive reports on average for *Parents Satisfied with School*, for both *Principal's Report, Academic Emphasis* and *Teacher's Report, Academic Emphasis*, as well as for both *Principal's Report, Disorder* and *Teacher's Report, School Discipline*, and *Teacher Job Satisfaction*.

Some disconnect is implied by teachers who reported that they are, on average, using engaging teaching practices more often than teachers in schools in other clusters (*Engaging Teaching Practices*), whereas students in these schools reported, on average, more moderate

levels of instructional clarity (*Instructional Clarity, Mathematics and Science*) and classroom disruption (*Disruptive Lessons, Mathematics and Science*). For *Students Like Mathematics and Science*, far fewer Cluster 4 schools have means in the “High” category compared to Cluster 3 (28% vs. 88% and 39% vs. 84%, respectively). For *Student Sense of Belonging*, less than half of Cluster 4 schools have means in the “High” category (45%).

While student reports, on average, were more moderate in Cluster 4, reports from adults were overwhelmingly positive. This difference is the defining characteristic of **generational incongruence**. Characterizing this cluster with respect to school climate health themes is difficult without separating students from the adults within the school systems. Teachers and principals in these schools evaluated the quality of the relationships within the schools as if they were operating in different spaces than the students. The teachers and principals report positively on their own practices and the state of discipline while average reports from students might suggest disengagement – they reported classroom experiences characterized by disruptions and less clarity from their teachers. Reports related to school environment follow a similar pattern. Principals and teachers in these schools reported high academic emphasis, teachers in these schools reported high job satisfaction and that schools are safe. Meanwhile, students reported less belonging and experiences of bullying. The generational split is also evident in reports related to stakeholder satisfaction. Student reports in Cluster 4 schools are much less positive about their confidence or affinity for mathematics and science whereas all groups of adult stakeholders gave more positive evaluations.

School Climate Health Clusters and School Socioeconomic Status

One question arising from the results in the previous section is the extent to which school climate health relates to the socioeconomic composition of schools. For example, is the healthy school system cluster simply capturing higher-socioeconomic status schools? Exhibit 2.5 shows the average school-level score on the TIMSS 2023 *Home Socioeconomic Status* scale by cluster within each of the TIMSS 2023 countries.^e Higher average scores represent higher average socioeconomic status.

^e This parent-reported scale was aggregated to the school level for these analyses. For further information see Appendix C.

Exhibit 2.5: School Climate Health Clusters and School-Level Average Home Socioeconomic Status

	Cluster 1: Experiencing Difficulties		Cluster 2: System Misalignment		Cluster 3: Healthy Social System		Cluster 4: Generational Incongruence	
	Percent of Schools	Average Home SES Scale Score	Percent of Schools	Average Home SES Scale Score	Percent of Schools	Average Home SES Scale Score	Percent of Schools	Average Home SES Scale Score
Albania	s	0 ~	0 ~	~ ~	90 (4.4)	8.2 (0.1)	10 (4.4)	9.0 (0.7)
Armenia		0 ~	0 ~	~ ~	92 (3.2)	9.4 (0.1)	8 (3.2)	9.9 (0.2)
Azerbaijan		17 (2.9)	1 ~	~ ~	58 (4.8)	8.6 (0.1)	23 (4.4)	8.7 (0.2)
Bahrain		11 (2.2)	0 ~	~ ~	41 (3.2)	10.0 (0.1)	48 (3.7)	10.0 (0.1)
Belgium (Flemish)		33 (4.4)	54 (5.0)	10.9 (0.1)	5 (2.1)	10.8 (0.4)	8 (3.2)	10.7 (0.3)
Belgium (French)		48 (4.7)	45 (4.9)	11.1 (0.1)	3 ~	~ ~	5 (2.4)	11.0 (0.4)
Bosnia & Herzegovina	s	48 (7.1)	14 (5.2)	9.9 (0.4)	2 ~	~ ~	36 (6.6)	9.5 (0.1)
Brazil		55 (5.2)	0 ~	~ ~	5 (1.6)	7.5 (0.2)	39 (5.1)	8.9 (0.2)
Bulgaria		18 (5.2)	7.7 (0.4)	4 (1.8)	8.9 (1.4)	62 (5.5)	8.7 (0.2)	17 (2.5)
Chile	s	75 (3.9)	1 ~	~ ~	6 (2.4)	8.7 (0.2)	18 (4.1)	9.7 (0.2)
Chinese Taipei		6 (2.0)	82 (4.8)	10.2 (0.1)	4 ~	~ ~	7 (2.4)	10.1 (0.2)
Cyprus		46 (4.8)	9 (3.2)	10.6 (0.3)	5 (2.3)	10.5 (0.3)	39 (5.1)	10.9 (0.1)
Czech Republic	r	25 (4.3)	60 (5.2)	11.0 (0.1)	5 ~	~ ~	9 (2.6)	10.6 (0.5)
Denmark		14 (3.9)	80 (4.8)	10.9 (0.1)	2 ~	~ ~	4 ~	~ ~
Finland		8 (3.0)	11.0 (0.3)	89 (3.0)	11.0 (0.1)	0 ~	2 ~	~ ~
France		24 (3.6)	65 (4.2)	10.9 (0.1)	8 (2.5)	10.5 (0.2)	3 ~	~ ~
Georgia		2 ~	0 ~	~ ~	78 (4.7)	10.0 (0.1)	20 (4.6)	10.2 (0.2)
Germany	s	34 (3.3)	57 (3.9)	10.8 (0.1)	4 (1.8)	10.7 (0.2)	6 (2.2)	11.1 (0.3)
Hong Kong SAR	r	7 (2.7)	80 (3.2)	10.3 (0.2)	3 ~	~ ~	10 (1.9)	11.9 (0.2)
Hungary	s	30 (4.9)	47 (4.3)	10.8 (0.1)	11 (2.9)	11.1 (0.8)	11 (3.8)	10.9 (0.3)
Iran, Islamic Rep. of		10 (2.7)	2 ~	~ ~	74 (4.5)	8.6 (0.1)	15 (3.5)	8.6 (0.3)
Ireland		22 (4.3)	18 (4.8)	11.3 (0.3)	1 ~	~ ~	59 (5.8)	11.3 (0.1)
Italy		56 (5.6)	8 (2.3)	10.1 (0.2)	13 (2.9)	10.0 (0.2)	23 (5.0)	10.3 (0.1)
Japan		0 ~	100 (0.4)	10.1 (0.1)	0 ~	~ ~	0 ~	~ ~
Jordan		27 (4.5)	1 ~	~ ~	57 (5.8)	8.8 (0.1)	15 (4.0)	9.4 (0.2)
Kazakhstan		6 (1.4)	8 (2.8)	9.3 (0.2)	11 (3.1)	9.5 (0.1)	74 (4.3)	9.4 (0.1)
Korea, Rep. of		5 (1.7)	77 (7.1)	11.1 (0.2)	2 ~	~ ~	17 (6.8)	10.9 (0.3)
Kosovo	s	6 (3.5)	0 ~	~ ~	85 (4.7)	9.0 (0.2)	9 (3.0)	9.2 (0.5)
Kuwait		33 (4.0)	0 ~	~ ~	45 (3.7)	9.8 (0.1)	22 (3.1)	9.9 (0.1)
Latvia	s	44 (4.8)	50 (5.7)	10.4 (0.2)	2 ~	~ ~	4 ~	~ ~
Lithuania	s	42 (4.9)	43 (5.3)	10.5 (0.1)	4 ~	~ ~	11 (2.3)	10.8 (0.2)
Macao SAR		20 (4.7)	71 (5.5)	9.9 (0.1)	2 ~	~ ~	7 ~	~ ~
Montenegro	r	14 (3.0)	1 ~	~ ~	58 (4.1)	9.5 (0.1)	26 (3.3)	9.8 (0.2)
Morocco		16 (3.4)	0 ~	~ ~	80 (3.7)	6.9 (0.1)	4 (1.6)	7.4 (0.9)
North Macedonia		9 (2.8)	0 ~	~ ~	75 (3.7)	9.1 (0.1)	16 (3.1)	9.4 (0.2)
Norway (5)	s	21 (4.1)	73 (4.8)	11.4 (0.1)	1 ~	~ ~	5 ~	~ ~
Oman		50 (7.1)	0 ~	~ ~	17 (3.4)	9.6 (0.2)	32 (6.8)	9.6 (0.2)
Poland		14 (4.0)	84 (4.5)	10.7 (0.1)	0 ~	~ ~	3 ~	~ ~
Portugal		23 (5.6)	1 ~	~ ~	35 (3.8)	10.2 (0.1)	41 (5.4)	10.8 (0.2)
Qatar	s	19 (2.3)	1 ~	~ ~	18 (2.2)	10.0 (0.1)	62 (2.8)	10.3 (0.1)
Romania	s	10 (3.9)	0 ~	~ ~	52 (6.2)	9.2 (0.2)	39 (5.6)	9.5 (0.3)
Saudi Arabia	r	18 (2.8)	0 ~	~ ~	51 (4.0)	9.1 (0.1)	31 (3.7)	9.3 (0.1)
Serbia		14 (3.6)	3 (1.2)	10.8 (0.2)	55 (4.4)	8.9 (0.3)	27 (4.4)	9.9 (0.1)
Singapore	r	13 (2.6)	73 (3.2)	11.2 (0.0)	2 ~	~ ~	13 (2.5)	11.2 (0.1)
Slovak Republic		34 (4.9)	42 (5.0)	10.5 (0.1)	15 (4.2)	8.3 (0.5)	10 (2.8)	10.1 (0.3)
Slovenia		28 (4.3)	70 (4.5)	10.8 (0.1)	1 ~	~ ~	1 ~	~ ~
South Africa (5)	r	52 (3.8)	4 ~	~ ~	17 (2.6)	8.7 (0.2)	28 (3.9)	9.0 (0.2)
Spain		39 (3.6)	3 (0.9)	10.7 (0.2)	5 (1.2)	10.0 (0.2)	53 (3.5)	10.3 (0.1)
Sweden	r	13 (3.1)	80 (4.2)	11.3 (0.1)	3 ~	~ ~	4 (1.8)	11.4 (0.3)
Türkiye (5)		29 (6.6)	1 ~	~ ~	46 (6.5)	9.1 (0.3)	24 (6.1)	10.1 (0.4)
Uzbekistan		3 (1.2)	0 ~	~ ~	86 (3.3)	8.9 (0.1)	11 (3.1)	8.8 (0.2)
Canada	x	51 (3.3)	8 (1.4)	11.4 (0.1)	12 (2.3)	11.5 (0.1)	30 (3.3)	11.4 (0.2)
United Arab Emirates	x	21 (1.3)	0 ~	~ ~	38 (1.7)	10.2 (0.0)	40 (1.6)	10.5 (0.1)
New Zealand	y	72 (4.6)	3 (1.2)	~ ~	5 ~	~ ~	20 (4.6)	~ ~
Australia		--	--	--	--	--	--	--
England		--	--	--	--	--	--	--
Netherlands		--	--	--	--	--	--	--
United States		--	--	--	--	--	--	--

Standard errors appear in parentheses (). Because of rounding some results may appear inconsistent.

An “r” indicates home socioeconomic status data are available for at least 70% but less than 85% of the students. An “s” indicates home socioeconomic status data are available for at least 50% but less than 70% of the students. An “x” indicates home socioeconomic status data are available for at least 40% but less than 50% of the students—interpret with caution. A “y” indicates home socioeconomic status data are available for less than 40% of the students. A tilde (~) indicates insufficient data to report result. A dash (-) indicates comparable data not available.

Exhibit 2.5 does not report average scores on the *Home Socioeconomic Status* scale for clusters including fewer than 5 schools within a given country. Many countries with, on average, higher *Home Socioeconomic Status* scale scores (see [TIMSS 2023 International Results](#)) do not have enough schools assigned to the *healthy social system* cluster to have an average score reported for those schools. In contrast, countries with, on average, lower *Home Socioeconomic Status* scale scores often have relatively high percentages of schools assigned to this cluster. In addition, only 7.6 percent of the variance in school-level socioeconomic status is explained by the clusters on average across countries.^f This provides evidence that the school climate health clusters are not simply proxies for the socioeconomic composition of the school.

Among the relatively small number of countries with average *Home Socioeconomic Status* scores reported for all four clusters, there is some evidence that schools in the *experiencing difficulties* cluster may have, on average, students with lower socioeconomic status; however, there is little difference among the other three clusters. For example, in Belgium (Flemish), the confidence intervals calculated from the standard errors of the average socioeconomic status scale scores overlap for the *system misalignment*, *healthy social system*, and *generational incongruence* clusters, providing evidence that the averages are not significantly different from one another. Germany, Hungary, Italy, Kazakhstan, and Spain, which also have socioeconomic status scores reported in all four clusters, show the same pattern.

As it is true in all international comparisons of self-reports, students, teachers, and principals may show some cultural differences in how they interpret questions, and how they answer.^{23,24} Hence, the proportions of schools assigned to clusters within countries may be somewhat affected by cultural differences in item interpretation or the prevalence of certain response tendencies within particular cultural contexts.²⁵ For example, there is some evidence of cross-cultural variation in individuals' proclivity for potential socially desirable responding.²⁶ Further research can explore the extent to which these phenomena might have contributed to the proportions of schools assigned to clusters in individual countries.

School Climate Health Clusters and Grade 4 Mathematics and Science Achievement

This report section connects the school climate health clusters to mathematics and science achievement. Exhibit 3.1 shows each country's school-level average mathematics achievement within the four clusters. It is important to keep in mind the distribution of schools across clusters within countries, as well as overall average achievement in mathematics, when interpreting these results. Many of the countries showing high overall levels of mathematics achievement in the [TIMSS 2023 International Results](#) had large proportions of schools assigned to the *system misalignment* cluster. Also worth noting is the relatively small percentages of schools assigned to certain clusters within some groups of countries; with smaller percentages, there is less confidence in estimates of average achievement. Similar to the socioeconomic status analyses presented above, average achievement is not estimated for clusters with fewer than five schools assigned to them.

^f Only countries where home socioeconomic status data were available for at least 50% of the students are included in this analysis. See Appendix C for more details.

Exhibit 3.1: School Climate Health Clusters and Average Achievement in Grade 4 Mathematics

	Cluster 1: Experiencing Difficulties		Cluster 2: System Misalignment		Cluster 3: Healthy Social System		Cluster 4: Generational Incongruence	
	Percent of Schools	Average Achievement	Percent of Schools	Average Achievement	Percent of Schools	Average Achievement	Percent of Schools	Average Achievement
Albania	0 ~	~ ~	0 ~	~ ~	89 (4.0)	523 (7.8)	11 (4.0)	507 (16.8)
Armenia	0 ~	~ ~	0 ~	~ ~	92 (3.2)	527 (5.1)	8 (3.2)	491 (13.0)
Australia	54 (4.8)	498 (5.6)	5 (1.7)	558 (15.6)	3 (1.4)	561 (15.7)	38 (4.9)	542 (5.4)
Azerbaijan	17 (2.8)	414 (8.4)	1 ~	~ ~	60 (4.5)	496 (5.7)	22 (4.2)	467 (11.2)
Bahrain	12 (2.3)	429 (13.9)	0 ~	~ ~	41 (3.2)	481 (6.5)	48 (3.7)	462 (6.3)
Belgium (Flemish)	34 (4.4)	491 (7.1)	54 (4.9)	533 (2.7)	5 (2.1)	529 (9.6)	8 (3.1)	517 (12.4)
Belgium (French)	48 (4.7)	474 (4.8)	45 (4.8)	506 (3.8)	3 ~	~ ~	5 (2.3)	499 (17.6)
Bosnia & Herzegovina	46 (6.2)	427 (7.1)	11 (4.2)	465 (12.3)	9 (5.6)	495 (44.7)	34 (5.6)	450 (5.1)
Brazil	55 (5.2)	382 (5.8)	0 ~	~ ~	5 (1.6)	430 (21.8)	39 (5.1)	421 (8.9)
Bulgaria	18 (5.2)	448 (18.6)	4 (1.8)	456 (66.5)	62 (5.5)	505 (9.3)	17 (2.5)	541 (9.2)
Canada	52 (3.0)	483 (3.9)	8 (1.2)	525 (5.3)	12 (2.0)	513 (12.5)	29 (3.1)	513 (4.0)
Chile	73 (3.8)	428 (4.8)	1 ~	~ ~	4 (1.8)	452 (8.1)	21 (4.1)	462 (10.8)
Chinese Taipei	6 (2.0)	575 (10.2)	82 (4.8)	595 (3.8)	4 ~	~ ~	7 (2.4)	603 (6.4)
Cyprus	46 (4.8)	502 (4.8)	9 (3.2)	521 (10.9)	5 (2.3)	538 (9.1)	39 (5.1)	521 (5.4)
Czech Republic	25 (4.3)	508 (7.5)	60 (5.2)	532 (4.3)	5 ~	~ ~	9 (2.6)	561 (19.3)
Denmark	14 (3.9)	490 (10.8)	80 (4.8)	524 (3.6)	2 ~	~ ~	4 ~	~ ~
England	40 (5.1)	529 (4.4)	7 (3.1)	576 (11.0)	3 ~	~ ~	50 (4.7)	566 (4.2)
Finland	8 (3.0)	478 (17.6)	89 (3.0)	529 (4.2)	0 ~	~ ~	2 ~	~ ~
France	25 (3.4)	459 (7.2)	64 (3.9)	493 (5.1)	7 (2.4)	506 (8.2)	3 ~	~ ~
Georgia	2 ~	~ ~	0 ~	~ ~	78 (4.7)	506 (8.7)	20 (4.6)	501 (12.2)
Germany	36 (3.4)	488 (15.4)	55 (3.8)	532 (3.1)	4 (1.7)	543 (9.7)	5 (2.1)	563 (11.0)
Hong Kong SAR	7 (2.6)	550 (15.5)	80 (3.1)	577 (7.5)	2 ~	~ ~	10 (1.9)	628 (19.1)
Hungary	32 (4.9)	467 (11.2)	45 (4.2)	516 (7.6)	11 (2.8)	544 (21.0)	13 (4.0)	546 (22.4)
Iran, Islamic Rep. of	10 (2.7)	364 (18.5)	2 ~	~ ~	74 (4.5)	423 (6.8)	15 (3.5)	405 (14.8)
Ireland	22 (4.3)	518 (6.9)	18 (4.8)	575 (6.3)	1 ~	~ ~	59 (5.8)	551 (4.6)
Italy	55 (5.6)	504 (4.6)	8 (2.2)	547 (11.3)	12 (2.9)	522 (6.3)	25 (5.1)	531 (6.5)
Japan	0 ~	~ ~	100 (0.4)	586 (2.8)	0 ~	~ ~	0 ~	~ ~
Jordan	27 (4.5)	405 (11.7)	1 ~	~ ~	57 (5.8)	451 (9.5)	15 (4.0)	435 (13.6)
Kazakhstan	6 (1.4)	476 (10.5)	8 (2.8)	502 (27.8)	11 (3.1)	504 (15.8)	74 (4.3)	485 (7.2)
Korea, Rep. of	5 (1.7)	537 (29.7)	77 (7.1)	582 (6.8)	2 ~	~ ~	16 (6.8)	603 (17.6)
Kosovo	5 (3.4)	350 (41.7)	0 ~	~ ~	86 (4.6)	445 (6.3)	9 (3.0)	418 (11.7)
Kuwait	33 (4.0)	350 (14.7)	0 ~	~ ~	45 (3.7)	385 (6.9)	21 (3.1)	401 (13.8)
Latvia	44 (4.8)	506 (6.1)	50 (5.7)	531 (5.5)	2 ~	~ ~	4 ~	~ ~
Lithuania	42 (4.9)	517 (6.2)	43 (5.3)	560 (5.8)	4 ~	~ ~	11 (2.3)	563 (7.0)
Macao SAR	20 (4.7)	524 (9.8)	71 (5.5)	572 (4.3)	2 ~	~ ~	7 ~	~ ~
Montenegro	13 (2.9)	453 (8.6)	2 ~	~ ~	58 (4.1)	493 (5.2)	26 (3.3)	481 (6.7)
Morocco	16 (3.4)	348 (8.1)	0 ~	~ ~	80 (3.7)	413 (6.9)	4 (1.6)	382 (24.8)
Netherlands	24 (4.7)	514 (7.2)	59 (5.3)	544 (2.8)	5 (2.4)	540 (16.4)	11 (3.5)	539 (3.4)
New Zealand	73 (4.1)	474 (4.1)	3 (1.1)	545 (6.4)	4 ~	~ ~	20 (4.1)	522 (8.5)
North Macedonia	9 (2.8)	457 (17.5)	0 ~	~ ~	75 (3.7)	482 (7.0)	16 (3.1)	470 (9.6)
Norway (5)	22 (4.0)	518 (5.3)	73 (4.6)	535 (4.6)	1 ~	~ ~	5 ~	~ ~
Oman	52 (7.0)	411 (9.5)	0 ~	~ ~	16 (3.4)	485 (23.0)	32 (6.4)	457 (29.3)
Poland	14 (4.0)	518 (6.4)	84 (4.5)	543 (4.5)	0 ~	~ ~	3 ~	~ ~
Portugal	23 (5.6)	490 (12.3)	1 ~	~ ~	35 (3.7)	532 (4.7)	41 (5.4)	536 (6.4)
Qatar	19 (2.2)	444 (6.7)	1 ~	~ ~	18 (2.2)	477 (7.5)	62 (2.8)	474 (4.0)
Romania	11 (2.9)	498 (16.7)	0 ~	~ ~	54 (4.8)	538 (8.1)	34 (4.7)	536 (7.6)
Saudi Arabia	18 (2.8)	388 (7.7)	0 ~	~ ~	50 (4.0)	449 (7.6)	32 (3.8)	404 (8.4)
Serbia	14 (3.6)	490 (10.1)	3 (1.2)	534 (7.9)	55 (4.4)	508 (20.1)	27 (4.4)	527 (15.5)
Singapore	13 (2.6)	576 (5.1)	73 (3.2)	613 (3.1)	2 ~	~ ~	13 (2.5)	624 (9.0)
Slovak Republic	34 (4.9)	489 (10.7)	42 (5.0)	531 (3.9)	15 (4.2)	487 (19.7)	10 (2.8)	532 (18.0)
Slovenia	28 (4.2)	497 (4.0)	70 (4.3)	516 (2.3)	1 ~	~ ~	1 ~	~ ~
South Africa (5)	52 (3.4)	334 (5.6)	3 ~	~ ~	16 (2.4)	355 (7.6)	28 (3.5)	385 (13.9)
Spain	39 (3.5)	481 (4.9)	3 (0.9)	496 (13.1)	5 (1.2)	487 (10.3)	53 (3.5)	505 (3.1)
Sweden	13 (3.0)	487 (9.7)	80 (4.0)	538 (2.9)	3 ~	~ ~	4 (1.7)	534 (16.9)
Türkiye (5)	29 (6.6)	531 (10.3)	1 ~	~ ~	46 (6.4)	531 (9.2)	24 (6.0)	585 (9.3)
United Arab Emirates	21 (1.2)	427 (4.3)	0 ~	~ ~	38 (1.7)	526 (3.0)	40 (1.5)	502 (3.5)
United States	54 (5.8)	489 (3.8)	3 (1.0)	543 (9.5)	2 (0.9)	557 (16.0)	41 (6.0)	530 (11.8)
Uzbekistan	3 (1.2)	395 (15.4)	0 ~	~ ~	86 (3.3)	445 (3.9)	11 (3.1)	423 (8.9)

Standard errors appear in parentheses (). Because of rounding some results may appear inconsistent.

A tilde (~) indicates insufficient data to report result.

Exhibit 3.2 provides an overview of the clusters in which the TIMSS 2023 countries had their highest and lowest average mathematics scores. About 40 percent of countries (24 of 58) have their highest average mathematics achievement in the *healthy school climate* cluster. In 53 countries, the school cluster with the lowest achievement is *experiencing difficulties*. This suggests that while dysfunctional school climate is associated with lower achievement in most countries, the picture is more complicated and results are not as clear when considering school climate characteristics associated with higher achievement.

Exhibit 3.2: TIMSS 2023 Highest and Lowest Average Mathematics Achievement by School Climate Health Cluster

		Cluster			
		Experiencing Difficulties	System Misalignment	Healthy Social System	Generational Incongruence
Number of Countries	Highest Average Achievement in this Cluster	0	17	24	16
	Lowest Average Achievement in this Cluster	53	0	2	3

Average achievement for Japan could only be estimated for the *system misalignment* cluster, so it is excluded from these counts.

Türkiye had equivalent lowest average achievement for the *experiencing difficulties* and *healthy social system* clusters.

It is worth noting that many countries with the highest overall average mathematics achievement in TIMSS 2023 have less than five schools assigned to the *healthy social system* cluster and so average achievement cannot be estimated. These countries still evidently have overall very high-achieving students. Even among higher-achieving countries where average achievement can be estimated for the *healthy social system* cluster, there is not a clear indication that those are the highest achieving schools. For example, in the Netherlands, highest average mathematics achievement is observed among schools in the *system misalignment* cluster. In Türkiye, average achievement is identical for schools assigned to the *healthy social system* and *experiencing difficulties* clusters, with the highest average achievement for schools assigned to the *generational incongruence* cluster.

Exhibit 3.3: School Climate Health Clusters and Average Achievement in Grade 4 Science

	Cluster 1: Experiencing Difficulties		Cluster 2: System Misalignment		Cluster 3: Healthy Social System		Cluster 4: Generational Incongruence	
	Percent of Schools	Average Achievement	Percent of Schools	Average Achievement	Percent of Schools	Average Achievement	Percent of Schools	Average Achievement
Albania	0 ~	~ ~	0 ~	~ ~	89 (4.0)	498 (7.4)	11 (4.0)	492 (16.9)
Armenia	0 ~	~ ~	0 ~	~ ~	92 (3.2)	474 (5.3)	8 (3.2)	451 (21.5)
Australia	54 (4.8)	531 (4.5)	5 (1.7)	574 (13.0)	3 (1.4)	572 (12.8)	38 (4.9)	565 (3.6)
Azerbaijan	17 (2.8)	353 (10.2)	1 ~	~ ~	60 (4.5)	421 (5.2)	22 (4.2)	408 (12.1)
Bahrain	12 (2.3)	430 (13.8)	0 ~	~ ~	41 (3.2)	500 (6.4)	48 (3.7)	472 (7.0)
Belgium (Flemish)	34 (4.4)	457 (7.8)	54 (4.9)	501 (3.6)	5 (2.1)	510 (10.3)	8 (3.1)	491 (12.1)
Belgium (French)	48 (4.7)	465 (4.7)	45 (4.8)	501 (4.0)	3 ~	~ ~	5 (2.3)	496 (12.7)
Bosnia & Herzegovina	46 (6.2)	428 (8.3)	11 (4.2)	465 (11.3)	9 (5.6)	485 (29.7)	34 (5.6)	453 (4.1)
Brazil	55 (5.2)	404 (6.8)	0 ~	~ ~	5 (1.6)	441 (16.3)	39 (5.1)	451 (8.7)
Bulgaria	18 (5.2)	432 (24.6)	4 (1.8)	421 (95.5)	62 (5.5)	498 (13.1)	17 (2.5)	545 (10.3)
Canada	52 (3.0)	508 (3.9)	8 (1.2)	524 (3.9)	12 (2.0)	534 (10.8)	29 (3.1)	534 (3.2)
Chile	73 (3.8)	464 (4.9)	1 ~	~ ~	4 (1.8)	478 (7.5)	21 (4.1)	495 (10.1)
Chinese Taipei	6 (2.0)	544 (7.9)	82 (4.8)	560 (3.9)	4 ~	~ ~	7 (2.4)	569 (6.4)
Cyprus	46 (4.8)	466 (5.5)	9 (3.2)	484 (17.5)	5 (2.3)	516 (11.9)	39 (5.1)	495 (7.0)
Czech Republic	25 (4.3)	504 (7.5)	60 (5.2)	530 (3.9)	5 ~	~ ~	9 (2.6)	561 (20.7)
Denmark	14 (3.9)	485 (9.9)	80 (4.8)	525 (4.0)	2 ~	~ ~	4 ~	~ ~
England	40 (5.1)	538 (4.0)	7 (3.1)	574 (9.0)	3 ~	~ ~	50 (4.7)	569 (4.2)
Finland	8 (3.0)	494 (16.6)	89 (3.0)	543 (4.2)	0 ~	~ ~	2 ~	~ ~
France	25 (3.4)	463 (8.6)	64 (3.9)	498 (4.7)	7 (2.4)	517 (10.9)	3 ~	~ ~
Georgia	2 ~	~ ~	0 ~	~ ~	78 (4.7)	481 (8.4)	20 (4.6)	464 (12.5)
Germany	36 (3.4)	478 (16.1)	55 (3.8)	525 (3.9)	4 (1.7)	538 (12.0)	5 (2.1)	561 (20.4)
Hong Kong SAR	7 (2.6)	501 (14.2)	80 (3.1)	527 (7.7)	2 ~	~ ~	10 (1.9)	587 (19.0)
Hungary	32 (4.9)	481 (9.6)	45 (4.2)	522 (6.0)	11 (2.8)	542 (40.8)	13 (4.0)	548 (23.1)
Iran, Islamic Rep. of	10 (2.7)	372 (20.1)	2 ~	~ ~	74 (4.5)	434 (7.3)	15 (3.5)	423 (16.8)
Ireland	22 (4.3)	509 (7.2)	18 (4.8)	544 (10.2)	1 ~	~ ~	59 (5.8)	539 (5.1)
Italy	55 (5.6)	504 (4.5)	8 (2.2)	536 (8.2)	12 (2.9)	518 (5.5)	25 (5.1)	522 (6.9)
Japan	0 ~	~ ~	100 (0.4)	551 (3.3)	0 ~	~ ~	0 ~	~ ~
Jordan	27 (4.5)	395 (12.3)	1 ~	~ ~	57 (5.8)	441 (9.0)	15 (4.0)	428 (11.1)
Kazakhstan	6 (1.4)	477 (13.4)	8 (2.8)	457 (17.9)	11 (3.1)	491 (19.5)	74 (4.3)	461 (7.3)
Korea, Rep. of	5 (1.7)	526 (31.6)	77 (7.1)	576 (4.2)	2 ~	~ ~	16 (6.8)	601 (18.3)
Kosovo	5 (3.4)	312 (32.9)	0 ~	~ ~	86 (4.6)	398 (6.9)	9 (3.0)	370 (11.9)
Kuwait	33 (4.0)	328 (16.6)	0 ~	~ ~	45 (3.7)	384 (8.4)	21 (3.1)	398 (16.2)
Latvia	44 (4.8)	503 (7.1)	50 (5.7)	525 (6.6)	2 ~	~ ~	4 ~	~ ~
Lithuania	42 (4.9)	495 (6.8)	43 (5.3)	536 (5.1)	4 ~	~ ~	11 (2.3)	538 (7.6)
Macao SAR	20 (4.7)	471 (12.1)	71 (5.5)	527 (5.2)	2 ~	~ ~	7 ~	~ ~
Montenegro	13 (2.9)	436 (9.3)	2 ~	~ ~	58 (4.1)	475 (4.7)	26 (3.3)	464 (7.0)
Morocco	16 (3.4)	339 (10.4)	0 ~	~ ~	80 (3.7)	411 (7.9)	4 (1.6)	360 (20.4)
Netherlands	24 (4.7)	492 (7.8)	59 (5.3)	524 (3.7)	5 (2.4)	515 (22.5)	11 (3.5)	519 (4.2)
New Zealand	73 (4.1)	503 (3.5)	3 (1.1)	572 (7.4)	4 ~	~ ~	20 (4.1)	547 (7.1)
North Macedonia	9 (2.8)	418 (20.2)	0 ~	~ ~	75 (3.7)	450 (5.3)	16 (3.1)	437 (10.4)
Norway (5)	22 (4.0)	516 (6.7)	73 (4.6)	536 (5.0)	1 ~	~ ~	5 ~	~ ~
Oman	52 (7.0)	416 (8.2)	0 ~	~ ~	16 (3.4)	498 (23.5)	32 (6.4)	459 (23.4)
Poland	14 (4.0)	525 (6.8)	84 (4.5)	545 (4.7)	0 ~	~ ~	3 ~	~ ~
Portugal	23 (5.6)	494 (7.5)	1 ~	~ ~	35 (3.7)	520 (3.9)	41 (5.4)	524 (6.3)
Qatar	19 (2.2)	441 (7.3)	1 ~	~ ~	18 (2.2)	490 (7.6)	62 (2.8)	481 (3.9)
Romania	11 (2.9)	488 (14.2)	0 ~	~ ~	54 (4.8)	522 (7.5)	34 (4.7)	521 (6.9)
Saudi Arabia	18 (2.8)	382 (8.8)	0 ~	~ ~	50 (4.0)	465 (8.1)	32 (3.8)	411 (8.2)
Serbia	14 (3.6)	482 (7.9)	3 (1.2)	518 (6.9)	55 (4.4)	506 (20.5)	27 (4.4)	511 (13.5)
Singapore	13 (2.6)	572 (5.0)	73 (3.2)	606 (3.0)	2 ~	~ ~	13 (2.5)	614 (8.1)
Slovak Republic	34 (4.9)	491 (13.2)	42 (5.0)	542 (4.7)	15 (4.2)	479 (22.7)	10 (2.8)	541 (18.2)
Slovenia	28 (4.2)	511 (4.3)	70 (4.3)	528 (2.6)	1 ~	~ ~	1 ~	~ ~
South Africa (5)	52 (3.4)	271 (8.4)	3 ~	~ ~	16 (2.4)	298 (15.3)	28 (3.5)	339 (20.2)
Spain	39 (3.5)	487 (5.2)	3 (0.9)	502 (14.2)	5 (1.2)	499 (7.9)	53 (3.5)	513 (3.1)
Sweden	13 (3.0)	482 (11.8)	80 (4.0)	539 (3.2)	3 ~	~ ~	4 (1.7)	532 (20.3)
Türkiye (5)	29 (6.6)	551 (9.3)	1 ~	~ ~	46 (6.4)	556 (6.7)	24 (6.0)	599 (8.9)
United Arab Emirates	21 (1.2)	412 (5.1)	0 ~	~ ~	38 (1.7)	526 (3.5)	40 (1.5)	497 (3.9)
United States	54 (5.8)	507 (4.0)	3 (1.0)	554 (9.2)	2 (0.9)	563 (15.1)	41 (6.0)	553 (8.5)
Uzbekistan	3 (1.2)	355 (9.5)	0 ~	~ ~	86 (3.3)	416 (4.0)	11 (3.1)	393 (11.9)

Standard errors appear in parentheses (). Because of rounding some results may appear inconsistent.
A tilde (~) indicates insufficient data to report result.

Exhibit 3.3 shows the results for science. The pattern is similar to what was observed in mathematics. Average science achievement cannot be reliably estimated for the *healthy social system* cluster in many higher-achieving countries because of the small number of schools assigned to this cluster. Higher-achieving countries where average achievement is estimated for this cluster have their highest average science achievement in other clusters—for example, *system misalignment* (Australia) or *generational incongruence* (Türkiye).

Exhibit 3.4 provides an overview of clusters where the TIMSS 2023 countries had their highest and lowest average science achievement. As in mathematics, about 40 percent of countries (24 of 58) have their highest mean science achievement in schools assigned to *healthy social system* and in 51 countries the school cluster with the lowest average science achievement is *experiencing difficulties*.

Exhibit 3.4: TIMSS 2023 Highest and Lowest Average Science Achievement by School Climate Health Cluster

		Cluster			
		Experiencing Difficulties	System Misalignment	Healthy Social System	Generational Incongruence
Number of Countries	Highest Average Achievement in this Cluster	0	17	24	17
	Lowest Average Achievement in this Cluster	51	2	1	3

Average achievement for Japan could only be estimated for the *system misalignment* cluster, so it is excluded from these counts. Canada had equivalent highest average achievement for the *healthy social system* and *generational incongruence* clusters.

The results observed for mathematics and science suggest an overall pattern that in many countries, schools in healthier school climate clusters tend to have higher average achievement than those in poor school climate health clusters. In the majority of countries, the lowest average achievement in both subjects is in the *experiencing difficulties* cluster. As noted above, it is important to consider the proportions of schools in each cluster by country, as well as overall country achievement, and take care in making inferences at the country level. It is also important to consider the potentially confounding role of demographic characteristics on school climate health cluster assignment. These results require further investigation at the local level to expand understanding of possible relationships between school environmental conditions and achievement.

Key Takeaways

- **The TIMSS 2023 context questionnaires include a great variety of variables related to school climate health and can be used to create profiles characterizing school environments based on reports of multiple groups of stakeholders.** Response patterns on these variables distinctly separated schools into practically meaningful typologies of school climate health—with student variables being the most discriminating overall. For variables found to be influential in determining cluster assignment, intervention can be targeted to support those schools where community members report they are not flourishing.

- **The four clusters highlight the importance of accounting for reports from multiple stakeholders in evaluating the quality of school environments.** Two of the clusters identified with the help of MLCA can be described by their overall low mean scores relative to schools in the sample (Cluster 1: Experiencing Difficulties) or their relatively high scores (Cluster 3: Healthy Social System) across all respondent types. Cluster 2 (System Misalignment) contains schools where reports from different stakeholders do not align. Cluster 4 (Generational Incongruence) is similar, with the added pattern that the adults involved are, on average, more positive where students are muted in their satisfaction and evaluations of relationships and the school environment. **This suggests that efforts to positively shape school climate should account for the varying experiences of different stakeholders within school systems rather than obtaining information from a single source (for example, from only students or only teachers).**
- **Overall, there is a limited association between cluster membership and the socioeconomic status of students within schools.** However, several countries whose students have, on average, higher socioeconomic status had relatively small numbers of schools in the *healthy social system* cluster. While schools assigned to the *experiencing difficulties* cluster showed on average somewhat lower socioeconomic status scores, there was little difference observed among the averages of the other three clusters. It is important to note that with, on average, less than eight percent of the variance in socioeconomic status explained by the four clusters, there is substantial variance in socioeconomic status levels also within each cluster.
- **There appears to be a relationship between cluster membership and achievement in mathematics and science, although this should be interpreted with some caution.** Within most countries, schools assigned to the *experiencing difficulties* cluster had the lowest average achievement compared to schools assigned to other clusters. Further investigation is encouraged to clarify the variation in mathematics and science achievement across school climate health clusters. There is at least some evidence of a negative association between poor school climate health and school average mathematics and science achievement within schools; however, the association between positive school climate and increased achievement is less clear. Therefore, the relationship between cluster membership and achievement merits additional exploration in light of potential intervening factors within countries. Cluster proportions vary across countries, as expected and explicitly enabled by the flexibility of the analytic approach; however, it is noteworthy that many high-achieving countries had very few schools assigned to the *healthy social system* cluster and relatively more schools assigned to the *system misalignment* cluster. This should not be interpreted to imply that system misalignment necessarily creates better learning conditions, especially in light of the finding that the *healthy social system* cluster was associated with higher achievement at a higher rate across countries than other clusters. Researchers interested in disentangling complex relationships between school climate and academic achievement should focus on within-country relationships rather than on how a given country compares with high-achieving countries.

Future Directions

The results of the TIMSS Insights report suggest several areas for further investigation of school climate health and its relationship to student achievement. This report is an important step in acknowledging the complexity of measuring school climate and explicitly accounting for schools as complex social systems with different groups of stakeholders enacting varying roles.

One area for future research involves validating the findings of this report by investigating how school climate health may manifest in lower-secondary schools using Grade 8 TIMSS 2023 data. Even with the absence of parent satisfaction as a measure,⁹ multidimensional profiles of Grade 8 schools based on multiple stakeholders (students, principals, teachers) could still be established with the same methodologies. Similarly, this study's analytic approach could be replicated using results from the Progress in International Reading and Literacy Study (PIRLS) to see how the context of a different school subject might influence the profiles of Grade 4 schools described in this report or their relationship with achievement.

Another avenue of future research is investigating potential confounding school characteristics and response patterns that may have influenced the results in this report. It is possible that variables other than those included in the model can explain the association between the schools assigned to the same school climate health cluster. This does not discredit the usefulness of a multidimensional, multi-participant construct like school climate health, but can add to the systems-level understanding of forces acting on school learning environments. This report includes an initial exploration of the relationship between school-level socioeconomic status and school climate health; this investigation could be extended or applied to other school characteristics.

Further confoundedness could also be identified in cultural differences in response patterns to self-report surveys. Some countries that have the highest proportions of their schools in *healthy social systems* may also be more likely, on average, to respond more positively to survey questions compared to other countries. It could also be possible that relatively disengaged respondents choose the most positive option for all items to which they respond. The extent to which these habits impact the magnitude of positive or negative responses may reduce discrimination in some variables and is worthy of further investigation.

Finally, these results suggest room for improvement in the measurement of factors related to school climate in the TIMSS questionnaires in future assessment cycles. Many of the most discriminating items in this report were at the student level, and most of the variables included were from them as well. At the teacher or principal level, more information could be collected about the interactions between teachers and principals, the collegial environment among teachers, and the congruence of how values are communicated from school leadership to teachers before being communicated to students.

⁹ The Home Questionnaire was not administered at Grade 8 in TIMSS 2023.

Appendix A: About TIMSS 2023

The 2023 cycle of the Trends in International Mathematics and Science Study (TIMSS 2023) aims to measure and compare the mathematics and science achievement of fourth- and eighth-grade students around the world. TIMSS provides valuable insights into mathematics and science achievement differences across and within countries and helps policymakers and educators understand factors contributing to student learning.

The data presented in this report come from the TIMSS 2023 mathematics and science assessment, as well as the Student, Home, School, and Teacher Questionnaires. TIMSS 2023 also collected national policy-level data from National Research Coordinators from the participating countries and benchmarking entities.

Many resources are available for obtaining more information about TIMSS 2023, accessing TIMSS 2023 data, and conducting additional analyses. These resources include the following.

- [*TIMSS 2023 Assessment Frameworks*](#): provides information about the TIMSS 2023 mathematics and science assessments, context questionnaires, and assessment design. This includes a special supplement describing the measurement of environmental attitudes and behaviors in the TIMSS 2023 context questionnaires.
- [*TIMSS 2023 International Results in Mathematics and Science*](#): provides a summary of fourth- and eighth-grade students' mathematics and science achievement, as well as information about how contextual factors relate to students' achievement.
- [*TIMSS 2023 Encyclopedia*](#): contains information about national contexts for countries participating in TIMSS 2023, including a chapter written by each country describing its educational system and responses to the TIMSS 2023 Curriculum Questionnaires.
- [*TIMSS 2023 Technical Report: Methods and Procedures*](#): details instrument development, sample design and implementation, operations, and analytical procedures.
- [*TIMSS 2023 Context Questionnaires*](#): show the international versions of the contextual items to which students, parents, school principals, teachers, and National Research Coordinators responded.
- [*TIMSS 2023 International Database*](#): includes data collected from the mathematics and science assessments and context questionnaires and achievement and contextual scale estimates for the countries and benchmarking entities that participated in TIMSS 2023. The accompanying User Guide provides information about database contents and analysis resources.
- [*TIMSS 2023 Insights Series*](#): is a set of reports exploring different educational research areas using TIMSS 2023 data. Reports in the series provide an in-depth examination of particular topics beyond what can be covered in the International Results.

Appendix B: Context Questionnaire Items Included in Analysis

Healthy Relationships

Teacher Questionnaire Items

Variable Label (MLCA Variable Name)	Individual Variable Names	Item Stem and Response Options	Item Text
Engaging Teacher Practices (ATBG12)	ATBG12A	How often do you do the following in teaching this class? • Every or almost every lesson • About half the lessons • Some lessons • Never	a) Relate the lesson to students' daily lives
	ATBG12B		b) Ask students to explain their answers
	ATBG12C		c) Communicate goals or objectives for the lesson to the students
	ATBG12D		d) Ask students to complete challenging exercises that require them to go beyond the instruction
	ATBG12E		e) Encourage classroom discussions among students
	ATBG12F		f) Link new content to students' prior knowledge
	ATBG12G		g) Ask students to decide their own problem solving procedures

Student Questionnaire Items

Variable Label (MLCA Variable Name)	Individual Variable Names	Item Stem and Response Options	Item Text
Disruptive Lessons, Mathematics (ASBGDML)	ASBM04A ASBM04B ASBM04C ASBM04D ASBM04E ASBM04F	How often do these things happen in your mathematics lessons? • Every or almost every lesson • About half the lessons • Some lessons • Never	a) Students don't listen to what the teacher says b) There is too much noise for students to work well c) My teacher has to wait a long time for students to be quiet d) Students interrupt the teacher e) Students do not follow the classroom rules f) Other students' behavior makes it hard for me to concentrate
Disruptive Lessons, Science (ASBGDSL)	ASBS09A ASBS09B ASBS09C ASBS09D ASBS09E ASBS09F	How often do these things happen in your science lessons? • Every or almost every lesson • About half the lessons • Some lessons • Never	a) Students don't listen to what the teacher says b) There is too much noise for students to work well c) My teacher has to wait a long time for students to be quiet d) Students interrupt the teacher e) Students do not follow the classroom rules f) Other students' behavior makes it hard for me to concentrate
Instructional Clarity, Mathematics (ASBGICM)	ASBM03A ASBM03B ASBM03C ASBM03D ASBM03E ASBM03F ASBM03G	How much do you agree with these statements about your mathematics lessons? • Agree a lot • Agree a little • Disagree a little • Disagree a lot	a) My teacher makes it clear what we should learn in each lesson b) My teacher is easy to understand c) My teacher has clear answers to my questions d) My teacher is good at explaining mathematics e) My teacher does a variety of things to help us learn f) My teacher explains a topic again when we don't understand g) My teacher gives me helpful feedback on my work
Instructional Clarity, Science (ASBGICS)	ASBS08A ASBS08B ASBS08C ASBS08D ASBS08E ASBS08F ASBS08G	How much do you agree with these statements about your science lessons? • Agree a lot • Agree a little • Disagree a little • Disagree a lot	a) My teacher makes it clear what we should learn in each lesson b) My teacher is easy to understand c) My teacher has clear answers to my questions d) My teacher is good at explaining science e) My teacher does a variety of things to help us learn f) My teacher explains a topic again when we don't understand g) My teacher gives me helpful feedback on my work

School Questionnaire Items

Variable Label (MLCA Variable Name)	Individual Variable Names	Item Stem and Response Options	Item Text
Principal's Report, Disorder (ACBGDAS)	ACBG14A	To what degree is each of the following a problem among <fourth grade> students in your school? • Not a problem • Minor problem • Moderate problem • Serious problem	a) Arriving late at school
	ACBG14B		b) Absenteeism (i.e., unjustified absences)
	ACBG14C		c) Classroom disturbance
	ACBG14D		d) Cheating
	ACBG14E		e) Profanity
	ACBG14F		f) Vandalism
	ACBG14G		g) Theft
	ACBG14H		h) Intimidation or verbal abuse among students (including texting, emailing, etc.)
	ACBG14I		i) Physical fights among students
	ACBG14J		j) Intimidation or verbal abuse of teachers or staff (including texting, emailing, etc.)

Supportive Environments

Teacher Questionnaire Items

Variable Label (MLCA Variable Name)	Individual Variable Names	Item Stem and Response Options	Item Text
Teacher's Report, Academic Emphasis (ATBGEAS)	ATBG06A	How would you characterize each of the following within your school? • Very high • High • Medium • Low • Very low	a) Teachers' understanding of the school's curricular goals
	ATBG06B		b) Teachers' degree of success in implementing the school's curriculum
	ATBG06C		c) Teachers' expectations for student achievement
	ATBG06D		d) Teachers' ability to inspire students
	ATBG06E		e) Parental involvement in school activities
	ATBG06F		f) Parental commitment to ensure that students are ready to learn
	ATBG06G		g) Parental expectations for student achievement
	ATBG06H		h) Parental support for student achievement
	ATBG06I		i) Students' desire to do well in school
	ATBG06J		j) Students' ability to reach school's academic goals
Teacher's Report, Students Not Ready for Instruction (ATBGLSN)	ATBG06K		k) Students' respect for classmates who excel academically
	ATBG13A	In your view, to what extent do the following limit how you teach this class? • Not at all • Some • A lot	a) Students lacking prerequisite knowledge or skills
	ATBG13B		b) Students suffering from lack of basic nutrition
	ATBG13C		c) Students suffering from not enough sleep
	ATBG13D		d) Students absent from class
	ATBG13E		e) Disruptive students
	ATBG13F		f) Uninterested students
	ATBG13G		g) Distracted students
	ATBG13H		h) Students with mental, emotional, or psychological impairment
	ATBG13I		i) Students with difficulties understanding the language of instruction
Teacher's Report, School Discipline (ATBGSOS)	ATBG07A	How much do you agree or disagree with the following statements about your current school? • Agree a lot • Agree a little • Disagree a little • Disagree a lot	a) I feel safe at this school
	ATBG07B		b) This school's security policies and practices are sufficient
	ATBG07C		c) The students behave in an orderly manner
	ATBG07D		d) The students are respectful of the teachers
	ATBG07E		e) The students respect school property
	ATBG07F		f) This school has clear rules about student conduct
	ATBG07G		g) This school's rules are enforced in a fair and consistent manner

Student Questionnaire Items

Variable Label (MLCA Variable Name)	Individual Variable Names	Item Stem and Response Options	Item Text
Student Sense of Belonging (ASBGSSB)	ASBG13A ASBG13B ASBG13C ASBG13D ASBG13E ASBG13G	What do you think about your school? Tell how much you agree with these statements. • Agree a lot • Agree a little • Disagree a little • Disagree a lot	a) I like being in school b) I feel safe when I am at school c) I feel like I belong at this school d) Teachers at this school care about me e) I am proud to go to this school g) Students at this school like me the way I am
Students Arrive Hungry/Tired (ASBG09)	ASBG09A ASBG09B	How often do you feel this way when you arrive at school? • Every day • Almost every day • Sometimes • Never	a) I feel tired b) I feel hungry
Student Bullying (ASBGSB)	ASBG14A ASBG14B ASBG14C ASBG14D ASBG14E ASBG14F ASBG14G ASBG14H ASBG14I ASBG14J ASBG14K	During this school year, how often have other students from your school done any of the following things to you in person, through messaging, or through social media? • At least once a week • Once or twice a month • A few times a year • Never	a) Made fun of me or called me names b) Left me out of their games or activities c) Spread lies about me d) Stole something from me e) Damaged something of mine on purpose f) Hit or hurt me (e.g., shoving, hitting, kicking) g) Made me do things I didn't want to do h) Sent me nasty or hurtful messages online i) Shared nasty or hurtful messages about me online j) Shared embarrassing photos of me online k) Threatened me

School Questionnaire Items

Variable Label (MLCA Variable Name)	Individual Variable Names	Item Stem and Response Options	Item Text
Principal's Report, Academic Emphasis (ACBGEAS)	ACBG12A	How would you characterize each of the following within your school? • Very high • High • Medium • Low • Very low	a) Teachers' understanding of the school's curricular goals
	ACBG12B		b) Teachers' degree of success in implementing the school's curriculum
	ACBG12C		c) Teachers' expectations for student achievement
	ACBG12D		d) Teachers' ability to inspire students
	ACBG12E		e) Parental involvement in school activities
	ACBG12F		f) Parental commitment to ensure that students are ready to learn
	ACBG12G		g) Parental expectations for student achievement
	ACBG12H		h) Parental support for student achievement
	ACBG12I		i) Students' desire to do well in school
	ACBG12J		j) Students' ability to reach school's academic goals
	ACBG12K		k) Students' respect for classmates who excel academically
Principal's Report, Limited Resources (ACBGMRs)	ACBG11AA	How much is your school's capacity to provide instruction affected by a shortage or inadequacy of the following? [General School Resources]	a) Instructional materials (e.g., textbooks)
	ACBG11AB		b) Supplies (e.g., papers, pencils, materials)
	ACBG11AC		c) School buildings and grounds
	ACBG11AD		d) Heating/cooling and lighting systems
	ACBG11AE		e) Instructional space (e.g., classrooms)
	ACBG11AF		f) Technologically competent staff
	ACBG11AG		g) Audio-visual resources for delivery of instruction (e.g., interactive white boards, digital projectors)
	ACBG11AH		h) Digital devices for student use
	ACBG11BA	How much is your school's capacity to provide instruction affected by a shortage or inadequacy of the following? [Resources for Mathematics Instruction]	a) Teachers with a specialization in mathematics
	ACBG11BB		b) Computer software/applications for mathematics instruction
	ACBG11BC		c) Library resources relevant to mathematics instruction
	ACBG11BD		d) Calculators for mathematics instruction
	ACBG11BE		e) Concrete objects or materials to help students understand quantities or procedures

Stakeholder Satisfaction

Teacher Questionnaire Items

Variable Label (MLCA Variable Name)	Individual Variable Names	Item Stem and Response Options	Item Text
Teacher Job Satisfaction (ATBGTJS)	ATBG08A	How often do you have these feelings about being a teacher? • Very often • Often • Sometimes • Never or almost never	a) I am content with my profession as a teacher
	ATBG08B		b) I find my work full of meaning and purpose
	ATBG08C		c) I am enthusiastic about my job
	ATBG08D		d) My work inspires me
	ATBG08E		e) I am proud of the work I do
	ATBG08F		f) I feel appreciated as a teacher
	ATBG08G		g) I enjoy the challenges of teaching
Teacher's Report Non-Instruction Responsibilities (ATBG09)	ATBG09A	How much do you agree or disagree with the statements below? • Agree a lot • Agree a little • Disagree a little • Disagree a lot	a) There are too many students in the classes
	ATBG09B		b) I have too much material to cover in class
	ATBG09C		c) I have too many teaching hours
	ATBG09D		d) I need more time to prepare for class
	ATBG09E		e) I need more time to assist individual students
	ATBG09F		f) I feel too much pressure from parents
	ATBG09G		g) I have difficulty keeping up with all of the changes to the curriculum
	ATBG09H		h) I have too many administrative tasks

Student Questionnaire Items

Variable Label (MLCA Variable Name)	Individual Variable Names	Item Stem and Response Options	Item Text
Students Like Mathematics (ASBGSLM)	ASBM02A ASBM02D ASBM02E ASBM02F ASBM02G ASBM02H ASBM02I	How much do you agree with these statements about learning mathematics? • Agree a lot • Agree a little • Disagree a lot • Disagree a little	a) I enjoy learning mathematics d) I learn many interesting things in mathematics e) I like mathematics f) I like any schoolwork that involves numbers g) I like to solve mathematics problems h) I look forward to mathematics lessons i) Mathematics is one of my favorite subjects
Students Confident, Mathematics (ASBGSCM)	ASBM05A ASBM05B ASBM05C ASBM05D ASBM05E ASBM05F ASBM05G	How much do you agree with these statements about mathematics? • Agree a lot • Agree a little • Disagree a lot • Disagree a little	a) I usually do well in mathematics b) Mathematics is harder for me than for many of my classmates c) I am just not good at mathematics d) Mathematics is easy for me e) I am good at working out difficult mathematics problems f) Mathematics is harder for me than any other subject g) Mathematics makes me confused
Students Like Science (ASBGSLS)	ASBS07A ASBS07B ASBS07C ASBS07D ASBS07E ASBS07F ASBS07G ASBS07H ASBS07I	How much do you agree with these statements about learning science? • Agree a lot • Agree a little • Disagree a lot • Disagree a little	a) I enjoy learning science d) I learn many interesting things in science e) I like science f) I look forward to learning science in school g) Science teaches me how things in the world work h) I like to do science experiments i) Science is one of my favorite subjects
Students Confident, Science (ASBGSCS)	ASBS10A ASBS10B ASBS10C ASBS10D ASBS10E ASBS10F	How much do you agree with these statements about science? • Agree a lot • Agree a little • Disagree a lot • Disagree a little	a) I usually do well in science b) Science is harder for me than for many of my classmates c) I am just not good at science d) Science is easy for me e) Science is harder for me than any other subject f) Science makes me confused

School Questionnaire Items

Variable Label (MLCA Variable Name)	Individual Variable Names	Item Stem and Response Options	Item Text
Principal's Report, Teacher Absenteeism (ACBG15)	ACBG15A ACBG15B	To what degree is each of the following a problem among teachers in your school? • Not a problem • Minor problem • Moderate problem • Serious problem	a) Arriving late or leaving early b) Absenteeism

Home Questionnaire Items

Variable Label (MLCA Variable Name)	Individual Variable Names	Item Stem and Response Options	Item Text
Parents Satisfied with School (ASBHPSP)	ASBH08A ASBH08B ASBH08C ASBH08D ASBH08E ASBH08F ASBH08G ASBH08H	What do you think of your child's school? • Agree a lot • Agree a little • Disagree a lot • Disagree a little	a) My child's school does a good job including me in my child's education b) My child's school provides a safe environment c) My child's school cares about my child's progress in school d) My child's school does a good job informing me of his/her progress e) My child's school promotes high academic standards f) My child's school does a good job in helping him/her become better in reading g) My child's school does a good job in helping him/her become better in mathematics h) My child's school does a good job in helping him/her become better in science

Appendix C: Technical Details

Multilevel Latent Class Analysis

Latent class analyses are statistical models that extract group similarities by identifying subgroups that share response patterns on indicator variables for the model.²⁷ Schools in the same country are likely to provide more similar responses than schools across countries.²⁸ In the context of international large-scale assessments such as TIMSS, where data are collected using complex sampling designs involving nested structures (classrooms within schools, and schools within countries), a multilevel specification of latent class models is more appropriate.²⁹ This is because schools within the same country are likely to exhibit more similar response patterns due to shared curricula, educational policies, and cultural factors, compared to schools across different countries.³⁰ As such, multilevel extensions of latent structure models are well-suited for capturing these hierarchical dependencies.

In this report, Level 1 represents schools and Level 2 represents countries in the multilevel structure of the data. Individual-level response data from teachers, students, and parents is aggregated up to the school level for our analysis. Ultimately each school climate health latent cluster in the final model was described based on variation in the indicator variables from the TIMSS 2023 Context Questionnaires. Compared to a latent class analysis performed on Level-1 units sampled randomly from a population, MLCA also groups Level-2 units based on similarities in the Level-1 characteristics within. The MLCA model is based on a probabilistic model that allows the log-odds of Level-1 cluster proportions to vary across countries. The clusters identified by this methodology are mutually exclusive. Just as no school acts as a representative of multiple countries, the schools at Level 1 are assumed to be members of just one of the clusters that are identified in the analysis. The steps to finalize the model are to determine the number of Level-2 classes and then re-estimate the Level-1 clusters under that restriction.

Given Y_{ijk} is the aggregated school measure for school i within country j on item k , where the number of countries is J , the number of schools in each country j is n_j and the total number of items is K . A response on item k is given by s_k and the maximum possible value is denoted by S_k . The latent class variable for schools is X_{ij} and an individual class is t with the number of latent classes given as T . Suppose Y_{ij} is the vector of responses of school i in country j , and s a single possible response pattern, then the latent class model is expressed as,

$$P(Y_{ij}=s) = \sum_{t=1}^T P(X_{ij}=t)P(Y_{ij}=s|X_{ij}=t) = \sum_{t=1}^T P(X_{ij}=t) \prod_{k=1}^K P(Y_{ijk}=s_k|X_{ij}=t)$$

The above definition applies to both standard and multilevel latent cluster analyses.³¹ Distinctions are evident in the logit equations, below first assuming independence across countries J ,

$$P(X_{ij}=t) = \frac{\exp(\gamma_t)}{\sum_{r=1}^T \exp(\gamma_r)} P(Y_{ijk}=s_k|X_{ij}=t) = \frac{\exp(\beta_{s_k t}^k)}{\sum_{r=1}^{S_k} \exp(\beta_{rt}^k)}$$

Then, the addition of j to the subscripts of β and γ terms distinguishes the multilevel equations from the otherwise identical definition,

$$P(X_{ij}=t) = \frac{\exp(\gamma_{tj})}{\sum_{r=1}^T \exp(\gamma_{rj})} P(Y_{ijk}=s_k | X_{ij}=t) = \frac{\exp(\beta_{s_k t j}^k)}{\sum_{r=1}^{S_k} \exp(\beta_{r t j}^k)}$$

Non-parametric model. Two main approaches exist in specifying an MLCA: the parametric model and the non-parametric model. The parametric approach introduces random effects rather than estimating separate fixed effects for each Level-2 unit.^{32,33} In other words, the effects specific to each Level-2 unit are treated as random variables drawn from a specified distribution, typically a normal distribution. In contrast, the non-parametric approach avoids the need to assume a specific distribution form. Instead, it specifies an additional latent class model at Level-2, where the Level-2 units are clustered into a discrete number of latent classes, each associated with different Level-1 class membership distributions. In doing so, distributional assumptions like normality are relaxed and computational burden is reduced. Therefore, this study adopts the non-parametric approach.

Suppose W_j is the class assignment for country j given the discrete mixture distribution, then the model for the latent class probability is

$$P(X_{ij}=t | W_j=m) = \frac{\exp(\gamma_{tm})}{\sum_{r=1}^T \exp(\gamma_{rm})},$$

where m denotes a particular mixture component.

The analysis followed roughly the following steps for the non-parametric model, adapted from Bakk, di Mari, Oser, and Kuha (2022):³⁴

- Step 1: Simple Latent Class Analysis
 - Find the number of latent classes in the data without any clustering structure assumed to be present
- Step 2a: Estimate classes at Level 2 keeping Level 1 fixed
- Step 2b: Estimate classes at Level 1 with classes at Level 2 included
- Step 2c: Choose a model

Hereafter, we refer to Level-1 latent classes as “clusters” and Level-2 latent classes as “classes.” The true assignment of observations (schools) to clusters is unknown but can be estimated by calculating the posterior probabilities of belonging to a cluster based on the response pattern observed for each school. Thus, for each school the probability is calculated that its true cluster X is a specific cluster t given that its response pattern (Y) is observed to be y :

$$P(X=t | Y=y) = \frac{P(X=t)P(Y=y | X=t)}{P(Y=y)}$$

For each school, the posterior probability of belonging to each cluster is calculated, and the assignment is determined using the proportional or modal method. Proportional assignment places schools sharing response patterns y into clusters proportional to the probabilities calculated for each cluster. In contrast, the modal approach assigns all schools sharing y to the single cluster with the highest posterior probability. This research employed the modal method as the number of indicators specified and the multilevel specification resulted in singular response

patterns. Classification error is present in both methods of assignment and both approaches are frequently employed.³⁵ At Level 2, countries are assigned to classes based on the extent two countries have similar proportions of cluster membership across their schools.

Preparing the Data

Each of the 22 indicator variables required recodes for the MLCA analysis. Although the original response variables for each questionnaire item are categorical, calculating scale scores and aggregating to the school level resulted in either continuous variables or categorical variables with too many categories for meaningful analysis. Exhibits C.1-C.3 give unweighted descriptive statistics for the input variables aggregated to the school level.

Exhibit C.1: Descriptive Statistics (Unweighted) – Principal Indicators

	ACBG15	ACBGDAS	ACBGEAS	ACBGMRS
Valid	10863	11000	11050	11063
Missing	1152	1015	965	952
Mean	1.58	9.79	10.10	10.21
Standard Deviation	0.80	1.89	2.19	2.42
Minimum	1	3.74	1.43	2.77
Median	1.50	9.74	9.84	10.16
Maximum	4	12.58	16.36	15.36

Exhibit C.2: Descriptive Statistics (Unweighted) – Teacher Indicators

	ATBG09	ATBG12	ATBGEAS	ATBGLSN	ATBGSOS	ATBGTJS
Valid	11200	11395	11428	11387	11397	11385
Missing	815	620	587	628	618	630
Mean	2.22	1.58	10.23	9.79	10.28	10.19
Standard Deviation	0.54	0.40	2.12	1.86	1.92	1.77
Minimum	1	1	2.18	3.23	4.16	4.59
Median	2.19	1.57	10.07	9.82	10.39	10.37
Maximum	4	4	17.02	15.17	13.06	12.30

Exhibit C.3: Descriptive Statistics (Unweighted) – Student Indicators

	ASBG09	ASBGDML	ASBGDSL	ASBGICM	ASBGICS	ASBGSB
Valid	11970	11781	11779	11964	11962	11968
Missing	45	234	236	51	53	47
Mean	2.48	9.92	9.89	10.04	10.04	9.85
Standard Deviation	0.34	1.05	1.06	0.89	0.88	0.90
Minimum	1	6.16	6.44	2.46	2.57	4.39
Median	2.47	9.80	9.77	10.05	10.07	9.85
Maximum	4	14.84	13.92	12.07	11.90	12.77

	ASBGSCM	ASBGSCS	ASBGSLM	ASBGSLS	ASBGSSB	ASBHPS
Valid	11959	11958	11964	11960	11969	10864
Missing	56	57	51	55	46	1151
Mean	9.98	9.99	10.13	10.11	10.07	10.19
Standard Deviation	0.70	0.82	0.93	0.90	0.94	0.99
Minimum	6.68	5.97	4.93	4.42	6.72	4.46
Median	9.93	9.93	10.11	10.12	10.02	10.28
Maximum	14.24	13.79	12.59	12.38	12.86	12.18

The desired balance in executing the recodes was to maintain variability in the response patterns, but also support interpretability in evaluating the model results. Thus, cutpoints were determined such that the recodes would result in three regions of responses that were reasonably equal in size. These cutpoints informed the resulting regions (0, 1, 2) as summarized per respondent category in Exhibits C.4-C.6. In the interpretation, these categories are interpreted as “Low,” “Medium,” and “High,” respectively. All 22 indicator variables were included as ordinal variables with three categories each following the recode described earlier. The weighting variable appropriate for school-level analysis (schwt) was included.

Exhibit C.4: Cutpoints for Principal Indicators for Recode (Weighted)

	ACBG15	ACBGDAS	ACBGEAS	ACBGMRS
Cutpoint 1	1.00	9.15	8.86	9.13
Cutpoint 2	1.50	10.45	10.50	10.74

Exhibit C.5: Cutpoints for Teacher Indicators for Recode (Weighted)

	ATBG09	ATBG12	ATBGEAS	ATBGLSN	ATBGSOS	ATBGTJS
Cutpoint 1	2.00	1.36	8.99	8.86	9.32	9.15
Cutpoint 2	2.50	1.71	10.62	10.34	11.10	11.21

Exhibit C.6: Cutpoints for Student Indicators for Recode (Weighted)

	ASBG09	ASBGDML	ASBGDSL	ASBGICM	ASBGICS	ASBGSB
Cutpoint 1	2.28	9.35	9.31	9.68	9.68	9.49
Cutpoint 2	2.63	10.19	10.23	10.37	10.42	10.23
	ASBGSCM	ASBGSCS	ASBGSLM	ASBGSLS	ASBGSSB	ASBHPSP
Cutpoint 1	9.56	9.61	9.73	9.77	9.57	9.82
Cutpoint 2	10.21	10.33	10.59	10.61	10.37	10.78

Item-level missingness was common throughout and was treated as missing when aggregating the data to the school level or when calculating scores on items not included in scales. Missingness before aggregation could be missing responses for entire sets of items or scales or be missing for some items within a group of items or by respondent. Any available data were used in assigning schools to clusters; if data was missing for a particular variable, schools were assigned to clusters based on the variables where data were available.

Model Selection Details

When comparing specified models to evaluated different numbers of clusters or classes estimated, there is no target value or indicator of sufficiency. Instead, criteria such as the Bayesian Information Criterion (BIC; Schwarz, 1978³⁶) can be used to compare models, with lower BIC values indicating improvement in fit. A turning point in BIC can signal that adding more clusters no longer meaningfully improves the model. The classification error is computed as a weighted average of the probability that each school was not assigned to its true cluster, approximated as one minus the maximum posterior probability averaged across all schools.³⁷ Modeling decisions can also be supported by entropy-based measures of classification quality, such as entropy R^2 . With a range from 0 to 1, entropy R^2 gives a summary of the qualities of the model to correctly assign cases to clusters. Values over 0.8 are often considered indicative of good classification accuracy.³⁸ Different indicators may be more or less influential in deciding on final models. The

first step of this analysis, as outlined by Bakk, di Mari, Oser, and Kuha, (2022),³⁹ is to judge the fit of the models specifying different numbers of clusters at the school level (Level 1). In this analysis, the BIC, classification error, and entropy R^2 were compared to guide the model selection process.⁴⁰

For the Level 1, models specifying up to ten clusters were estimated—results are included as Exhibit C.7. The BIC decreases with each additional cluster, which suggests improved balance between model fit and parsimony, although the improvements became notably smaller after the 4-cluster model. The other diagnostic statistics provide additional information in deciding what model might be best. Classification error is lowest in the simplest model (2-cluster; 0.044), and the entropy R^2 is highest here too (2-cluster; 0.845), but none are below the entropy R^2 threshold of 0.8 such that no model in the set is flagged for poor performance based on the indicator alone. With each added cluster, the classification error goes up but not at a constant rate. There is a relatively large jump between the 2-cluster and 3-cluster models (0.044 to 0.072) and again between 3-clusters and 4-clusters (0.072 to 0.094). The entropy R^2 is steady across the 4-, 5-, and 6-cluster models (0.826, 0.824, 0.821, respectively). As noted in the main text of this report, the 4-cluster model was selected at this stage as it had favorable diagnostic characteristics and more complex models specified clusters that would contain very few schools.

Exhibit C.7: Level-1 Modeling Results

	LL	BIC(LL)	Npar	L ²	df	Class.Err.	Entropy R ²
1-Cluster	-276458	553330	44	385328	11971	0	1
2-Cluster	-262996	526622	67	358404	11948	0.044	0.845
3-Cluster	-258171	517187	90	348753	11925	0.072	0.837
4-Cluster	-254977	511015	113	342365	11902	0.094	0.826
5-Cluster	-252684	506646	136	337779	11879	0.110	0.824
6-Cluster	-251176	503846	159	334764	11856	0.123	0.821
7-Cluster	-249782	501274	182	331976	11833	0.121	0.826
8-Cluster	-248542	499010	205	329496	11810	0.133	0.824
9-Cluster	-247604	497350	228	327619	11787	0.149	0.817
10-Cluster	-246737	495832	251	325885	11764	0.153	0.820

Fixing the number of Level-1 clusters to 4, models with 1-5 Level-2 group classes (GClass) were specified (results in Exhibit C.8). Evaluating the model diagnostics with group-level classes follows a similar approach. Both the BIC and the classification error decrease with each additional country class specified. The entropy R^2 increases between the 2-GClass and 3-GClass models but is steady at around $R^2 = 0.85$ for more complex models.

Exhibit C.8: Level-2 Modeling Results

	LL	BIC(LL)	Npar	Class.Err.	Entropy R ²
1-GClass 4-Cluster	-9544239	19089950	113	0.094	0.826
2-GClass 4-Cluster	-9457708	18916939	117	0.084	0.844
3-GClass 4-Cluster	-9400684	18802943	121	0.081	0.852
4-GClass 4-Cluster	-9392181	18785990	125	0.080	0.855
5-GClass 4-Cluster	-9386358	18774396	129	0.079	0.857

While the 3- and 4-GClass models are similar in terms of model diagnostic statistics, the model assignment of countries to classes reveals a clearer difference. For the 4-Gclass

model, the fourth country class only has eight countries which is about 14 percent of the total number of countries. With 58 countries in the sample, one group with eight countries would be much smaller than the other three and would stand out simply for being a more homogenous group of countries. Whereas the three-class solution results in groups of relatively equal size with diversity within the groups as well, a solution with one substantially smaller group could introduce more distraction than information. To avoid the groupings being too small, we decided to select the 3-Gclass model.

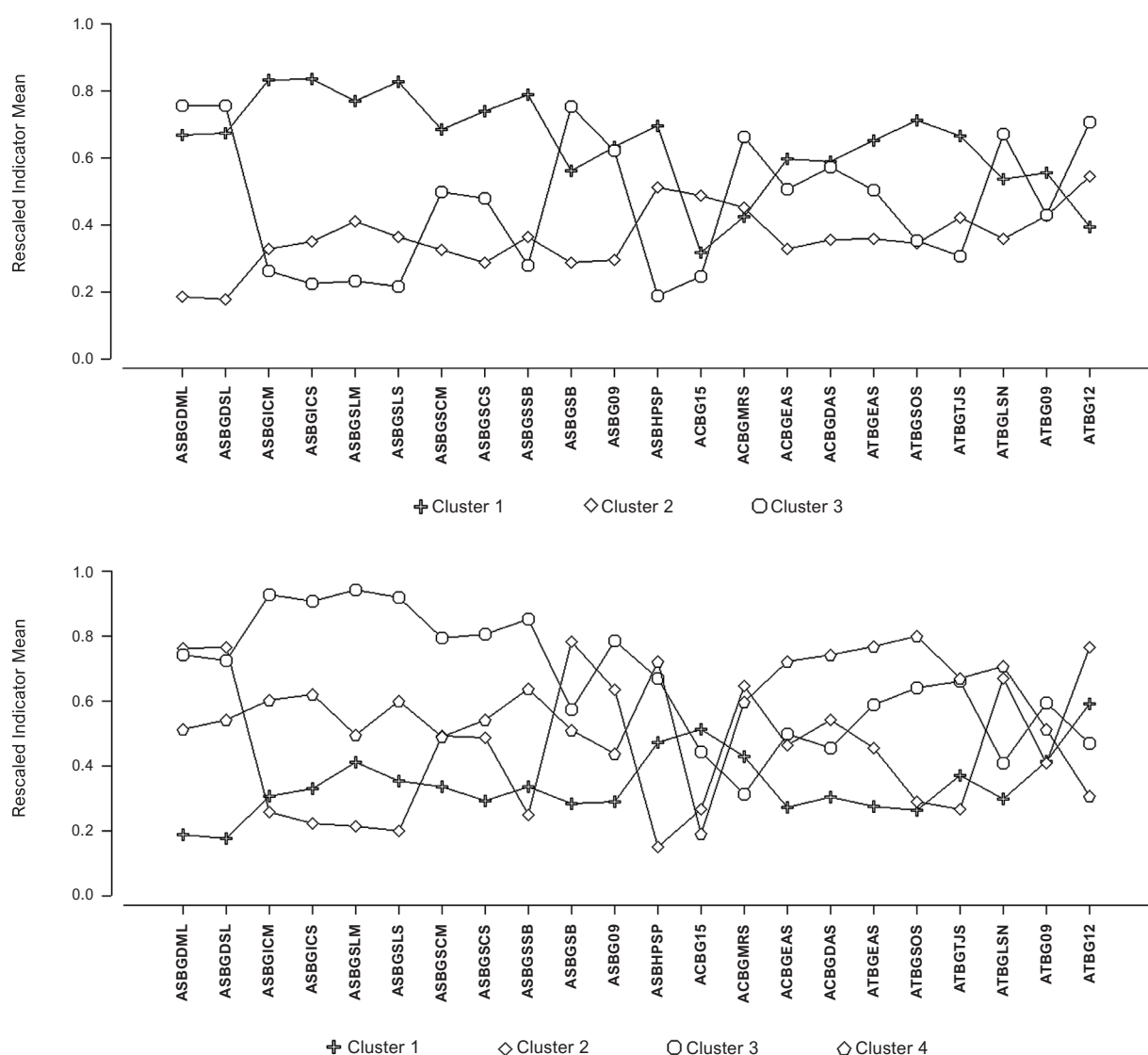
Re-estimating the school-level models at Level 1 with the 3-GClass specification is the last stage of the modeling process. The resulting model fit statistics are included in Exhibit C.9. The 4-Cluster, 3-GClass model has the lowest BIC, but the highest classification error compared to the less complex models. The entropy R^2 is roughly the same for the 2- and 3-Cluster model with 3-GClasses but dips for the 4-Cluster model.

Exhibit C.9: Re-estimating Level-1 Model Results

	LL	BIC(LL)	Npar	Class.Err.	Entropy R^2
3-GClass 1-Cluster	-10348332	20697262	46	0	1
3-GClass 2-Cluster	-9777234	19555392	71	0.038	0.863
3-GClass 3-Cluster	-9546470	19094190	96	0.060	0.864
3-GClass 4-Cluster	-9400684	18802943	121	0.081	0.852

The Level-1 profile graph of the 3-GClass 3- and 4- cluster solution is helpful in making a determination. The additional cluster does not impact the shared characteristics between the 3- and 4- cluster figures. Interpreting the 4-cluster graph at once is less straightforward than the apparent “high,” “middle,” “low” pattern of the 3-cluster graphs (Exhibit C.10). However, isolating the “high” cluster in the 4-cluster version alongside the fourth cluster illuminates its distinctions. From interpreting the graph, the additional cluster describes schools where students are less satisfied and content than in the “high” cluster, whereas the teachers, parents, and principals are the most satisfied of any cluster.

Exhibit C.10: 3-GClass 3- and 4- Cluster Solutions



* Cluster numbers in this exhibit reflect LatentGOLD's internal class labels for each model specification. Because these labels are assigned independently, the numeric cluster identifiers do not align across the 3- and 4-cluster solutions. The interpretation in the text links clusters based on their shared profile characteristics, rather than on the numeric cluster IDs shown in the plots.

This differential in experience is an interesting result worthy of further investigation given that it describes a phenomenon directly referenced in the literature—that overall evaluations of quality are complicated by differing views across stakeholder categories. Thus, the 4-Cluster, 3-Class model is selected as the final model from the analysis. Readers interested in conducting further explorations with the school cluster classifications can download the data and accompanying documentation below.

[Download the school cluster classifications.](#)

Contributions of Individual Indicators

The extent that a variable contributed to the distinctness of the cluster definitions is helpful to understand before describing the clusters themselves based on the school-level response data from principals, teachers, students, and parents. For each variable, the pseudo- R^2 can be interpreted as a measure of ‘variance explained’ and is defined as the ratio of between-cluster variation in a given variable and the total variance (for further detail, see Vermunt & Magidson, 2006).⁴¹ These results are presented in descending order by R^2 in Exhibit C.11.

Exhibit C.11: Indicator Variables by R^2

Indicator	R^2
Students Like Science	0.42
Instructional Clarity, Mathematics	0.41
Instructional Clarity, Science	0.41
Students Like Mathematics	0.40
Disruptive Lessons, Science	0.36
Disruptive Lessons, Mathematics	0.35
Student Sense of Belonging	0.33
Teacher's Report, School Discipline	0.29
Parents Satisfied with School	0.29
Students Arrive Hungry/Tired	0.22
Student Bullying	0.21
Students Confident, Science	0.21
Teacher's Report, Academic Emphasis	0.20
Teacher Job Satisfaction	0.19
Teacher's Report, Students Not Ready for Instruction	0.18
Students Confident, Mathematics	0.16
Principal's Report, Academic Emphasis	0.15
Engaging Teacher Practices	0.15
Principal's Report, Disorder	0.15
Principal's Report, Limited Resources	0.10
Principal's Report, Teacher Absenteeism	0.09
Teacher's Report, Non-Instruction Responsibilities	0.03

Most of the indicator variables included in the model discriminated well between the clusters in the model selected for this report. The student variables (aggregated to the school level) had the highest R^2 values, with teacher, parent and school variables following with lower values. Low R^2 values, such as for *Teacher's Report, Non-Instruction Responsibilities* with only three percent of variance explained, do not imply that it should be excluded from the model, but instead that further variation exists in the data that is not explained by the separation of schools across clusters alone. By contrast, variance in the indicators with high R^2 is better explained by the school groupings—the groups are distinct in their observed response pattern on that variable to account for more of the overall difference observed in that indicator.

School Climate Health Clusters and Achievement

After schools were assigned to clusters based on the modal approach described earlier, the average mathematics and science achievement for each cluster within each country was estimated using the IEA IDB Analyzer (version 5.0.48).⁴² The mean achievement for each cluster

was calculated as the weighted average across plausible values, and the corresponding standard error incorporated both sampling and imputation variance components.

To ensure the reliability of the reported statistics, achievement estimates (mean and standard error) and the standard error of the percentage were not reported when fewer than five schools in a country were assigned to a given cluster. This threshold is based on the rationale that data from fewer than five schools are insufficient to produce stable estimates of variance and cluster-level average achievement.

School Climate Health Clusters and School-Level Average Home Socioeconomic Status

The analysis of school climate health clusters and school-level average home socioeconomic status follows an approach similar to that used for achievement. The home socioeconomic status variable was derived from the *Home Socioeconomic Status* scale in the Home Questionnaire, which was based on parents' reports of the number of books and children's books in the home, as well as their self-reported levels of education and occupation. The student-level scale scores were aggregated to the school level by computing the weighted mean of students' scores within each school, representing the overall home background of students enrolled in that school. Using the IEA IDB Analyzer, the average school-level socioeconomic status and its standard error were then estimated for each cluster within each country. Consistent with the reporting rule applied to achievement data, socioeconomic status estimates (mean and standard error) and the standard error of the percentage were not reported when fewer than five schools in a country were assigned to a given cluster.

To quantify the proportion of variance in the school average home socioeconomic status explained by the clusters, weighted one-way analyses of variance were conducted separately for each country using the IEA IDB Analyzer. The resulting country-specific explained variance values were then averaged across countries, excluding countries where home socioeconomic status data were available for less than 50 percent of the students.

References

- 1 Müller, F., Denk, A., Lubaway, E., Sälzer, C., Kozina, A., Perše, T.V., Rasmusson, M., Jugović, I., Nielsen, B. L., Rozman, M., Ojsteršek, A., & Jurko, S. (2020). Assessing social, emotional, and intercultural competences of students and school staff: A systematic literature review. *Educational Research Review*, 29, 100304. <https://doi.org/10.1016/j.edurev.2019.100304>
- 2 Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, 95(3), 542–575. <https://doi.org/10.1037/0033-2909.95.3.542>
- 3 Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist*, 55(1), 34–43. <https://doi.org/10.1037/0003-066X.55.1.34>
- 4 Ryff, C.D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological wellbeing. *Journal of Personality and Social Psychology*, 57(6), 1069–1081. <https://doi.org/10.1037/0022-3514.57.6.1069>
- 5 Seligman, M. E. P. (2002). *Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment*. The Free Press.
- 6 Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate Research. *Review of Educational Research*, 83(3), 357–385. <https://doi.org/10.3102/0034654313483907>
- 7 Ramelow, D., Currie, D., & Felder-Puig, R. (2015). The assessment of school climate: Review and appraisal of published student-report measures. *Journal of Psychoeducational Assessment*, 33(8), 731–743. <https://doi.org/10.1177/0734282915584852>
- 8 Aldrich, C.E.A. (2024). *Towards holistic evaluation of education systems: Using TIMSS 2023 context data to classify schools by school climate health* (Order No. 31487997). [Doctoral dissertation, Boston College]. ProQuest Dissertations & Theses Global.
- 9 Zandvliet, D.B., Stanton, A., & Dhaliwal, R. (2019). Design and validation of a tool to measure associations between the learning environment and student well-being: The healthy environments and learning practices survey (HELPS). *Innovative Higher Education*, 44(4), 283–297. <https://doi.org/10.1007/s10755-019-9462-6>
- 10 Gates, E. F., Walton, M., Vidueira, P., & McNall, M. (2021). Introducing systems- and complexity-informed evaluation. *New Directions for Evaluation*, 170, 13–25. <https://doi.org/10.1002/ev.20466>
- 11 Zazueta, A. E., Le, T. T., & Bahramalian, N. (2021). Development trajectories and complex systems-informed theories of change. *American Journal of Evaluation*, 42(1), 110–129. <https://doi.org/10.1177/1098214020947782>
- 12 Diener, E., Lucas, R. E., & Oishi, S. (2018). Advances and open questions in the science of subjective wellbeing. *Collabra: Psychology*, 4(1), 15. <https://doi.org/10.1525/collabra.115>
- 13 Allen, K.A., Gallo Cordoba, B., Ryan, T., Arslan, G., Slaten, C. D., Ferguson, J. K., Bozoglan, B., Abdollahi, A., & Vella-Brodrick, D. (2022). Examining predictors of school belonging using a socio-ecological perspective. *Journal of Child and Family Studies*. <https://doi.org/10.1007/s10826-022-02305-1>
- 14 Bronfenbrenner, U. (1989). Ecological systems theory. *Annals of Child Development*, 6, 187–249.
- 15 Bronfenbrenner, U. (1992). Ecological systems theory. In *Six theories of child development: Revised formulations and current issues* (pp. 187–249). Jessica Kingsley Publishers.
- 16 Bronfenbrenner, U. (2000). Ecological systems theory. In *Encyclopedia of Psychology*, Vol. 3. (pp. 129–133). American Psychological Association. <https://doi.org/10.1037/10518-046>
- 17 UNESCO Office Bangkok and Regional Bureau for Education in Asia and the Pacific. (2016). *Happy schools! A framework for learner wellbeing in the Asia-Pacific*. <https://unesdoc.unesco.org/ark:/48223/pf0000244140>
- 18 Vermunt, J.K. (2003). 7. Multilevel latent class models. *Sociological Methodology*, 33(1), 213–239. <https://doi.org/10.1111/j.0081-1750.2003.t01-1-00131.x>

- 19 Galvez-Nieto, J.L., García, J.A., Vera-Bachmann, D., Trizano-Hermosilla, I., & Polanco, K. (2020). Multilevel latent class cluster analysis of school climate: Individual, family and community factors. *Revista de Psicodidáctica (English Ed.)*, 25(2), 85–92. <https://doi.org/10.1016/j.psicoe.2020.01.002>
- 20 Allison, K.R., Adlaf, E.M., Irving, H.M., Schoueri-Mychasiw, N., & Rehm, J. (2016). The search for healthy schools: A multilevel latent class analysis of schools and their students. *Preventive Medicine Reports*, 4, 331–337. <https://doi.org/10.1016/j.pmedr.2016.06.016>
- 21 Aldrich, C. E. A. *Towards holistic evaluation of education systems*.
- 22 Vermunt, J.K. and Magidson J. (2016). *Technical guide for Latent GOLD 5.1: Basic, Advanced, and Syntax*. Statistical Innovations Inc.
- 23 Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.
- 24 van de Vijver, F. J. R. & Leung, K. (2021). *Methods and data analysis for cross-cultural research*. 2nd ed. V. H. Fetvadjev, J. He, & J. R. J. Fontaine, Eds. Cambridge University Press.
- 25 He, J. & van de Vijver, F. J. R. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture*, 2(2). <https://doi.org/10.9707/2307-0919.1111>
- 26 Johnson, T.P. & van de Vijver, F. J. R. (2003). Social desirability in cross-cultural research. *Cross-Cultural Survey Methods*, 325, 195–204.
- 27 Bakk, Z., Tekle, F.B., & Vermunt, J.K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology*, 43(1), 272–311. <https://doi.org/10.1177/0081175012470644>
- 28 Bakk, Z., Di Mari, R., Oser, J., & Kuha, J. (2022). Two-stage multilevel latent class analysis with covariates in the presence of direct effects. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(2), 267–277. <https://doi.org/10.1080/10705511.2021.1980882>
- 29 von Davier, M. (2007). Hierarchical general diagnostic models. ETS Research Report Series, 2007: i-19. <https://doi.org/10.1002/j.2333-8504.2007.tb02061.x>
- 30 Bakk, Z., Di Mari, R., Oser, J., & Kuha, J. Two-stage multilevel latent class analysis.
- 31 Vermunt, J.K. 7. Multilevel latent class models.
- 32 Henry, K. L., & Muthén, B. (2010). Multilevel latent class analysis: An application of adolescent smoking typologies with individual and contextual predictors. *Structural Equation Modeling: a Multidisciplinary Journal*, 17(2), 193–215. <https://doi.org/10.1080/10705511003659342>
- 33 Vermunt, J.K. 7. Multilevel latent class models.
- 34 Bakk, Z., Di Mari, R., Oser, J., & Kuha, J. (2022). Two-stage Multilevel Latent Class Analysis with Covariates in the Presence of Direct Effects. *Structural Equation Modeling: A Multidisciplinary Journal*, 29(2), 267–277. <https://doi.org/10.1080/10705511.2021.1980882>
- 35 Bakk, Z., Di Mari, R., Oser, J., & Kuha, J. Two-stage multilevel latent class analysis.
- 36 Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464.
- 37 Vermunt, J.K. & Magidson J. *Technical guide for Latent GOLD 5.1*.
- 38 Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(3), 329–341. <https://doi.org/10.1080/10705511.2014.915181>
- 39 Bakk, Z., Di Mari, R., Oser, J., & Kuha, J. Two-stage multilevel latent class analysis.
- 40 Vermunt, J.K. & Magidson J. *Technical guide for Latent GOLD 5.1*.
- 41 Vermunt, J.K., & Magidson J. (2005). *Latent GOLD 4.0 User's Guide*. Statistical Innovations Inc.
- 42 IEA (2022). *Help Manual for the IEA IDB Analyzer (Version 5.0)*. www.iea.nl